Universidade Estadual de Londrina

JOÃO MARCOS FÁVARO LOPES

MULTIMODAL SPECTROSCOPIC ANALYSIS FOR EROSION ASSESSMENT IN AGRICULTURAL SOILS

LONDRINA
2026

JOÃO MARCOS FÁVARO LOPES

MULTIMODAL SPECTROSCOPIC ANALYSIS FOR EROSION ASSESSMENT IN AGRICULTURAL SOILS

Thesis presented to the Postgraduate Program in Physics at the State University of Londrina, as a partial requirement for obtaining the doctor's degree.

Supervisor:	Prof. Dr. Avacir Casanova Andrello
Co-supervisor:	Prof. Dr. Fábio Luiz Melquiades

LONDRINA
2026

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Lopes, João Marcos Fávaro.
Multimodal Spectroscopic Analysis for Erosion Assessment in Agricultural Soils / João Marcos Fávaro Lopes. - Londrina, 2026.
146 f. : il.

Orientador: Avacir Casanova Andrello.
Coorientador: Fábio Luiz Melquiades.
Tese (Doutorado em Física) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Física, 2026.
Inclui bibliografia.

1. Física Nuclear Aplicada - Tese. 2. Análise Multivariada - Tese. 3. Fusão de Dados - Tese. 4. Redistribuição de Solo - Tese. I. Andrello, Avacir Casanova. II. Melquiades, Fábio Luiz. III. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Física. IV. Título.

CDU 53

JOÃO MARCOS FÁVARO LOPES

MULTIMODAL SPECTROSCOPIC ANALYSIS FOR EROSION ASSESSMENT IN AGRICULTURAL SOILS

Thesis presented to the Postgraduate Program in Physics at the State University of Londrina, as a partial requirement for obtaining the doctor's degree.

Examination Board

Prof. Dr. Avacir Casanova Andrello
Universidade Estadual de Londrina

Prof. Dr. Eduardo Inocente Jussiani
Universidade Estadual de Londrina

Prof. Dr. Paulo Sérgio Parreira
Universidade Estadual de Londrina

Prof. Dr. Felipe Rodrigues dos Santos
Universidade Estadual do Centro-Oeste

Dra. Lina Du
International Atomic Energy Agency

Londrina, 25 June 2026.

"The most important step a man can take.
It's not the first one, is it? It's the next one.
Always the next step."
- Dalinar Kholin, in *Oathbringer* by
Brandon Sanderson.

ACKNOWLEDGEMENTS

Writing the acknowledgements section of a work that spanned four years is no easy task. So many people were part of the development of this thesis in one way or another, and there is always the risk of forgetting to mention someone. However, I will try my best in the following pages.

First and foremost, I would like to thank my supervisor and co-supervisor, Professors Avacir Andrello and Fábio Melquiades, respectively, for the years of mentorship, guidance, and friendship since the beginning of my master's degree. I cannot put into words how much I value the knowledge you have shared with me, nor how grateful I am for the opportunities you have provided throughout this journey.

While I am still talking about the laboratory, I would also like to thank my friends for the companionship, friendship, philosophical discussions, bad jokes, coffee breaks, lunch breaks, ice cream breaks, game breaks, and everything in between. I would also like to thank Professors Carlos Appoloni and Eduardo Inocente for the many fruitful and pleasant conversations over the years.

I am deeply grateful to my parents and sister, Maria Regina Fávaro, Silvio Lopes, and Rafaela Lopes, for their constant support during this period. I know I can be difficult to deal with when I am stressed, so thank you for your patience, understanding, and for always encouraging me to pursue yet another degree. I am also thankful to my cats, Mimi (R.I.P.) and Yoda, for their faithful attendance at so many presentation rehearsals over the years.

I would also like to acknowledge the friends I made during my bachelor's degree, Lucas Maquedano and Pedro Haerter. Thank you for a friendship that has endured over the years. Who would have imagined that the students we were back then would one day be traveling around the world and pursuing our PhDs?

Regarding my time abroad, I am deeply grateful to the International Atomic Energy Agency (IAEA) for welcoming me and for providing an environment that not only supported my research, but also broadened my perspective of the world. From the Nuclear Sciences and Instrumentation Laboratory (NSIL), I would like to thank Alessandro Migliori, Daljeet Gahle, and Kalliopi Kanaki for the guidance, teachings, and opportunities. From the Soil and Water Management and Crop Nutrition Laboratory (SWMCNL), I would like to thank Gerd Dercon and Magdeline Vlasimsky for their support throughout my stay. I am also sincerely thankful for the connections and friendships I built during my time at the IAEA, which made both the work and daily routine far more enjoyable.

Beyond academia, I would like to thank the friends and fellows from the Associação Londrinense de Tiro com Arco (ALTA) for the camaraderie, support, and invaluable moments outside academia that contributed to my well-being during this period. Who would have guessed that shooting carbon rods into EVA targets could be so soothing and satisfying?

This work was also made possible through the support of the Laboratório de Física Nuclear Aplicada (LFNA) and the Universidade Estadual de Londrina (UEL) for providing the infrastructure, academic environment, and institutional support that made this work possible. Financial support from CAPES, CNPq, and the INCT-FNA is also acknowledged, both through scholarships and funding, as well as through access to equipment and research facilities essential to the development of this work. I further thank the International Atomic Energy Agency (IAEA), the Abdus Salam International Centre for Theoretical Physics (ICTP), and the Comissão Nacional de Energia Nuclear (CNEN) for the opportunity and support provided through the Sandwich Training Education Programme (STEP) fellowship.

At last, I would like to thank the reader for taking an interest in this work, as well as the members of the examination board who kindly accepted the invitation to evaluate and contribute to it.

It is remarkable how a single decision — choosing to pursue a master's degree at UEL back in 2020 — led to so many changes in my life. If I could go back five years and tell myself everything I would experience along the way, I would hardly believe it. These past years have been chaotically productive, and I am deeply grateful to everyone who has been a part of this journey.

I may have gotten a little carried away but, without further ado, on with the show!

LOPES, J. M. F.. **Análise Espectroscópica Multimodal Aplicada à Avaliação da Erosão em Solos Agrícolas**. 2026. 146f. Tese (Doutorado em Física) – Universidade Estadual de Londrina, Londrina, 2026.

RESUMO

A erosão do solo em áreas agrícolas é um fenômeno natural caracterizado pela perda da camada fértil, podendo ser agravada por atividades humanas e resultando em impactos sociais e econômicos significativos. Dada a importância da conservação do solo, a adoção de práticas de manejo sustentável e o desenvolvimento de metodologias capazes de avaliar tais práticas de maneira rápida e precisa são essenciais. Nos últimos anos, técnicas de análise espectral têm se consolidado como ferramentas capazes de fornecer informações relevantes sobre a saúde e o manejo do solo, exigindo preparação mínima de amostras e sem a geração de poluentes.

Nesse contexto, este trabalho apresenta dois objetivos principais: (i) desenvolver metodologias baseadas em Espectrometria de Raios Gama (GRS) que permitam o uso de detectores de baixa resolução na avaliação da redistribuição do solo; e (ii) empregar GRS, Fluorescência de Raios X por Dispersão de Energia (EDXRF) e Espectroscopia de Infravermelho Médio (MIRS) para avaliar a composição nuclear, elementar e molecular de solos e, por meio de fusão de dados e técnicas de aprendizado de máquina, aprimorar a compreensão dos processos erosivos e sua relação com a composição do solo.

Para o primeiro objetivo, foram desenvolvidas duas metodologias: uma baseada na deconvolução espectral de medidas realizadas com detectores de NaI(Tl) para a quantificação de ^{7}Be; e outra baseada na transferência de calibração entre detectores de HPGe e NaI(Tl) via *Target Transformation Factor Analysis* (TTFA), visando aumentar a acurácia das análises conduzidas com detectores de baixa resolução.

Para o segundo objetivo, quatro abordagens foram aplicadas: (i) análise de componentes principais (PCA) para avaliar correlações entre os resultados quantitativos de GRS e EDXRF, indicando comportamento semelhante entre a redistribuição do solo e a concentração de elementos oriundos de fertilizantes; (ii) fusão de dados espectrais de GRS e EDXRF utilizando Análise de Dimensões Comuns (ComDim), demonstrando potencial para a determinação da proveniência do solo; (iii) fusão de dados quantitativos de GRS, EDXRF e química analítica via PCA multibloco, revelando correlações entre marcadores de redistribuição, composição elementar e parâmetros de fertilidade; e (iv) fusão de dados MIRS e EDXRF para a construção de modelos preditivos de parâmetros de fertilidade, oferecendo uma alternativa rápida aos métodos convencionais de química analítica.

Esta abordagem integrada demonstra que a sinergia entre diferentes técnicas espectrométricas e análise multivariada amplia significativamente a capacidade de avaliação da dinâmica do solo e da produtividade agrícola.

Palavras-chave: Redistribuição de Solo. Fusão de Dados. Espectrometria de Raios Gama. Fluorescência de Raios X. Espectroscopia do Infravermelho Médio.

LOPES, J. M. F.. **Multimodal Spectroscopic Analysis for Erosion Assessment in Agricultural Soils**. 2026. 146p. PhD Thesis (PhD in Physics)Tese (Doutorado em Física) – State University of Londrina, Londrina, 2026.

ABSTRACT

Soil erosion in agricultural areas is a natural phenomenon characterized by the loss of fertile soil, which can be exacerbated by human activities and result in significant social and economic impacts. Given the importance of soil conservation, the adoption of sustainable management practices and the development of methodologies capable of evaluating such practices in a rapid and accurate manner are essential. In recent years, spectral analysis techniques have become established as tools capable of providing relevant information on soil health and management, requiring minimal sample preparation and generating no pollutants.

In this context, this work presents two main objectives: (i) to develop methodologies based on Gamma-Ray Spectrometry (GRS) that enable the use of low-resolution detectors for the assessment of soil redistribution; and (ii) to employ GRS, Energy-Dispersive X-Ray Fluorescence (EDXRF), and Mid-Infrared Spectroscopy (MIRS) to evaluate the nuclear, elemental, and molecular composition of soils and, through data fusion and machine learning techniques, to improve the understanding of erosive processes and their relationship with soil composition.

For the first objective, two methodologies were developed: one based on the spectral deconvolution of measurements performed with NaI(Tl) detectors for the quantification of ^{7}Be; and another based on calibration transfer between HPGe and NaI(Tl) detectors using Target Transformation Factor Analysis (TTFA), aiming to increase the accuracy of analyses conducted with low-resolution detectors.

For the second objective, four distinct approaches were applied: (i) principal component analysis (PCA) to evaluate correlations between quantitative GRS and EDXRF results, which indicated similar behavior between soil redistribution and the concentration of elements originating from fertilizers; (ii) fusion of GRS and EDXRF spectral data using Common Dimensions Analysis (ComDim) for soil provenance determination; (iii) fusion of quantitative GRS, EDXRF, and wet chemistry data via multiblock PCA to identify correlations among erosion markers and fertility parameters; and (iv) the fusion of MIRS and EDXRF data to build predictive models for soil fertility parameters, providing a faster alternative to traditional wet chemistry methods.

This integrated approach demonstrates that the synergy between different spectrometric techniques and multivariate analysis significantly enhances the assessment of soil dynamics and agricultural productivity.

Keywords: Soil Redistribution. Data Fusion. Multisensor Integration. Gamma-ray Spectrometry. X-Ray Fluorescence. Mid-Infrared Spectroscopy.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

CRM	Certified Reference Material
CT	Calibration Transfer
EC	Electronic Capture
FRN	Fallout Radionuclide
FWHM	Full Width at Half Maximum
GRS	Gamma-Ray Spectrometry
HPGe	High-Purity Germanium
IAEA	International Atomic Energy Agency
IPEN	Instituto de Pesquisas Energéticas e Nucleares
LLD	Lower Limit of Detection
LLQ	Lower Limit of Quantification
MDA	Minimum Detectable Activity
MIRS	Mid-Infrared Spectroscopy
MCA	Multichannel Analyzer
NaI(Tl)	Thallium-Activated Sodium Iodide
NORM	Naturally Occurring Radioactive Material
PCA	Principal Component Analysis
PLSR	Partial Least Squares Regression
PMT	Photomultiplier Tube
RF	Random Forest
ROI	Region of Interest
SDD	Silicon-Drift Detector
SVD	Singular Value Decomposition
TTFA	Target Transformation Factor Analysis
XRF	X-Ray Fluorescence

CONTENTS

1 INTRODUCTION AND OBJECTIVES

Erosion is a naturally occurring degradation agent that involves three types of processes: detachment, movement, and deposition of soil particles [1, 2]. Albeit natural, erosive processes might be accelerated by human activities that leads to negative impacts on agricultural production [3, 4]. The dislodgment of fertile topsoil in agricultural areas is associated with irretrievable loss of nutrients that plants require to grow, leading to the decrease of quality and quantity of food production.

Different types of land degradation affects over 1.5 billion people, specially in developing countries [5]. Unsustainable human activities, such as overgrazing and deforestation, have accelerated soil erosion for over a thousand times [6]. Soil is the source of 95 percent of all agricultural production [6] and approximately a quarter of the world's population relies on food produced at already degraded lands and this rate increases every year [5]. As of now, soil is degrading faster than it can be replenished. Overall, soil erosion is more common on steep, slopping land [5].

This way, soil erosion is considered to be one of the major threats to soil stability and productivity. Despite the seriousness of this problem, erosion monitoring remains a challenge [7]. For the past six decades, fallout radionuclides (FRN) have been studied as tracers to evaluate soil erosion in different times frames and different types of soil [4, 7–12]. FRN are radionuclides that are found in the Earth's atmosphere for a given period and eventually are deposited on the ground, where they are adsorbed by clay particles thats constitutes soil.

Gamma-Ray Spectrometry (GRS) is an analytical non-destructive technique that may be employed to detect and quantify radionuclides present in the environment and, therefore, evaluate soil redistribution based on the FRN concentrations. In the context of the cosmogenic radioactive isotope that is formed in the stratosphere and troposphere by the spallation of Nitrogen and Oxygen [4]; and Caesium-137, a anthropogenic radionuclide released in the atmosphere due to nuclear explosions [1]. The main objectives of this work were:

- to optimize the soil erosion estimation procedure using GRS coupled with multivariate analysis;
- to develop a methodology that allows for soil erosion estimation using ^{7}Be as a tracer employing Thallium-Activated Sodium Iodide (NaI(Tl)) detectors; and
- to evaluate the short-term soil redistribution for three agricultural areas.

Additionally, the GRS results were combined with data from X-ray Fluorescence (XRF) and Mid-Infrared Spectroscopy (MIRS) to:

- investigate correlations between radioactive, elemental, and molecular properties in eroded soils;
- employ multivariate and machine learning methods to predict soil fertility parameters; and
- assess the complementarity of these spectrometric techniques in evaluating soil redistribution processes.

The results of the present study are organized in two major branches: (1) **Methodological developments in Gamma-Ray Spectrometry** (Chapter 6), addressing detector calibration, spectral deconvolution, and data transfer between detection systems; and (2) **Multitechnique approaches for soil redistribution studies** (Chapter 7), where GRS, XRF, and MIRS data are jointly analyzed to enhance the understanding of soil redistribution and its relationship with soil composition. Each Section within the respective Chapters describes a self-contained work, with its methodologies, results, and conclusions discussed independently.

By integrating nuclear and X-ray spectrometric techniques with modern multivariate tools, this work contributes to advancing soil erosion assessment, offering both methodological innovations and an interdisciplinary perspective on soil redistribution dynamics.

2 PRINCIPLES OF RADIATION AND DETECTION

2.1 Radiation Physics

Radioactivity is a nuclear phenomenon that arises when a atomic species has an unstable nucleus. This instability may be caused by an uneven distribution between the protons and neutrons that compose the nucleus, or by interactions with incident particles or sufficiently energetic waves [13]. For lighter elements, the amount of protons and neutrons necessary to achieve stability is somewhat similar, while for heavier elements these quantities mays differ greatly. This behavior leads to an abundance of radioactive isotopes of heavy elements in comparison with lighter ones.

Unstable nuclei must undergo radioactive decay to achieve stability, which might happen as the emission of particles (α, β) or electromagnetic waves (γ, X-rays). Corpuscular emissions leads to nuclear transmutation since the amount of protons within the nuclear species will change and, for energy to be conserved, the final state's energy has to be lower than the initial [13].

The alpha decay (α) consists in the emission of 2 protons and 2 neutrons (a He nucleus) from the nucleus. The α emission reaction is described in Equation (2.1). This type of decay is particularly common in nuclei with $A \geq 200$, where the electrostatic repulsion between the protons is stronger [14].

$$ {}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} \rightarrow {}^{\mathrm{A}-4}_{\mathrm{Z}-2}\mathrm{Y} + \alpha \tag{2.1} $$

The beta decay (β) might be positive or negative. In the β^- decay a neutron is converted into a proton, an electron and an antineutrino. The electron and the antineutrino are then emitted from the nucleus. Analogously, the β^+ decay consists in the conversion of a proton into a neutron, a positron and a neutrino, where the last two are emitted. Both processes are shown in Equation 2.2.

$$ \begin{aligned} {}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} &\rightarrow {}_{\mathrm{Z}+1}{}^{\mathrm{A}}\mathrm{Y} + \beta^- + \bar{\nu} \\ {}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} &\rightarrow {}_{\mathrm{Z}-1}{}^{\mathrm{A}}\mathrm{Y} + \beta^+ + \nu \end{aligned} \tag{2.2} $$

Unlike α or γ emission, the β decay has a continuos energy spectrum. Alternatively to the β^+ decay, another mechanism that might happen inside a unstable nucleus is the electron capture (EC). In this process, a nuclear proton "captures" an electron from the K electronic shell and is then converted into a neutron and a neutrino. The EC process is shown in Equation 2.3.

$$ {}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} + e^{-} \rightarrow {}_{\mathrm{Z}-1}{}^{\mathrm{A}}\mathrm{Y} + \nu \tag{2.3} $$

The γ emission frequently happens after one of the corpuscular decays, since the daughter nucleus is often found in an excited state and needs to release the energy excess before reaching stability. This energy is released as an electromagnetic wave know as gamma radiation. The process is described in Equation (2.4).

$$ {}^{\mathrm{A}}\mathrm{X}^{*} \rightarrow {}^{\mathrm{A}}\mathrm{X} + \gamma \tag{2.4} $$

The * indicates that the nucleus is in an excited state. And, unlike α, β or EC, the γ emission is not associated with nuclear transmutation. Each γ transition is characterized by a well-defined energy, commonly ranging from 10 keV to several MeV. The half-lives of the excited states are usually very short, ranging from 10^{-12} to 10^{-9} seconds [14]. A few exceptions might occur, and some isotopes may have unusually long half-lives for their excited states, ranging from milliseconds to years, in extraordinary cases. These cases are called metastable states, and a few states, alongside their half-lives, are shown in Table 1.

Table 1 – A few excited metastable states and their respective half-lives.

Metastable State	Half-Life
^{24m}Na	20.20 ms
^{60m}Co	10.47 min
^{99m}Tc	60.2 h
^{137m}Ba	2.52 min
^{242m}Am	141 y

Source: [14], adapted.

Analogously to EC, electrons from the K shell might also aid in the energy release from excited nuclei. Sometimes the dissipated energy from the nucleus might be transferred to an atomic electron, which leads to the phenomenon of internal conversion (IC). Internal conversion is a radioactive decay process where an unstable, excited nucleus transfers its excess energy directly to an orbital electron. This causes the electron to be ejected from the atom rather than emitting a gamma-ray photon. Nonetheless, IC is followed by the emission of X-Rays. Since the electron that absorbs the released energy leaves a vacancy in the atom's electronic shell that shall be filled by another atomic electron.

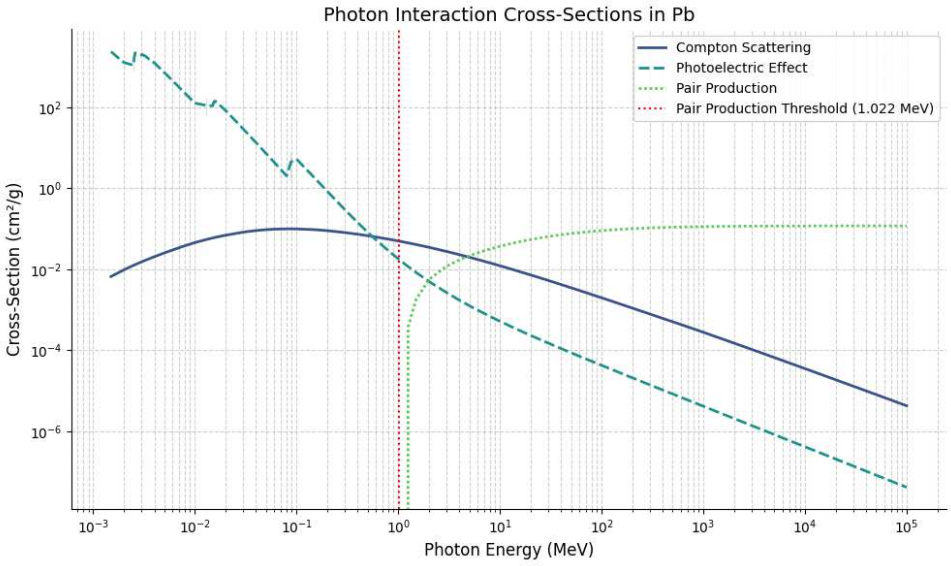
Despite being an atomic process, X-Rays are also a type of radiation and its discovery is entangled with the discovery of radioactive decays. In fact, detected X-Rays and Gamma-Rays in the same energy range are indistinguishable from one another, the only differences are their origin. X-Rays originate from the deexcitation of atomic electrons,

oftenly caused by the necessity of filling a vacancy in inner atomic shells. Besides IC, the photoelectric effect (Section 2.2.1) also causes atoms to emit X-Rays. This type of radiation has a narrower range of energy when compared with gamma radiation, ranging from a few eV to hundreds of keV [15].

2.2 Radiation Interaction with Matter

Ionizing radiation interacts with matter in different ways. The relative probability of each interaction occurring depends on the energy of the incident photon and the atomic number of the material with whom radiation is interacting [14]. Figure 1 illustrates how the probability of each interaction changes as the energy increases when interacting with lead. The Photoelectric Effect is dominant up to approximately 500 keV ($E < 50$ keV in light elements). Compton Scattering is the dominant effect in the region between 100 keV and 10 MeV. At last, Pair Production is the dominant interaction for energies above 5 MeV. It is worth noting that Pair Production has a minimum energy of 1.022 MeV to happen. More details in how each interaction happens will be discussed in the following Subsections. Other types of interactions, such as Rayleigh scattering, nuclear reactions, and photonuclear interactions, may also occur but are generally less significant in the energy range relevant to this work and the employed detectors [14].

Figure 1 – Photon interaction cross-sections for Pb as a function of the energy of the photon. Both axes are logarithmic.

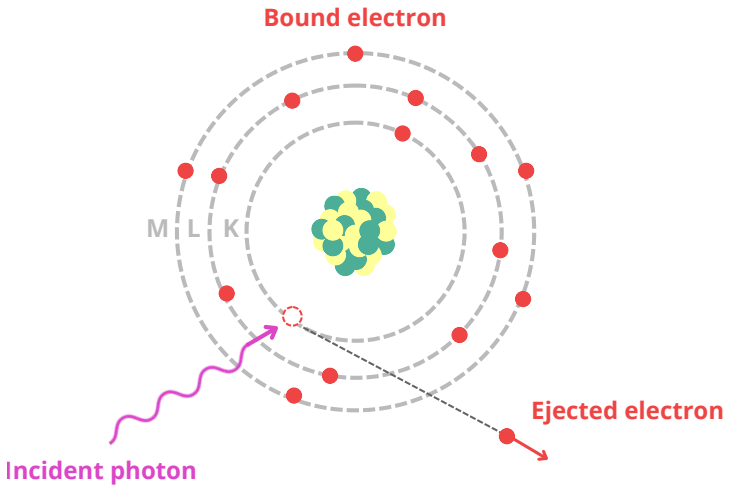


Source: self-elaboration with data sourced from the NIST XCOM database [16].

2.2.1 Photoelectric Effect

The photoelectric effect is a low-energy phenomenon that consists in the absorption of an incident photon by an atomic electron, followed by the ejection of this electron from the atomic shell [17], as illustrated in Figure 2. The interaction can not occur in free electrons since a third body is necessary to absorb the recoil momentum from the collision [14]. This effect might occur through the interaction of both ionizing and non-ionizing radiation with matter, the only requirement is that the incident photon has an energy greater than the electron's binding energy.

Figure 2 – Representation of the Photoelectric Effect using Bohr's atomic model for a Na atom. Electrons in the K shell have the greatest probability of absorbing an incident photon.



Source: self-elaboration.

The photoelectric effect was first observed by Heinrich Rudolf Hertz in 1887, who reported that illuminating two electrodes with a potential difference using ultraviolet light resulted in a change in the measured voltage. In 1902, this effect was explained in further details by Philipp Lenard, who reported that the cause of such changes is the emission of electrons.

The kinetic energy of the ejected electron (K) is defined as the difference between the kinetic energy of the incident photon ($E = h\nu$) and the electron's binding energy (E_{bind}), as shown in Equation (2.5).

$$K = h\nu - E_{\text{bind}} \tag{2.5}$$

Electrons from different atomic shells may absorb the incident photon and be ejected. The probability of photoelectric absorption depends on the electron binding energy and orbital, with inner-shell electrons (particularly from the K and L shells) being

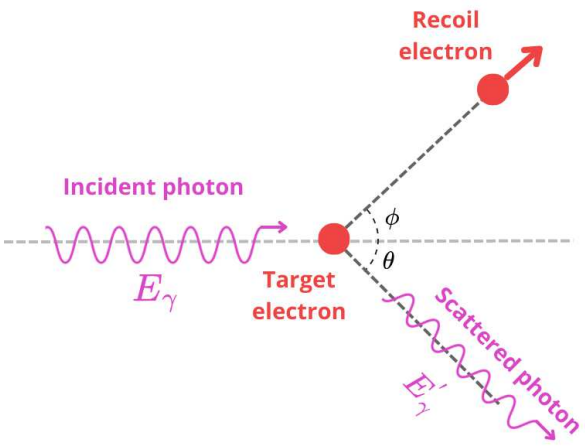
much more likely to interact. This occurs because these electrons are more strongly bound and have a greater overlap with the atomic nucleus. After the ejection of an inner-shell electron, the resulting vacancy is filled by an electron from a higher shell, leading to the emission of characteristic X-rays or Auger electrons.

2.2.2 Compton Effect

The classical electromagnetic interpretation dictates that electromagnetic radiation can be elastically dispersed by atomic electrons [14]. When electromagnetic radiation interacts with atomic electrons, these will emit a secondary radiation with the same energy as the incident radiation. This phenomenon is known as the Thomson effect, and it is dominant for low-energy photons ($E_\gamma < 100$ keV) since different effects will happen when the radiation energy is similar of the electrons rest mass ($m_0c^2 = 511$ keV$/c^2$) [14].

The Compton effect consists of the inelastic scattering of photons by atomic electrons that are weakly bounded [18], and is a mid-energy phenomenon. The effect was first observed by Arthur Holly Compton in 1923 while researching the scattering of X-rays. An incident photon hits a weakly bound electron ($E_{\text{photon}} > E_{\text{bind}}$), scattering the electron and changing its original direction, as shown in Figure 3.

Figure 3 – Representation of the Compton Effect.



Source: self-elaboration.

During the collision, the photon loses part of its energy, which is transferred to the electron. The energies of the two bodies after the collision can be calculated as a function of the photon's initial energy (E_γ) and its scattering angle (θ). These functions are displayed in Equation (2.6), where $\alpha = E_\gamma/m_0c^2$.

$$E'_\gamma = \frac{E_\gamma}{1+\alpha(1-\cos\theta)}$$
$$E_e = \frac{E_\gamma^2(1-\cos\theta)}{m_0c^2+E_\gamma(1-\cos\theta)} \tag{2.6}$$

E'_γ and E_e takes continuous energy values according to the scattering angle, which ranges between 0 and π. When the energy of the incident photon is sufficiently high, its dependence on E'_γ becomes negligible. In that case, the energy of the scattered photon can be rewritten as Equation (2.7).

$$E'_\gamma \to \frac{m_0c^2}{1-\cos\theta} \tag{2.7}$$

This result is particularly interesting for $\pi/2 \leq \theta \leq \pi$, since this will lead to Equation (2.8). This result is useful for interpreting the response of gamma radiation detectors [19].

$$\frac{m_0c^2}{2} \leq E'_\gamma \leq m_0c^2 \tag{2.8}$$

To fully understand this result, it is important to look at the Klein-Nishina equation for the Compton cross-section (Equation (2.9)) that was first derived by Klein-Nishina in 1929 [20,21]. The equation describes the probability of a photon with energy of E_γ being scattered through an angle between θ and $\theta + \mathrm{d}\theta$.

$$\frac{\mathrm{d}\sigma(E_\gamma,\theta)}{\mathrm{d}\Omega} = \frac{r_0^2}{2}\left\{\frac{1+\cos^2\theta}{[1+\alpha(1-\cos\theta)]^2} + \frac{\alpha^2(1-\cos\theta)^2}{[1+\alpha(1-\cos\theta)]^3}\right\} \tag{2.9}$$

At low energies, $\alpha \to 0$, the distribution corresponds to that expected for the Thomson effect [14]. For low energies, the probability of forward ($0 \leq \theta \leq \pi/2$) and backwards ($\pi/2 \leq \theta \leq \pi$) scattering are similar. As the energy of the incident photon increases, the probability of backscatter diminishes. By integrating Equation (2.9) for all possible angles, it is possible to obtain the Compton scattering excitation function (Equation (2.10)).

$$\sigma_c = 2\pi r_0^2\left\{\frac{1+\alpha}{\alpha^2}\left[\frac{2(1+\alpha)}{1+2\alpha} - \frac{1}{\alpha}\ln(1+2\alpha)\right] + \frac{1}{2\alpha}\ln(1+2\alpha) - \frac{1+3\alpha}{(1+2\alpha)^2}\right\} \tag{2.10}$$

Equation (2.10) shows that the scattering cross-section decreases as the energy increases. The maximum energy transfer from the photon to the electron happens when $\theta = \pi$. This condition defines the upper limit of the continuous region in the gamma-ray spectrum known as the Compton continuum, and the endpoint is referred to as the

Compton edge. The energy at this edge can be derived from Equation (2.6) by setting the backscatter condition, leading to:

$$E_{e,\max} = E_\gamma\left(1 - \frac{1}{1+2\alpha}\right) \tag{2.11}$$

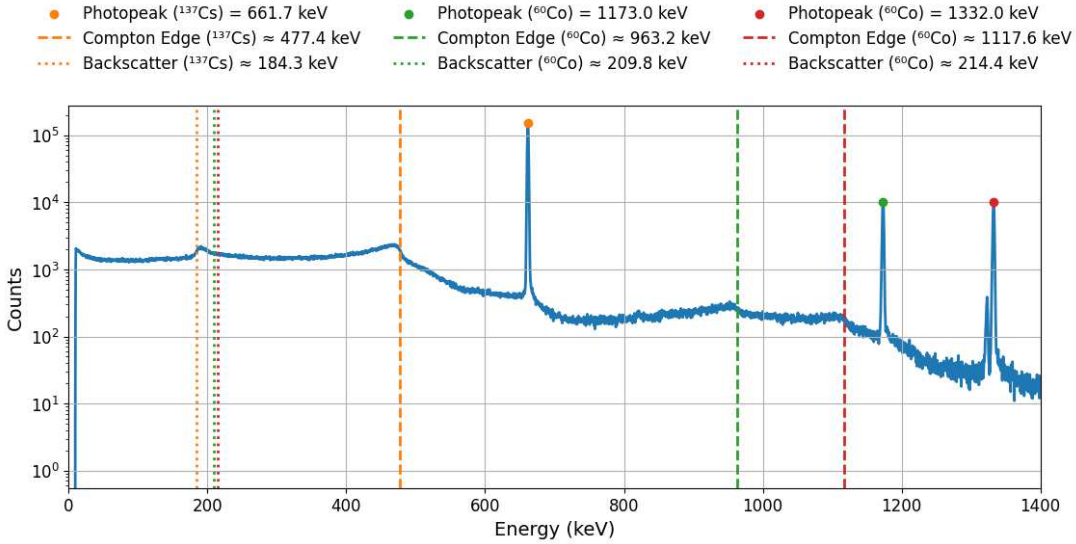
Despite this being the maximum possible energy transferred to the electron, the Klein-Nishina distribution indicates that back scattering events are statistically less probable as α increases. Therefore, although the Compton edge marks a clear energy limit in the spectrum, most scattered photons will interact at smaller angles, leading to the characteristic declining shape of the Compton continuum.

Additionally, photons that are scattered externally and reenter the detector can contribute to a backscatter peak. These photons have lost energy due to scattering at angles near $\theta = \pi$ before being detected, and their energy is given by:

$$E'_\gamma = \frac{E_\gamma}{1+2\alpha} \tag{2.12}$$

This feature appears as a distinct, lower-energy peak in the gamma spectrum, particularly in detectors surrounded by high-Z materials. Figure 4 illustrates the gamma ray spectrum of a ^{137}Cs ($E_\gamma = 661.7$ keV) and a ^{60}Co sources ($E_\gamma = 1173$ and $E_\gamma = 1332$ keV), obtained with a HPGe detector, indicating the position of the Compton edges (Equation (2.11)) and backscatter peaks (Equation (2.12)).

Figure 4 – Spectrum of ^{137}Cs and ^{60}Co sources indicating their photopeaks (dots), Compton edges (dashed lines), and backscatter peaks (dotted lines).



Source: self-elaboration.

The Compton edge appears in the spectrum as a sharp drop in the *Compton continuum*, which is the region between 0 and the edge, near the photopeak being scattered.

The Compton continuum is composed of the energies deposited by photons who undergo Compton scattering at small- and mid-angles, since the Klein-Nishina equation favors forward scattering. The backscatter peak, on the other hand, always appears with small energies, leading to an accumulation of counts in this region of the spectrum.

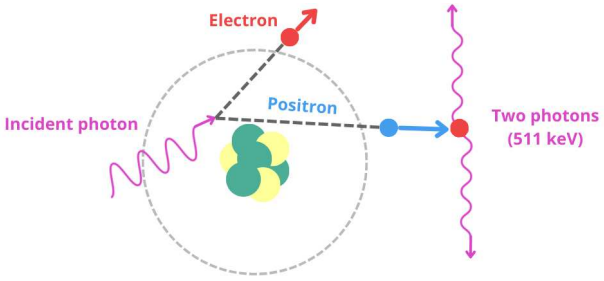
For compounds or materials emitting multiple gamma lines, the resulting spectrum shows overlapping Compton continua and edges, which can make feature identification and energy quantification more challenging. This is particularly relevant when interpreting environmental or mixed-source spectra, where multiple gamma emissions contribute to a complex continuum.

In Figure 4 a small peak can be seen in one of ^{60}Co emissions. This happens due to summed ^{137}Cs emissions, cause by the source's high activity. More on this effect will be discussed in Section 2.3.

2.2.3 Pair Production

While both photoelectric and Compton effects describes the interaction between photos and electrons, the effect of pair production consists in the interaction between a photon and the atomic nucleus [19]. When a photon with sufficiently high energy passes through the Coulomb field of an atom, this photon is converted into an electron-positron pair. The processes in illustrated in Figure 5.

Figure 5 – Representation of the pair production happening in the Coulomb field of an atom. An incident photon ($E \geq 1022$ keV) is converted into an electron-positron pair, and the positron undergo annihilation with an electron producing two photons with $E = 511$ keV.



Source: self-elaboration.

For the conversion to take place, it is necessary that the energy of the incident photon is equal or greater than twice the rest mass of an electron, 511 keV, making 1022 keV the threshold for this phenomenon to happen. The Coulomb field of an electron can also cause a photon to experience pair production, but the threshold for this interaction is equal to four times the rest mass of an electron, which deems the probability of this event happening negligible for common gamma ray spectrometry (0 to 3 MeV) [14, 19].

After the conversion, the positron forms a positronium with a free electron from the medium [20]. When the positronium collapses, two photons with energies $E \approx 511$ keV are produced and collinearly emitted in exactly opposed directions, to conserve momentum [19, 20]. The peak produced in a gamma ray spectrum by these photons is characteristically broader than regular decay-produced peaks due to *Doppler broadening*. In the moment of the annihilation, the resulting momentum from the positronium's orbit net momentum causes the energy of the produced photons to oscillate around 511 keV, increasing statistical uncertainty and widening the peak [19].

2.3 Gamma and X-Radiation Detectors

Despite the different types of detectors developed in the past decade and available nowadays, they all function under the same principle: an incident photon interacts with the detector mass, and its kinetic energy is partially or totally converted into a signal [17]. Before dabbing into the specificities of the detector types relevant to this work, it is important to discuss the general properties of radiation detectors.

A radiation detector must be able to produce usable signals for a given type of radiation and energy, therefore detectors are built to be sensitive to certain types of radiation in a given energy range [17]. The *sensitivity* of a detector is dependent of its mass and cross-section, since these will determine the probability of an incident photon producing ionization. Since the cross-section is proportional to the atomic number of the medium, overall detectors are built with elements or compounds with high atomic number [19].

Beyond detecting the presence of radiation, a detector must produce a *response* that provides information on the incident photon's energy [17]. Moreover, each type of detector has its *energy resolution*, which is the measurement of how well a detector can distinguish emission with close energy values. Ideally, every photon would produce a response similar to a delta function but, in reality, a Gaussian-shaped peak with finite width is produced [17]. A detector's resolution for a given energy is measured in terms of the peak's full width at half maximum (FWHM), as shown in Equation (2.13), and is measured as a percentage.

$$\text{Resolution} = \frac{\text{FWHM}}{\text{Energy}} \tag{2.13}$$

During a measurement, not all photons emitted by the source will interact with the detector's sensitive volume. Furthermore, the fraction of emitted photons that are detected depends on the energy of the emitted radiation [17, 19]. The ratio between the number of detected and emitted photons defines the *efficiency* (ε) of the detector for a

given energy, as shown in Equation (2.14). For detectors used to measure a broad range of photon energies, efficiency values are calculated at multiple energies to construct an *efficiency curve* encompassing the entire energy range.

$$\varepsilon = \frac{\text{detected photons}}{\text{emitted photons}} \tag{2.14}$$

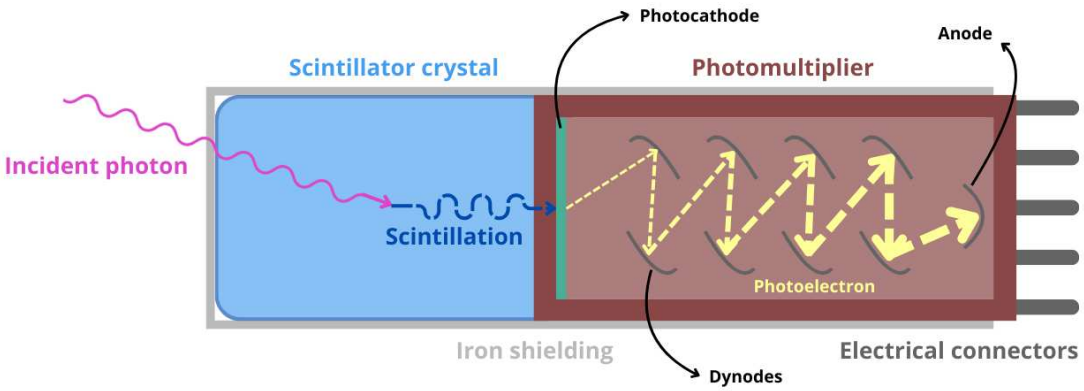
During a radiation measurement, the detection system requires a finite time to process each event, a period known as *dead time*, during which the detector is unable to process incident photons. At high count rates, dead time losses become more significant and can distort the observed spectrum. On the other hand, the *live time* corresponds to the interval in which the system is active and capable of recording data [17].

When two or more photons are detected at the same time, these incidences might be mistakenly interpreted as a single photon with greater energy. This is known as *pile-up* and might lead to the appearance of extra peaks in a spectrum. This can create false peaks in the spectrum at energies corresponding to the sum of two or more gamma lines, which can complicate the analysis and interpretation of the data [19].

2.3.1 Thallium-Activated Sodium Iodide Detectors

Sodium Iodide (NaI) detectors belong to a class of material called scintillation detectors, which receives this name due to its behavior when being excited. After radiation interacts with the detection medium by temporarily exciting and electron that de-excites by emitting electromagnetic radiation in, or near, optical wavelengths [19]. The schematics of a scintillation detector are displayed in Figure 6.

Figure 6 – Schematics of a scintillation detector. The incident photon produces scintillation inside the crystal, that strikes the photocathode, emitting photoelectrons whose intensities are increased by the dynodes and later converted into electronic signals.



Source: self-elaboration.

The produced wavelengths are absorbed by a photocathode and gives rise to photoelectrons, which are augmented by a photomultiplier tube (PMT). The intensity-enhanced photoelectrons are then converted into a weak electric pulse proportional to the incident radiation energy [17].

Normally, the de-excitation of secondary electrons in the NaI crystal are below the visible range. Moreover, the bulk of the crystal may absorb the produced photons before they reach the photomultiplier. The solution to this problem lies in the introduction of an *activator* when manufacturing the crystal. The addition of 10^{-3} mol fraction of Thallium changes the energy levels of the crystal, giving rise to a NaI(Tl) detector. In this new configuration, the de-excited electrons emit photons with lower energies — and, therefore, smaller wavelengths — than in the pure crystal. This emission will no longer match the absorption characteristics of the detector and more light will reach the photomultiplier. Not all the energy absorbed from the incident photon is re-emited as scintillation; NaI(Tl) detectors convert only 12% of the total energy [19].

The processes inside the PMT are as follows: (*i*) the scintillation photon strikes the *photocathode*, producing photoelectrons; (*ii*) the photoelectrons are focused onto a series of dynodes, structures that emit more electrons than they receive, increasing the input signal at each stage; and (*iii*) the amplified signal is collected at the anode and converted into electronic signals [19]. Most PMT's contains 10 to 14 dynode stages, leading to gains in intensity up to 10^7 [17].

One of the main limitations of NaI(Tl) detectors is their relatively poor energy resolution. This is primarily due to the multiple statistical processes involved in signal formation: from the production of scintillation photons, to their collection and conversion into photoelectrons in the PMT. The small number of photoelectrons generated per unit of deposited energy leads to significant pulse height fluctuations for monoenergetic radiation, resulting in broader peaks in the measured spectra [17, 19]. Although various scintillating materials are available, NaI(Tl) remains one of the most widely used in gamma-ray spectrometry due to its high detection efficiency, relatively low cost, and ease of operation. Its performance characteristics, however, make it more suitable for applications where high energy resolution is not essential.

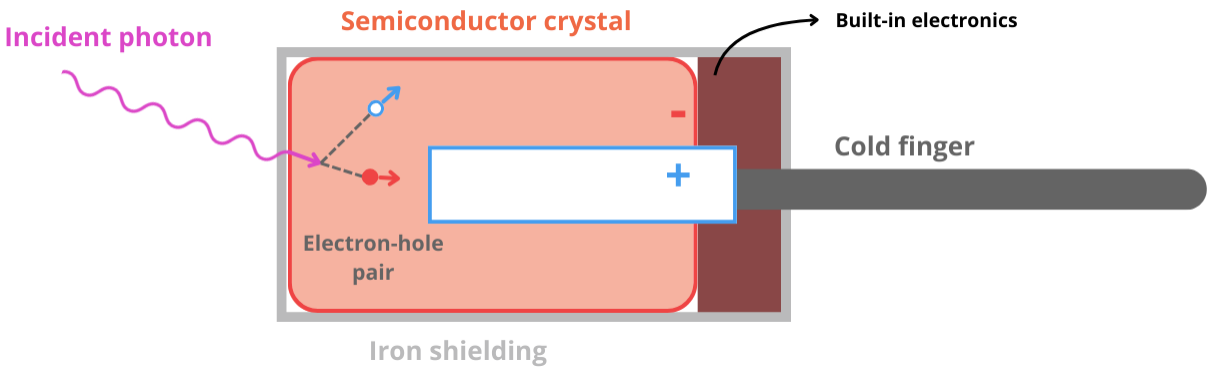
2.3.2 High-Purity Germanium Detectors

High-Purity Germanium (HPGe) detectors are solid-state detectors widely used in gamma-ray spectrometry due to their excellent energy resolution. Unlike scintillation detectors, HPGe detectors produce an electrical signal directly from the ionization of the detection medium. When an incident photon interacts with the germanium crystal, it transfers energy to the lattice, generating electron-hole pairs [17].

The germanium crystal is biased with a strong electric field. After interaction, electrons drift toward the positive contact, while holes drift toward the negative. The movement of charge carriers induces a current pulse proportional to the energy deposited. Because germanium has a narrow band gap (0.67 eV), thermal excitation can create noise at room temperature. For this reason, the crystal must be cooled to 77 K using a cold finger connected to a liquid nitrogen dewar [19].

Modern HPGe detectors also include a preamplifier close to the crystal to minimize signal loss and electronic noise. Figure 7 shows the schematic operation of the system. The incoming photon generates an electron-hole pair in the crystal, which is collected under the influence of the electric field. The signal is then amplified and processed.

Figure 7 – Schematics of an HPGe detector. The incident photon interacts with the semiconductor crystal, generating an electron-hole pair. The charges are separated by the electric field and collected at the contacts.



Source: self-elaboration.

The main advantage of HPGe detectors is their superior energy resolution, typically around 0.2% at 1.33 MeV [19]. This makes them suitable for applications requiring precise identification of gamma emitters, especially in complex spectra. However, they are more expensive, require continuous cooling, and have lower efficiency for high-energy photons compared to scintillators of similar volume.

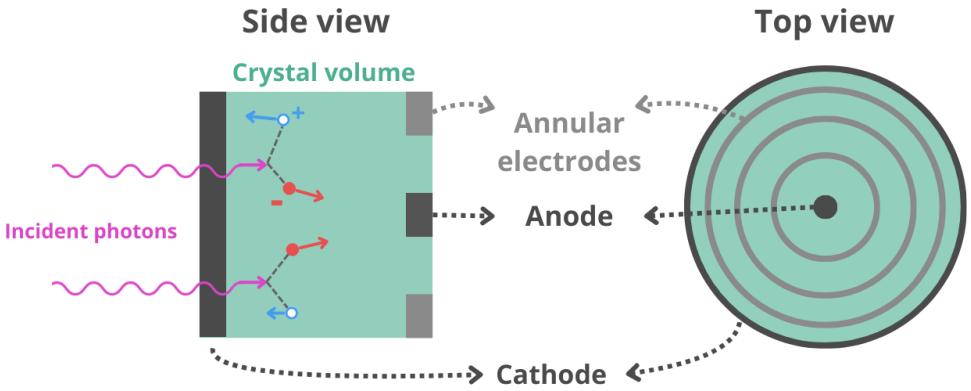
2.3.3 Silicon-Drift Detectors

Silicon Drift Detectors (SDDs) are semiconductor devices used for detecting low- to medium-energy photons, especially in X-ray fluorescence applications. Their design allows for fast signal processing, low electronic noise, and excellent energy resolution, making them well-suited for energy-dispersive X-ray fluorescence (EDXRF) measurements.

When an X-ray photon enters the sensitive volume of the silicon crystal, it interacts with the material and generates an electron-hole pair. The number of pairs is proportional to the energy deposited [17]. Unlike planar detectors, SDDs use a circular

drift-ring structure that guides electrons to a small, central anode. This geometry reduces the detector's capacitance and allows for fast signal rise times and low noise. The schematics of a SDD detector are displayed in Figure 8, indicating its top and side view and how the electron-hole pairs formed by incident radiation are moved inside the crystal.

Figure 8 – Schematic illustration of the operating principle of a SDD. In the side view, incident photons interact with the silicon crystal, producing electron-hole pairs whose drift directions are determined by the applied electric field. The top view shows the concentric annular electrodes responsible for shaping the electric field and directing electrons toward the central anode, while holes drift toward the cathode.



Source: self-elaboration.

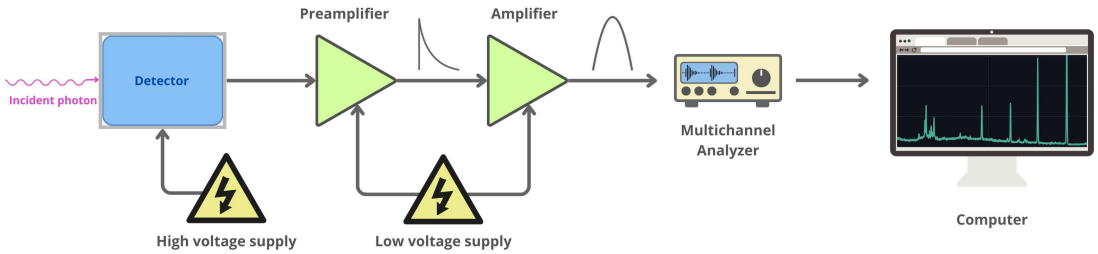
SDDs are equipped with a thermoelectric (Peltier) cooler to maintain a low operating temperature, minimizing thermal noise without the need for cryogenic cooling. The low capacitance, combined with integrated field-effect transistors (FETs), allows for excellent resolution at high count rates, often below 150 eV FWHM at 5.9 keV, the characteristic X-ray of Mn [19].

Due to their resolution and speed, SDDs are commonly used in portable or bench top EDXRF systems. While not suitable for high-energy gamma-ray detection, their performance in the X-ray range makes them ideal for qualitative and quantitative elemental analysis.

2.4 Electronics of Radiation Detection

The signal generated by radiation detectors is not immediately suitable for analysis. It must first undergo a series of steps to extract meaningful information about the radiation source. The output from a detector is a small amount of electrical charge proportional to the energy of the incident photon [19]. The typical signal processing path is shown in Figure 9.

Figure 9 – Diagram illustrating the steps for signal processing after a photon is detected. The preamplifier is placed as close to the detector as possible, while other components are typically connected via coaxial cables.



Source: self-elaboration.

The electronics used in a spectrometry system are designed to transfer and process the signal with minimal distortion. However, some amount of electrical noise is unavoidable, leading to fluctuations in the shape and amplitude of the signal. When these pulses are collected to form a spectrum, such fluctuations contribute to peak broadening.

A high-voltage power supply is connected to the detector to create an electric field strong enough to drive the generated charges toward the electrodes. The preamplifier, positioned as close as possible to the detector, amplifies the signal before it becomes degraded by external noise. In NaI(Tl) detectors, additional fluctuations can arise in the PMT, before the preamplifier.

After the initial amplification, the signal goes through the main amplifier. This component not only enhances the pulse intensity but also reshapes the signal — typically into a semi-Gaussian form — to suppress long tails, reducing the likelihood of pile-up, and optimize the signal-to-noise ratio.

The final stage is the Multichannel Analyzer (MCA), which digitizes the shaped pulses. The MCA measures the height of each pulse, bins them by amplitude (i.e., energy), and counts the number of events within each bin. Since the pulse height is proportional to the energy of the incident photon, the resulting histogram — which is the spectrum itself — reflects both the energy and intensity distribution of the detected radiation.

2.5 Gamma-Ray Spectrometry

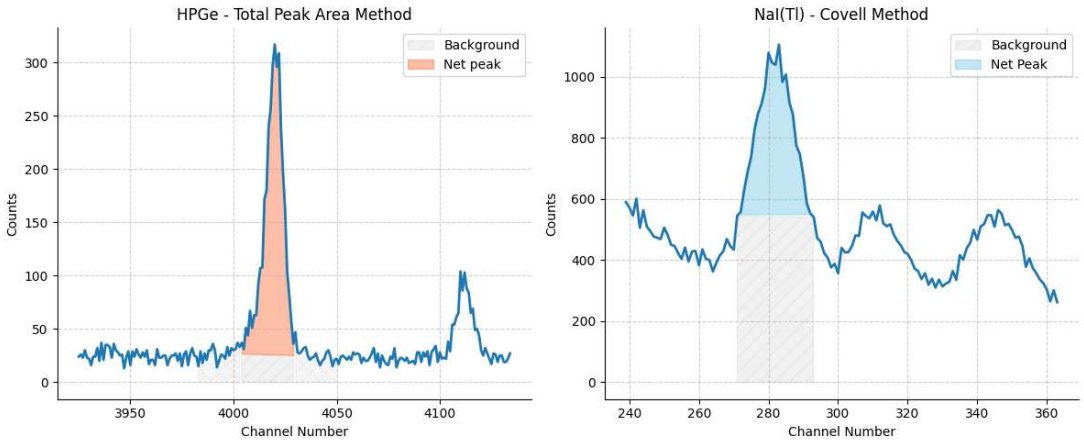
Gamma-ray spectrometry (GRS) is a non-destructive technique that allows the simultaneous detection and quantification of multiple radionuclides in a sample, with minimal sample preparation and no production of waste materials. Various sample carriers with different geometries and sizes may be used, as long as good detection efficiency is maintained for the photon energies of interest. The choice of detector — such as HPGe or NaI(Tl) — depends on the radionuclides under investigation and the required energy

resolution.

Before routine measurements, the detector must be calibrated in energy to correlate the MCA channels with the photon energy. This is done by placing standard sources with known emission energies near the detector and calculating a function — usually linear — that relates the centroid channel number to the corresponding gamma energy. At least two known emissions are required for this procedure, but three or more are preferable, ideally spanning the detector's sensitive range.

Radionuclide activity concentrations are determined from the *net area* under the photopeaks — defined as the total counts in the region of interest (ROI) minus the estimated background. The *Covell method* is typically applied to low-resolution detectors, while the *total peak area method* is more suitable for high-resolution detectors [19]. Both are illustrated in Figure 10.

Figure 10 – Illustration of the net area calculation for spectra obtained with HPGe and NaI(Tl) detectors.



Source: self-elaboration.

Unlike the total peak area method, the Covell method excludes the area beneath the baseline, reducing uncertainty in cases of poor resolution. In both approaches, the background is estimated using channels adjacent to the ROI, defined by a lower limit (L) and an upper limit (U). The gross area (G) is computed by summing the counts in all channels c_i between L and U. The background (B) is approximated by averaging counts in m channels on both sides of the ROI. The *net area* is then calculated as:

$$\begin{aligned} \text{Net Area} &= G - B \\ &= \sum_{i=L}^{U} c_i - \frac{n}{2m}\left(\sum_{i=L-m}^{L-1} c_i + \sum_{i=U+1}^{U+m} c_i\right) \end{aligned} \tag{2.15}$$

Here, n is the number of channels in the ROI, and m is typically between 2 and 5.

The variance of the net peak area is given by the sum of variances from both terms, with the uncertainty of the net area obtained by taking the square root:

$$\begin{aligned} \text{var(Net Area)} &= \sum_{i=L}^{U} C_i + \frac{n^2}{4m^2}\left(\sum_{i=L-m}^{L-1} C_i + \sum_{i=U+1}^{U+m} C_i\right) \\ \sigma_{\text{area}} &= \sqrt{\text{net area} + B\left(1+\frac{n}{2m}\right)} \end{aligned} \tag{2.16}$$

The lower limit of detection (LLD) is defined by Equation (2.17), and is used to calculate the *minimum detectable activity* (MDA):

$$LLD = 2.76 + 4.66\sigma_{\text{area}} \tag{2.17}$$

The activity concentration of a radionuclide, in $\text{Bq}\,\text{kg}^{-1}$, is then obtained from the net area using:

$$A = \frac{\text{net area}}{\varepsilon \times m \times t \times p_\gamma} \tag{2.18}$$

In Equation (2.18), m is the sample mass (kg), t is the live measurement time (s), p_γ is the emission probability of the gamma ray, and ε is the detector efficiency for that energy. The efficiency can be determined from spectra of reference materials with known activity. The minimum detectable activity (MDA) is calculated by substituting the net area with the LLD [22].

For the development of the present work, two GRS detectors were employed, and all measurements were performed inside lead shielding. The details of each measurement system are described in the following subsections. Both measurements systems are located at the Laboratório de Física Nuclear Aplicada (LFNA) from the State University of Londrina (UEL). In both cases, the energy calibration was accomplished using IPEN certified calibration sources. Namely, a ^{137}Cs and a ^{60}Co sources were employed.

2.5.1 Canberra GC6020

The first detector was a HPGe detector manufactured by Canberra (model GC6020), featuring a relative efficiency of 60.0% and an energy resolution of 0.3% at 661.7 keV (Figure 11a).

The detector was housed inside a Canberra 747 lead shield (Figure 11b) in order to reduce the contribution of ambient background radiation. Signal acquisition was performed using an Inspector 2000 multichannel analyzer, which provided spectral processing and data storage during the measurements.

Figure 11 – Photographs of the HPGe detector and respective lead shielding used in the experimental measurements.

(a) Canberra HPGe GC6020 detector.

(b) Canberra 747 lead shield.

Source: self-elaboration.

As shown in Figure 11a, the detector is connected to a liquid nitrogen reservoir through a cold finger, which is responsible for maintaining the germanium crystal at cryogenic temperatures during operation. This cooling is essential to suppress thermal noise and ensure optimal energy resolution throughout the measurement period.

2.5.2 Ortec 905–4

The second detector was a NaI(Tl) scintillation detector manufactured by Ortec (model 905–4), with an energy resolution of 7.1% at 661.7 keV (Figure 12a). The detector was coupled to a digiBASE photomultiplier tube, which integrates the PMT, high-voltage supply, and multichannel analyzer into a single compact unit for signal acquisition and spectral processing. To reduce the contribution of ambient background radiation, the detector was installed inside an Ortec HPLBS1 top-loading lead shield (Figure 12b).

Unlike HPGe detectors, NaI(Tl) scintillation detectors operate at room temperature and therefore do not require cryogenic cooling. Incident gamma rays interact with the scintillation crystal, producing flashes of visible light that are converted into electrical pulses by the photomultiplier tube. Although this detection principle results in lower energy resolution than semiconductor detectors, it provides a robust and cost-effective alternative for environmental gamma-ray spectrometry.

Figure 12 – Photographs of the NaI(Tl) detector, digiBASE PMT, and lead shielding used in the experimental measurements.

(a) Ortec NaI(Tl) 905–4 detector.

(b) Ortec HPLBS1 Top Loading Shield

Source: self-elaboration.

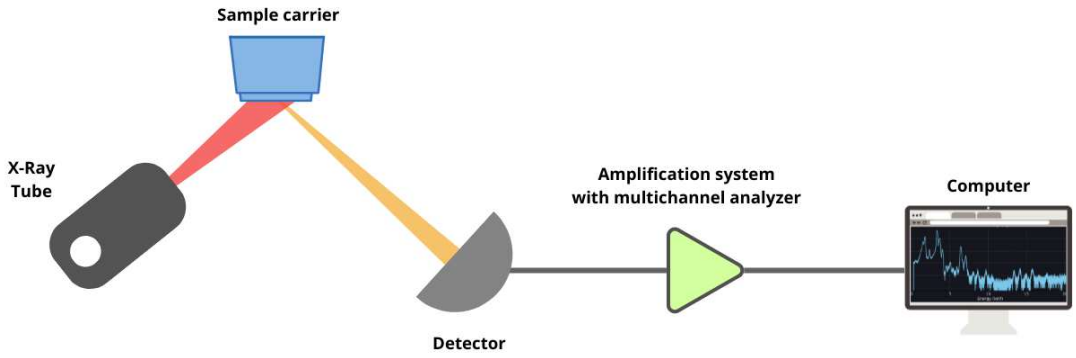
2.6 X-ray Fluorescence

X-ray fluorescence (XRF) is a versatile technique used for elemental analysis based on the interaction between X-rays and matter and, unlike GRS, relies on the excitation of the sample being analyzed. Given the non-destructive nature of this methodology, the technique may be employed in mineral, environmental, biological, and archeological contexts [23].

The excitation source in XRF systems is typically an X-ray tube, where high-energy electrons strike a metallic anode (e.g., Rh, Mo, or Ag), producing continuous and characteristic radiation. The geometry between the source, sample, and detector is carefully chosen to optimize excitation and minimize background scattering. Collimators and filters may be employed to improve spectral purity, and helium purging may be used for low-Z element detection.

XRF analysis consists of four basic stages: (i) excitation of the analyte by incident radiation from a source, resulting in the emission of one or more sets of characteristic X-rays; (ii) selection of characteristic emissions by energy or wavelength dispersion; (iii) detection of the selected emissions and measurement of their intensity; and (iv) conversion of emission intensity into elemental concentration [24]. The basic XRF schematics are displayed in Figure 13.

Figure 13 – Basic schematics for X-Ray fluorescence. In EDXRF, the different energies generate from the interaction between the incident beam and sample are dispersed inside the detector.



Source: self-elaboration.

Elements from Na to U are routinely measured using energy-dispersive X-ray fluorescence (EDXRF). Although direct measurements may be performed, adequate sample preparation generally yields results of higher quality and accuracy [23].

Quantitative X-ray analysis requires the application of suitable empirical or theoretical methods to produce satisfactory results. The measured X-ray intensity depends not only on the elemental concentration, but also on the matrix composition, sample geometry (shape, thickness), and preparation procedure, as well as on the measurement conditions — including, but not limited to, system geometry and detector efficiency [23]. Equation (2.19) expresses the relationship between radiation intensity, I_i, and the weight fraction of an element, W_i, in its simplest form.

$$I_i = k_i W_i + b_i \tag{2.19}$$

b_i corresponds to the spectral intensity when the analyte concentration is equal to zero, and the constant k_i corresponds to the sensitivity of the spectrometer, and is expressed in counts per second per unit of concentrations. Both these constants are determined by least-squares fit when calibrating the spectrometer, and their formulas are displayed in Equations (2.20a) and (2.20b). n corresponds to the amount of standards used for the calibration of the element i, W_{ij} is the weight fraction of the element i in the standard j, and I_{ij} is the intensity of the the element i in the standard j [23].

$$k_i = \frac{n\sum_j^n W_{ij}I_{ij} - \sum_i^n W_{ij}\sum_i^n I_{ij}}{n\sum_i^n W_{ij}^2 - \left(\sum_i^n W_{ij}\right)^2} \tag{2.20a}$$

$$b_i = \frac{\sum_i^n I_{ij} - K_i'\sum_i^n W_{ij}}{n} \tag{2.20b}$$

If calibration is performed using standards with elemental composition and concentrations similar to those of the analyzed matrix — employing the same sample preparation

procedure and measurement conditions — the system-related factors will be incorporated into k_i. In other words, matrix effects will be automatically accounted for. When calibration standards are unavailable, matrix effects can be corrected using fundamental parameter models or internal ratios such as the Compton peak intensity. [23].

The quantitative evaluation of emission peaks generally requires spectral deconvolution, especially in multi-element matrices where line overlaps are frequent (e.g., Kα–Kβ or L-series lines). Gaussian profiles are commonly fitted to isolate individual contributions and obtain net intensities [23, 24].

The energy of each emitted photon is characteristic of a specific element, whereas the emission intensity is proportional to its concentration. The resulting X-ray spectrum thus consists of a series of sharp peaks, each corresponding to characteristic electronic transitions of the elements present in the sample. In most EDXRF systems, these emissions are detected by a SDD, which measures both the energy and intensity of the incoming photons with high resolution [15].

The concentration LLD in XRF depends on the background noise, counting statistics, and excitation conditions, and is often expressed as three times the standard deviation of the background counts divided by the sensitivity, as show in Equation (2.21) [22].

$$\text{LLD} = \frac{3\sigma_b}{k_i} \tag{2.21}$$

2.6.1 Shimadzu EDX-720

The first EDXRF system employed in this work was a Shimadzu EDX-720 benchtop spectrometer [25], located at the Laboratório de Análises por Raio X (LARX) of UEL. This system is shown in Figure 14.

Figure 14 – Shimadzu EDX-720 EDXRF system.

Source: [25].

The Shimadzu EDX-720 is a laboratory-scale spectrometer designed for routine elemental analysis of solid and powdered samples. It operates over an elemental range

extending from sodium (^{11}Na) to uranium (^{92}U), allowing the detection of both light and heavy elements.

The system is equipped with a rhodium (Rh) X-ray tube as the excitation source and a silicon lithium-drifted semiconductor detector, Si(Li), for photon detection. This configuration provides adequate energy resolution for qualitative and quantitative elemental analysis in a wide range of matrices. Two measurement conditions were employed for the soil measurements performed with this system, one at 15 kV with a 100 s analysis time, suitable for Al-Ca elements; and another at 50 kV for 100 s, optimized for quantifying elements heavier than Ti [15].

2.6.2 Bruker Tracer 5i

The second XRF spectrometer employed in this work was a Bruker Tracer 5i, a portable XRF (pXRF) [26], located at LFNA–UEL. The system is shown in Figure 15.

Figure 15 – Bruker Tracer 5i pXRF system.

Source: [26].

The Tracer 5i is a handheld spectrometer designed for routine elemental analysis of solid and powdered samples. It operates over an elemental range extending from sodium (^{11}Na) to uranium (^{92}U), allowing the detection of both light and heavy elements.

The system is equipped with a rhodium (Rh) X-ray tube as the excitation source and a SDD detector. The device is also equipped with a set of internal calibrations that allow for direct qualitative and quantitative analysis. The measurements were performed in the GeoExploration mode, which encompasses three measurement conditions: (*i*) 30 kV and 13 μA applied to the tube with 25 μm Ti and 300 μm Al filters; (*ii*) 50 kV and 22.85 μA applied to the tube, with the two previously mentioned filters and a 75 μm Cu filter; and (*iii*) 15 kV and 12.14 μA applied to the tube without filters.

2.6.3 Panalytical Epsilon 5

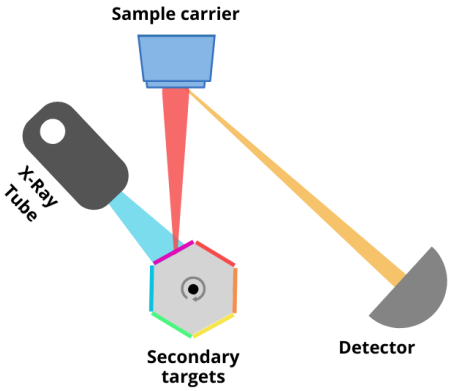
The third XRF system employed is this work was a Panalytical Epsilon 5, from the Nuclear Sciences and Instrumentation Laboratory (NSIL) at the IAEA Laboratories. The Epsilon 5 is designed for elemental analysis in solids, powders, and liquids, from sodium (^{11}Na) to uranium (^{92}U). The spectrometer is displayed in Figure 16a, and is equipped with a Sc/W tube, and Ge solid-state detector.

The optical path inside the Epsilon 5 is designed to polarize the X-rays, so that the scattered radiation from the tube does not reach the detectors. As a consequence, the background is reduced and the LLD becomes lower. Moreover, the spectrometer is equipped with up to 15 secondary targets that are excited by the tube's radiation, and whose fluorescence excites the sample. The Epsilon 5 schematics are displaced in Figure 16b.

Figure 16 – Photograph of the Panalytical Epsilon 5, and schematics of the optics inside the spectrometer.

(a) Panalytical Epsilon 5 Spectrometer.

(b) XRF schematics for secondary targets inside the Epsilon 5.



(a) Source: [27]. (b) Source: self-elaboration.

The calibration focused on the detection and quantification of light metals in soil matrices, as these are key indicators of soil fertility and nutrient availability. Therefore, only four measurement conditions were employed, and are described in Table 2.

Table 2 – XRF Spectrometer conditions employed to perform the measurements.

Target	**Detected Elements**	**Current (mA)**	**Tension (kV)**	**Live time (s)**
Al	Na, Mg	24	25	300
CaF_2	Al, Si, P, K	15	40	300
Fe	Ca, Ti, V, Cr	8	75	300
Ge	Mn, Fe, Ni, Cu, Zn	5	75	600

2.7 Mid-Infrared Spectroscopy

Unlike GRS and XRF, which rely on the interaction of ionizing radiation with matter, the basic principle of mid-infrared spectroscopy (MIRS) lies in the interaction of infrared radiation with the molecular bonds present in the sample. When the frequency of the incident radiation matches the natural vibrational frequency of a bond, energy absorption occurs, producing characteristic spectral features that reflect the molecular composition of the material. The technique is particularly useful for identifying functional groups and organic or mineral components containing hydroxyl, carbonate, or silicate structures [28].

In the mid-infrared region (approximately 4000–400 cm^{-1}), absorption bands correspond mainly to fundamental stretching and bending vibrations. The MIR spectrum can be divided into four regions: (*i*) the X–H stretching region (4000–2500 cm^{-1}); (*ii*) the triple-bond region (2500–2000 cm^{-1}); (*iii*) the double-bond region (2000–1500 cm^{-1}); and (*iv*) the fingerprint region (1500–600 cm^{-1}). Table 3 summarizes some of the main bands typically observed in MIRS spectra and their corresponding fingerprint regions.

Table 3 – Common absorption bands in the mid-infrared region and their corresponding fingerprint regions.

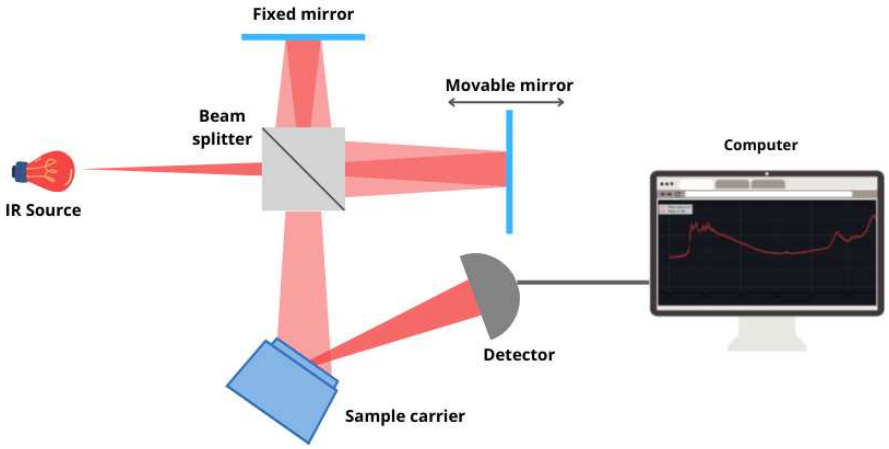
Bond	**Region (cm^{-1})**	**Note**
O−H	3700–3600	Broad
N−H	3400–3300	Narrow
C−H	3000–2850	
C≡C	2300–2050	Very weak
C≡N	2300–2200	Medium intensity
C=O	1830–1650	Most intense
C=C	Around 1650	Weak
C=N	Around 1650	Strong

The resulting spectrum yields a unique chemical fingerprint of the sample, containing information on both organic and inorganic constituents. Because MIRS measurements are rapid, non-destructive, and require little or no sample preparation, the technique is widely applied in soil, mineral, and environmental analyses [28].

Since H_2O and CO_2 are constantly present in the atmosphere, their absorption features must be taken into account when designing the spectrometer. These contributions are compensated for by using a beam splitter, in which one of the split beams interacts with the sample, while the other passes through a reference path. In Fourier-transform infrared (FTIR) spectroscopy, the difference in optical path length between the two beams produces an interferogram, which contains the spectral information of the sample as a function of the wavenumber [28, 29].

FTIR spectrometers are commonly equipped with a Globar[1] or Nersnt[2] source to emit radiation in the mid-infrared region. For the detection, usually a deuterated triglycine sulfate (DTGS) operating at room temperature is employed. Depending on the module configuration, radiation-sample interaction may occur by attenuated total reflectance (ATR) or by diffuse reflectance (DRIFTS), both suitable for solid and powdered materials. These reflection-based approaches are advantageous because they minimize alignment requirements and allow direct analysis of heterogeneous or rough surfaces. Figure 17 illustrates the basic schematics of a FTIR spectroscope.

Figure 17 – Basic schematics of a FTIR spectroscope.



Source: self-elaboration.

The relationship between absorbance and concentration of a compound in MIRS spectra is dictated by the Beer–Lambert law:

$$A = \varepsilon C l \tag{2.22}$$

where A is the absorbance, ε is the molar absorptivity, C is the concentration, and l is the effective optical path length. The absorbance corresponds to the logarithmic difference between the incident and transmitted intensities.

In complex matrices, such as soils, where spectral features of different components strongly overlap, multivariate calibration methods are applied to extract compositional information. Techniques such as PCR and PLSR are commonly used to relate spectral data to reference concentrations [30]. These empirical calibrations implicitly account for matrix effects, path-length variations, and other system-specific factors, analogous to the sensitivity constants obtained in XRF and GRS calibrations.

The analytical performance of MIRS depends on the detector sensitivity, optical configuration, and surface homogeneity of the sample. Although the technique is less

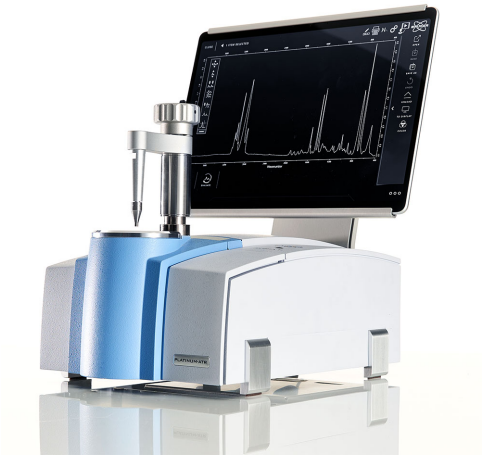
[1] Silicon carbide (SiC) rod, heated to temperatures between 1000–1650°C.
[2] Ceramic rod heated to incandescence.

element-specific than XRF or GRS, it provides complementary information on chemical bonding, organic matter composition, and mineralogical structure, offering a valuable framework for the integrated characterization of complex materials.

2.7.1 Bruker Alpha II

The FTIR spectroscope employed in this work was a Bruker Alpha II, from the Soil and Water Management and Crop Nutrition Laboratory (SWMCNL) at the IAEA Laboratories, and is shown in Figure 18. This compact spectrometer is suitable for studying the chemical composition of polymeric, forensic, cultural heritage, pharmacological, and environmental samples [31].

Figure 18 – Bruker Alpha II FTIR spectroscope.



Source: [31]

The system is equipped with a set of sampling modules that may be changed depending on the matrix to be analyzed. These modules allow for measurements of transmission, external, diffuse reflectance or attenuated total reflection.

3 ENVIRONMENTAL RADIOACTIVITY

Despite popular belief, that radiation is harmful and should be avoided, life has coexisted with exposure to radiation since the first unicellular organism came to existence. Although large doses are indeed harmful and may cause unhealthy and undesired effects, some amounts of radiation are harmless. In fact, the main sources of natural radioactivity are the radionuclides from the ^{238}U and ^{232}Th decay chains, as well as ^{40}K [32].

Part of the radiation all living things are constantly exposed to stem from natural radionuclides, that may also be called Naturally Occurring Radioactive Materials (NORMs). Those may be divided into two categories: primordial and cosmogenic radionuclides. Primordial radionuclides will be discussed in details in the following section, but the only cosmogenic radionuclide that presents relevance to this work, ^{7}Be, will be addressed alongside fallout radionuclides.

3.1 Primordial Radionuclides

The so-called primordial radionuclides are those whose half-lives surpasses or are comparable to the age of the Earth (4.7×10^9 years) [14]. This characteristic indicates that these elements might have been created during, or even prior to, the formation of the solar system itself [33].

Primordial radionuclides may be single, isolated elements or might be part of a decay chain whose father's half-life is sufficiently long. Table 4 shows a few known primordial radionuclides that are not a part of decay chain, alongside their half-lives and decay mode.

Table 4 – A few known primordial radionuclides, with their half-lives and decay modes.

Radionuclide	Half-Life (years)	Decay mode
^{40}K	1.25×10^9	β^-, EC
^{87}Rb	4.97×10^{10}	β^-
^{123}Te	$> 9.2 \times 10^{16}$	EC
^{144}Nd	2.29×10^{15}	α
^{147}Sm	1.06×10^{11}	α
^{174}Hf	2×10^{15}	α
^{176}Lu	3.76×10^{10}	β^-
^{187}Re	4.33×10^{10}	β^-
^{190}Pt	6.5×10^{11}	α

Source: [14], adapted.

Although only a few examples are shown here, the list goes on. The long half-lives may lead to difficulties in identifying these isotopes as radioactive instead of stable [14]. Primordial radionuclides may be found in association with specific geological systems, which makes them efficient chronological tools. Rock and mineral dating may be performed by evaluating the concentrations of ^{40}K, ^{87}Rb, ^{174}Hf, and ^{176}Lu, for example [34].

3.1.1 Natural Radioactive Series

Among the primordial radionuclides, ^{238}U, ^{235}U and ^{232}Th are particularly interesting. Not only their half-lives are long enough to be comparable with Earth's age, but they originate long decay chains whose daughter elements have variable half-lives. The three of them undergo several decays until reaching a stable Pb isotope. There is also a fourth natural radioactive series, the one from ^{237}Np, but this one is a special case which will be discussed later. Radioactive series are defined by the mass number of their components. For example, the ^{238}U family is called the $4n + 2$ series because the mass of all isotopes in the chain are divisible by 4 with a remainder of 2. The main characteristic of the four decay series are summarized in Table 5. For detailed information on the daughter nuclides and transitions experienced in each decay chain see Appendix A.

Table 5 – Main characteristics of the four natural radioactive series.

Mass	Series	Parent	Half-life (years)	Stable isotope
$4n$	Thorium	^{232}Th	1.41×10^{10}	^{208}Pb
$4n + 1$	Neptunium	^{237}Np	2.14×10^{6}	^{205}Pb
$4n + 2$	Uranium-Radium	^{238}U	4.51×10^{9}	^{206}Pb
$4n + 3$	Uranium-Actinium	^{235}U	7.18×10^{8}	^{207}Pb

Source: [14].

As can be seen in Table 5, ^{237}Np has a half-life that is 3 orders of magnitude smaller than the age of the Earth. This means that the only isotope from the $4n+1$ series found in nature is ^{205}Pb, the stable element at the end of the series.

All daughter radionuclides from the other three series may be found in nature. Despite their short half-lives in comparison with the Earth's age, the decay from the parent radionuclide allows for the continuos production of the other isotopes. In isolated systems, secular equilibrium may be reached, leading to all elements in the decay chain having equal activities at some point.

3.2 Fallout Radionuclides

Both natural and artificial radionuclides may be classified as fallout radionuclides (FRNs). This name is attributed to radioactive isotopes that, after some time in the

atmosphere, fall to the ground by wet or dry deposition where they are rapidly and strongly adsorbed by fine soil particles.

The basis of the FRN method is to consider that the spatial deposition is uniformly distributed — at least over a small area — and compare the vertical distribution at determined points. The differences in concentration and distribution of a given FRN between eroded and undisturbed regions are what allow to evaluate soil erosion using radionuclides.

The differences between half-lives and penetration in soil makes different FRN suitable for soil erosion evaluation in different timescales and situations. These details and differences will be addressed in the following sections.

3.2.1 Beryllium-7

Beryllium (Be) is the 4th element of the periodic table. The most notable unstable isotopes of this element are ^{7}Be and ^{10}Be, both of cosmogenic origin. Particularly, ^{7}Be has a half-life of 53.12 days and is continuously produced in the stratosphere and troposphere by nitrogen and oxygen spallation [35]. The processes that might lead to the creation of this radionuclide are shown in Equation 3.1 [36].

$$
\begin{aligned}
&{}^{12}_{6}\mathrm{C} + \mathrm{p} \longrightarrow {}^{7}_{4}\mathrm{Be} + {}^{6}_{3}\mathrm{Li} && {}^{12}_{6}\mathrm{C} + \mathrm{n} \longrightarrow {}^{7}_{4}\mathrm{Be} + {}^{6}_{2}\mathrm{He} \\
&{}^{14}_{7}\mathrm{N} + \mathrm{p} \longrightarrow {}^{7}_{4}\mathrm{Be} + 2\,{}^{4}_{2}\mathrm{He} && {}^{14}_{7}\mathrm{N} + \mathrm{n} \longrightarrow {}^{7}_{4}\mathrm{Be} + {}^{8}_{3}\mathrm{Li} \\
&{}^{16}_{8}\mathrm{O} + \mathrm{p} \longrightarrow \begin{cases} {}^{7}_{4}\mathrm{Be} + {}^{10}_{5}\mathrm{B} \\ {}^{7}_{4}\mathrm{Be} + {}^{7}_{3}\mathrm{Li} + {}^{3}_{2}\mathrm{He} \end{cases} && {}^{16}_{8}\mathrm{O} + \mathrm{n} \longrightarrow \begin{cases} {}^{7}_{4}\mathrm{Be} + {}^{10}_{4}\mathrm{Be} \\ {}^{7}_{4}\mathrm{Be} + {}^{6}_{2}\mathrm{He} + {}^{4}_{2}\mathrm{He} \end{cases}
\end{aligned}
\tag{3.1}
$$

The production of this isotope is related to solar activity, since cosmic rays originated from the Sun are the precursors of this reaction [4]. Moreover, the ^{7}Be production is influenced by the latitude showing higher concentrations near the poles due to radiation deflections towards these regions. The decay of ^{7}Be is described in Equation (3.2).

$$
{}^{7}\mathrm{Be} \xrightarrow{\mathrm{EC}} {}^{7}\mathrm{Li}^{m} \xrightarrow{\gamma} {}^{7}\mathrm{Li} \tag{3.2}
$$

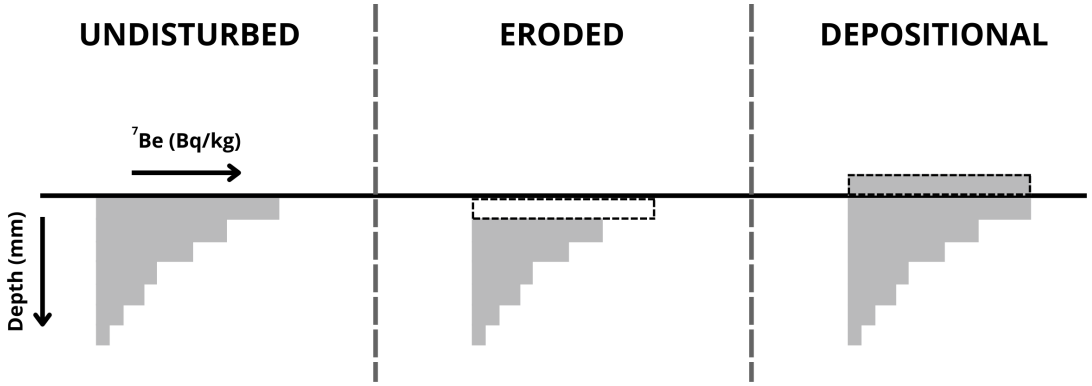
The photon emitted by this decay has a well-defined energy of 477.6 keV and may be resolved by most gamma-ray detectors available today.

After formation, ^{7}Be becomes associated with aerosols and get subjected to atmospheric transport processes. The fallout is mainly wet, i.e., associated with rain, with approximately 10% of the total deposition happening via dry fallout. After the fallout, the ionic form of this isotope is $^{7}\mathrm{Be}^{2+}$ which causes it to be quickly and strongly adsorbed by soil particles [4, 35].

Given the double positive charge of the ion, ^{7}Be is able to penetrate only a few centimeters in soil showing a decreasing exponential profile as the depth increases [37]. Moreover, given its short half-live there is little opportunity for further downward transfer [4]. Figure 19 shows the typical vertical distribution of ^{7}Be in soil for undisturbed, eroded and depositional sites.

Figure 19 – Vertical distribution of ^{7}Be in undisturbed, eroded and depositional soil sites.

Commonly, the maximum reached depths are no more than 30 mm. Assuming the three sites are nearby, all experienced the same amount of fallout, indicate by the black horizontal line. As soil redistribution happened, eroded sites lose the first few millimeters of soil, as well as the ^{7}Be bound to these soil particles. On the other hand, depositional sites gain soil and ^{7}Be from other places.



Source: [4], adapted.

^{7}Be short half-life and shallow penetration in soil allows this radionuclide to be suitable for short-term erosion assessment. This way, this type of analysis may be employed to evaluate the effectiveness of recent soil conservation measures or the impacts of land use change [38].

The IAEA recommends that the Profile Distribution model is employed to quantify soil losses using the ^{7}Be approach [38]. The exponential depth distribution of the tracer in soil is described as:

$$C(x) = ce^{-x/z_0} \tag{3.3}$$

where $C(x)$ (Bq kg^{-1}) is the ^{7}Be activity at mass depth x (kg m^{-2}), c is a constant value and z_0 is the relaxation mass depth. This last parameter is particularly important as it indicates the depth in which 63.2% of the ^{7}Be inventory is found.

The erosion rates (kg m^{-2}) experienced in a particular site can be estimated by comparing the local's ^{7}Be inventory, I (Bq m^{-2}), with the inventory measured at the reference site, I_{ref}. The relationship between these variables is displayed in Equation (3.4).

$$I = \int_{-z}^{\infty} C(x)\mathrm{d}x = I_{\text{ref}}e^{z/z_0} \tag{3.4}$$

The erosion rate, z, can, therefore, be calculated as:

$$z = z_0 \ln\left(\frac{I}{I_{\text{ref}}}\right) \tag{3.5}$$

and since z represents soil *loss* in a given point, its value is necessarily negative.

In case of inventory values greater than the reference, Equation (3.6) is employed to estimate the deposition rate, z' (kg m^{-2}).

$$h' = \frac{I - I_{\text{ref}}}{C_d} \tag{3.6}$$

The deposited concentration of ^{7}Be, C_d (Bq kg^{-1}), is calculated employing Equation (3.7), where S represents the sampling area.

$$C_d = \frac{1}{\int_S z \mathrm{d}S} \int_S C_e z \mathrm{d}S \tag{3.7}$$

Finally, C_e represents the concentration of eroded ^{7}Be and is calculated from the loss of inventory divided by the mass of the soil loss, as shown in Equation (3.8).

$$C_e = \frac{I_{\text{ref}}(1 - e^{-z/z_0})}{z} \tag{3.8}$$

3.2.2 Caesium-137

The most commonly employed FRN to evaluate soil erosion rates is ^{137}Cs [1]. Caesium (Cs) is the 55th atomic element of the periodic table and its only stable isotope is ^{133}Cs. ^{137}Cs is an artificial radionuclide, a sub-product of ^{235}U as shown in Equation (3.9). The main sources of ^{137}Cs in the environment originate from the nuclear weapons tests carried during 1950 and 1970, with a maximum release occurring in 1963 in the northern hemisphere, and slightly later for the southern hemisphere [1].

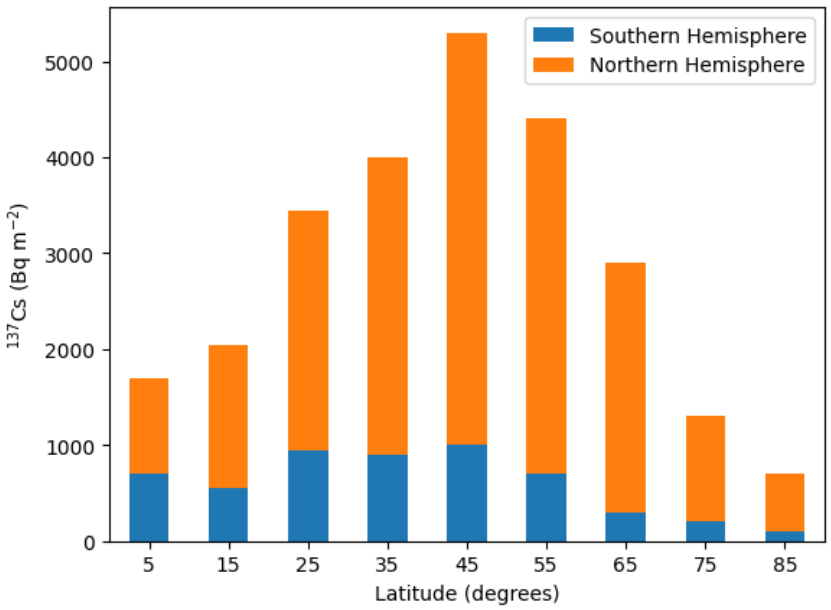
$$\begin{aligned} &^{235}\mathrm{U} + \mathrm{n} \longrightarrow {}^{236}\mathrm{U} \longrightarrow {}^{137}\mathrm{I} + \mathrm{X} \\ &^{137}\mathrm{I} \xrightarrow{\beta^-} {}^{137}\mathrm{Xe} \xrightarrow{\beta^-} {}^{137}\mathrm{Cs} \xrightarrow{\beta^-} {}^{137}\mathrm{Ba}^m \xrightarrow{\gamma} {}^{137}\mathrm{Ba} \end{aligned} \tag{3.9}$$

The photon emitted by ^{137}Ba^m has a well-defined energy of 661.66 keV and may be resolved by most gamma-ray detectors available today. With its half-life of 30.17 years, ^{137}Cs is a suitable tracer for soil redistribution between 1963 and nowadays.

Bomb-derived ^{137}Cs was released into the stratosphere, where it circulated globally before reaching the surface by fallout [1, 39]. Globally, most of the fallout happened due to precipitation, but locally, dry fallout was also significant [1, 8]. The spatial distribution of ^{137}Cs fallout was determined by the local of the weapons testing, the patterns of

stratospheric circulation and the annual precipitation following the explosions. The global distribution of ^{137}Cs is strongly correlated with the latitudinal rainfall distribution [1, 40] and, overall, the depositions in the northern hemisphere are more significant than in the southern hemisphere, as shown in Figure 20.

Figure 20 – Global distribution of bomb-derived fallout ^{137}Cs.



Source: [1], adapted.

This way, ^{137}Cs is more easily detected in the northern hemisphere. Low-resolution radiation detectors are not suitable for ^{137}Cs monitoring in the southern hemisphere due to the lower concentrations, that are normally below the lower limit of detection (LLD) and the emission can not be resolved from the background. In order to quantify ^{137}Cs in the southern hemisphere it is necessary to use high-resolution detectors with long measurement times.

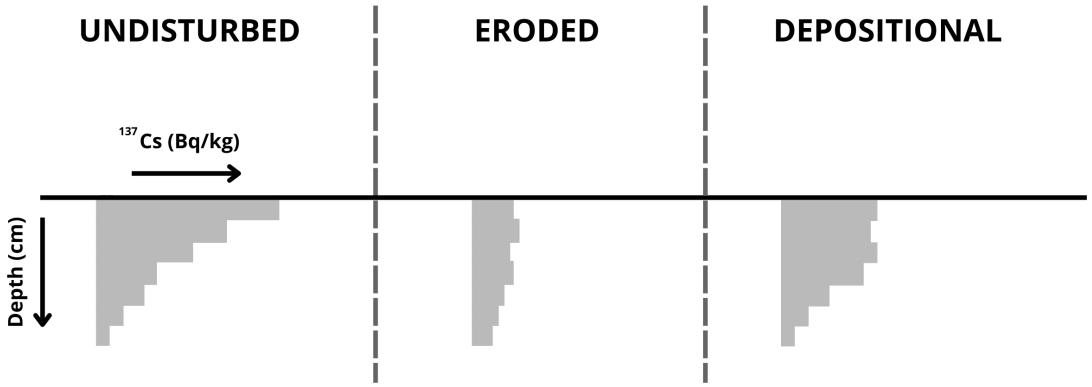
In undisturbed soils, ^{137}Cs vertical distribution in soil is similar to ^{7}Be, with the main difference being the reached depth. ^{137}Cs may reach up to 30 cm in agricultural areas [1]. Due to continuous plowing, the ^{137}Cs concentration in mechanized soils is dissolved almost uniformly, leading to a profile different than what is observed in undisturbed regions, as represented in Figure 21.

The use of the ^{137}Cs technique to evaluate soil redistribution offers several advantages when compared with traditional techniques [1], such as:

1. The methodology allows for retrospective evaluation;
2. The results obtained with the technique encompasses not only erosion caused by land use, but also by climatic effects;

3. The technique provides information on both erosion and deposition;
4. The estimates are based on individual samplings points distributed across the studied area, and can be extrapolated to generate patterns of redistribution;
5. The sampling is simple and not costly. Moreover, it is not necessary to perform long-term monitoring, a single visit to the study site is enough to provide the desired results;
6. The on-site disturbance while sampling is minimal and will not interfere with cultivation operations.

Figure 21 – Vertical distribution of ^{137}Cs in undisturbed and mechanized soil sites.



Source: [4], adapted.

On the other hand, the technique also has some disadvantages and limitations [1], such as:

1. In some area, the ^{137}Cs inventories are low, therefore the measurements require longer counting times;
2. In some regions, the additional fallout caused by the Chernobyl accident must be taken into account;
3. The technique is suitable for evaluation medium-term average erosion rates and can not be adjusted for short-term analysis;
4. It is necessary to take into account the necessity of representativeness and variability of the sampled points. Therefore, it is necessary to collect multiple or replicate samples to determine the inventory values.

Several mathematical models have been developed to convert ^{137}Cs concentrations in soils into quantitative estimates of erosion and deposition rates [9,41,42]. The simplest, and one of the most applied models, is the Proportional Model. The key consideration of

this model is that the fallout ^{137}Cs inputted in soil is completely mixed within the plow or cultivation layer and, therefore, any soil gain or loss is directly proportional to the fraction of the original ^{137}Cs inventory moved from the soil profile since the first deposition. The model provides either the mean annual loss or gain since the beginning of the ^{137}Cs fallout, and these values can be obtained with Equation (3.10) [1].

$$Y = 10\frac{Bd}{Tp}\left(\frac{I - I_{\text{ref}}}{I_{\text{ref}}}\right) \tag{3.10}$$

Y is the mean annual soil loss since 1963 (ton$\cdot$ ha^{-1} $\cdot$ year^{-1}), d is the cultivation depth (m), B is the soil bulk density (kg$\cdot$m^{-3}), I_{ref} is the reference inventory (Bq$\cdot$m^{-2}), I is the inventory of a given site (Bq$\cdot$m^{-2}), T is the time elapsed since 1963 (years), and p is the particle size corrections factor. By inverting I and I_{ref} in the fraction, it is possible to obtain the mean annual soil gain.

The particle size correction factor must be introduced to account the potential particle size selectivity of erosion and deposition processes. Overall, erosion favors the removal of finer soil particles while deposition favors the movement of coarser particles.

The main advantage of the Proportional Model relies on its simplicity, and in the little amount of information it requires. But, as a consequence, the assumptions of this model oversimplifies the behavior of ^{137}Cs in soil and often leads to inaccurate results.

4 DATA ANALYSIS AND CHEMOMETRIC TOOLS

Performing the measurements described in Sections 2.5 to 2.7 yields a set of spectral results with number of variables (energy, wavelength, or detector channel) ranging from 1024 to 8192. Moreover, the activity and elemental concentration calculations, described in Sections 2.5 and 2.6, yields a set of quantitative results. In such cases, it is necessary to use multivariate data analysis methods to simultaneously analyze the variables contained in the datasets and their relationships [43, 44].

In the field of multivariate analysis, two classes of methods were employed in this work: exploratory analysis and machine learning calibration. For exploratory methods, Principal Component Analysis (PCA) and Common Dimensions Analysis (ComDim) were employed. Regarding machine learning, Partial Least Squares Regression (PLSR) and Random Forest (RF) were performed to predict soil fertility attributes based on spectral data. At last, calibration transfer (CT) was also employed to transform spectral data. More specifically, Target Transformation Factor Analysis (TTFA) was applied.

When working with georeferenced data, it is useful to combine the analytical data with the coordinates of the sampling to create maps displaying the spatial distribution of a quantity of interest. From this perspective, Inverse Distance Weighting (IDW), a spatial interpolation method, was employed.

4.1 Exploratory Analysis

Exploratory analysis is the first step to be performed when analyzing a dataset. Essentially, it consist of the examination and characterization of data in order to find hidden patterns and relationships, possible outliers, and underlying characteristics.

4.1.1 Principal Component Analysis

PCA is one of the most established and extensively employed multivariate analysis techniques [43, 44]. Originally proposed by Karl Pearson in 1901 [45] and later formalized by Hotelling [46], PCA is an unsupervised method that aims to reduce the dimensionality of a dataset while retaining most of its original variance. This dimensionality reduction facilitates the visualization and interpretation of latent relationships within complex datasets. The decomposition of matrices in PCA is conventionally performed through Singular Value Decomposition (SVD) [44].

In PCA, the original correlated variables are transformed into a reduced set of uncorrelated latent variables, known as principal components (PCs), which represent the

directions of maximum variability in the data. The transformation from the original variable space to the principal component space is achieved by decomposing the data matrix $\mathbf{X}_{m\times n}$ into three matrices: the scores $\mathbf{V}_{m\times k}$, the loadings $\mathbf{P}_{k\times n}$, and the residuals $\mathbf{E}_{m\times n}$, where m denotes the number of samples, n the number of variables, and k the number of retained components. The value of k is defined a priori by the user, typically based on criteria such as the explained variance, interpretability of the components, or application-specific requirements. The relationships among these matrices are expressed in linear and matrix forms in Equations (4.1a) and (4.1b).

$$\mathbf{X}_{m\times n} = \mathbf{V}_{m\times k} \times \mathbf{P}^T_{k\times n} + \mathbf{E}_{m\times n} \tag{4.1a}$$

$$\begin{aligned}\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix} &= \begin{bmatrix} V_{11} & \dots & V_{1k} \\ V_{21} & \dots & V_{2k} \\ \vdots & \vdots & \vdots \\ V_{m1} & \dots & V_{mk} \end{bmatrix} \times \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ P_{k1} & P_{k2} & \dots & P_{kn} \end{bmatrix} \\ &+ \begin{bmatrix} E_{11} & E_{12} & \dots & E_{1n} \\ E_{21} & E_{22} & \dots & E_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ E_{m1} & E_{m2} & \dots & E_{mn} \end{bmatrix}\end{aligned} \tag{4.1b}$$

Each principal component possesses an associated value that quantifies its contribution to the total variance of the dataset, referred to as the explained variance ratio. The explained variance decreases progressively with each subsequent component, and the cumulative contribution of all components equals 100%.

The loadings matrix indicates the contribution of each original variable to the formation of the PCs, whereas the scores matrix contains the coordinates of the samples in the reduced-dimensional space. The residuals matrix represents the portion of information not captured by the selected PCs and becomes a null matrix when $k = m$.

In spectrometric applications, PCA is frequently employed to investigate how different spectral features contribute to the discrimination and characterization of samples. The analysis can be conducted using spectra-derived variables or directly from the raw spectra, enabling the identification of the most relevant energies (for XRF and GRS) or wavelengths (for IR-based techniques) in describing the system under study. In this work, PCA was performed in Python using the `PCA` function from the `scikit-learn` library, version 1.4.2 [47].

When the same set of samples is analyzed by two distinct techniques, the relationship between the resulting variables can be examined by concatenating the corresponding data matrices, $\mathbf{X}^{(1)}_{m\times n}$ and $\mathbf{X}^{(2)}_{m\times l}$, where n and l denote the number of variables in each

block. The resulting joint matrix may then be subjected to PCA to explore the covariance structure across both measurement domains. This approach, in which raw variables are merged prior to modeling, constitutes a *low-level data fusion* strategy.

4.1.1.1 Multi-Block PCA

Multi-Block Principal Component Analysis (MB-PCA) is an unsupervised multivariate technique conceptually analogous to standard PCA, but specifically designed to handle datasets composed of multiple blocks of variables. In addition to revealing the relationships among samples and variables, MB-PCA also quantifies the relative contribution of each data block to the formation of the principal components, analogous to the explained variance ratio in conventional PCA.

The method involves performing dimensionality reduction (e.g., via SVD) independently on each of the K data blocks. The resulting latent matrices are subsequently concatenated into a global matrix, which is then decomposed again through SVD to extract shared components. This procedure enables the investigation of the complementarity or synergy between different datasets, such as spectra-derived quantities or raw spectra obtained from multiple analytical techniques.

The number of block components, k_b, retained from each data block and the number of global components, k_g, extracted from the concatenated latent matrix are defined a priori by the user. MB-PCA is categorized as a *mid-level data fusion* technique, as it relies on the extraction of latent features from individual data blocks prior to their integration into a common low-dimensional representation. Since there is no native Python package to perform MB-PCA, a custom pipeline was built based upon the `PCA` function from the `scikit-learn` library, version 1.4.2 [47].

4.1.2 Common Dimensions Analysis

Common Dimensions Analysis (ComDim) aims to identify a shared latent space that best represents all blocks within a multivariate dataset [48, 49]. In this space, the method extracts the so-called common dimensions (CDs), which enable the simultaneous recognition of patterns across all measurement domains.

The multi-block dataset is composed of K data blocks, denoted as $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_K$. Each block corresponds to one analytical technique or data source. The decomposition defining each common dimension is expressed in Equation (4.2).

$$\boldsymbol{W}_k = \boldsymbol{X}_k \boldsymbol{X}_k^T = \sum_{j=1}^{n} \lambda_{k,j} \boldsymbol{q}_j \boldsymbol{q}_j^T \tag{4.2}$$

In this formulation, $\boldsymbol{W}_k$ $(k = 1, 2, \ldots, K)$ represents the cross-product matrix of

the kth data block, while $\lambda_{k,j}$ denotes the specific weight (or *salience*) assigned to block k for the jth common dimension. The vector $\boldsymbol{q}_j$ contains the global scores corresponding to the jth CD. Prior to decomposition, each covariance-variance matrix $\boldsymbol{X}_k\boldsymbol{X}_k^T$ is column-centered and normalized by its Frobenius norm [50, 51].

The estimation of common components and their associated saliences is achieved by minimizing the threshold function $\boldsymbol{T}$, defined in Equation (4.3) [52]. The algorithm proceeds sequentially: the first CD and its corresponding weights are extracted, followed by the second CD, and so forth. Components are ordered according to their importance, defined as the proportion of variance jointly explained by all blocks. This ordering enables the extraction process to be terminated once subsequent CDs represent only statistical noise or fluctuations [52, 53].

$$\boldsymbol{T} = \sum_{k=1}^{K} \left|\left| \boldsymbol{W}_k - \sum_{j=1}^{n} \lambda_{k,j}\boldsymbol{q}_j\boldsymbol{q}_j^T \right|\right|^2 \tag{4.3}$$

The extraction of the first common dimension involves finding a unit-length vector $\boldsymbol{q}_1$ that minimizes the threshold function for $j = 1$. The procedure follows these main steps:

(i) initialize all $\lambda_{k,1}$ $(k = 1, 2, \ldots, K)$ as unity;

(ii) determine the optimal vector $\boldsymbol{q}_1$ by computing the standardized eigenvector of $\sum_{k=1}^{K} \lambda_{k,1}\boldsymbol{X}_k\boldsymbol{X}_k^T$ associated with the largest eigenvalue;

(iii) update the block-specific weights as $\lambda_{k,1} = \boldsymbol{q}_1^T\boldsymbol{W}_k\boldsymbol{q}_1$; and

(iv) compute a measure of fit as $\boldsymbol{T}_1 = \sum_{k=1}^{K} ||\lambda_{k,1}\boldsymbol{q}_1\boldsymbol{q}_1^T||^2$.

Steps (ii)–(iv) are repeated iteratively until $\boldsymbol{T}_1$ reaches a predefined convergence criterion [52, 53]. Once the first CD is obtained, each data block is deflated according to Equation (4.4), where $\boldsymbol{X}_k^*$ denotes the deflated matrix and $\mathbb{I}$ is the identity matrix. The algorithm is then restarted to determine the subsequent CDs. This iterative deflation continues until the desired number of common dimensions is extracted [49, 51, 53, 54].

$$\boldsymbol{X}_k^* = (\mathbb{I} - \boldsymbol{q}_1\boldsymbol{q}_1^T)\boldsymbol{X}_k \tag{4.4}$$

The salience $\lambda_{k,j}$ indicates the proportion of variance from block $\boldsymbol{X}_k$ that is represented in the jth CD, and can thus be interpreted as the relative importance of each block for a given common dimension [51]. The global scores describe the projection of the samples within the CD space, while the block-specific loadings quantify the contribution of individual variables to each dimension [50, 51], analogously to the interpretation of PCA.

Consequently, by analyzing the scores, loadings, and saliences obtained for each common dimension, it becomes possible to assess the extent to which the different analytical techniques contribute to the overall characterization of the samples. These results can be visualized through two- or three-dimensional representations, facilitating the identification of latent relationships across the combined dataset while preserving the variance structure among samples and variables [55].

From a methodological perspective, ComDim shares conceptual similarities with MB-PCA, as both seek to identify common structures across several data blocks. However, a key distinction lies in the optimization criterion: MB-PCA maximizes the total explained variance of concatenated block scores subject to block scaling constraints, whereas ComDim explicitly identifies a set of common dimensions by optimizing shared variance across all blocks through salience weighting. Consequently, ComDim provides a more direct quantification of each block's contribution to the common latent structure, offering enhanced interpretability when the analytical sources exhibit different variances or measurement scales [55]. In this work, ComDim was performed in the Julia programming language, employing the `Jchemo` package [56].

4.2 Machine Learning Models

Machine learning (ML) models are commonly employed when a dataset of measurements, $\mathbf{X}_{m\times n}$, and a corresponding set of outcomes, $\mathbf{Y}_m$, are available, with the goal of estimating a function $\mathcal{F}$ that maps each input sample $\mathbf{X}_i$ to its corresponding outcome Y_i, such that

$$\mathcal{F}(\mathbf{X}_{m\times n}) = \mathbf{Y}_m. \tag{4.5}$$

These ML models are classified as *supervised learning* models, since the target outcomes $\mathbf{Y}_m$ associated with $\mathbf{X}_{m\times n}$ are known *a priori*. Different models assume different relationships between both datasets — i.e. linear or non-linear — and employ different strategies to determine $\mathcal{F}$.

4.2.1 Partial Least Squares Regression

Partial Least Squares Regression (PLSR) is a supervised multivariate regression technique widely used when the goal is to model linear relationships between two datasets measured on the same set of samples. In contrast to PCA, PLSR simultaneously analyzes the input matrix and a corresponding target matrix in order to maximize the covariance between their latent variables [57]. Depending on the dimensionality of the response matrix, PLSR can be implemented as Partial Least Squares 1 (PLS1), in which a separate model is constructed for each response variable, or Partial Least Squares 2 (PLS2), in which multiple response variables are modeled simultaneously within a single framework.

In this study, the PLS1 approach was adopted, as individual models were developed for each soil fertility parameter.

Analogously to PCA, PLSR relies on the decomposition of the input matrix; however, in PLSR both the input matrix $\mathbf{X}_{m\times n}$ and the target matrix $\mathbf{Y}_{m\times r}$ are decomposed using different optimization criteria[1]. These decompositions are expressed as

$$\mathbf{X}_{m\times n} = \mathbf{T}_{m\times f}\mathbf{P}^T_{f\times n} + \mathbf{E}_{\mathbf{X}m\times n} \tag{4.6a}$$

$$\mathbf{Y}_{m\times r} = \mathbf{U}_{m\times f}\mathbf{Q}^T_{f\times r} + \mathbf{E}_{\mathbf{Y}m\times r} \tag{4.6b}$$

where m denotes the number of samples, n and r the number of variables in the input and target matrices, respectively, and f the number of retained latent components. The matrices $\mathbf{T}$ and $\mathbf{U}$ contain the score vectors associated with $\mathbf{X}$ and $\mathbf{Y}$, while $\mathbf{P}$ and $\mathbf{Q}$ contain the corresponding loadings. The residual matrices $\mathbf{E_X}$ and $\mathbf{E_Y}$ represent the portions of the data not explained by the selected components [57].

Unlike PCA, the latent components in PLSR are obtained by maximizing the covariance between the scores of $\mathbf{X}$ and $\mathbf{Y}$. As a result, the extracted components are directly related to the predictive relationship between both datasets, making PLSR particularly suitable for regression problems involving multicollinear or high-dimensional data.

When applying a PLSR model for prediction, the score matrix $\mathbf{T}$ associated with an unknown input matrix $\mathbf{X}$ must first be estimated using the previously determined loading matrix $\mathbf{P}$. The predicted values of the target matrix are then obtained from the regression relationship between the scores of $\mathbf{X}$ and $\mathbf{Y}$. In this context, the predicted target values for an unknown sample can be expressed as

$$\mathbf{Y}_{\text{unknown}} = \mathbf{T}_{\text{unknown}}\mathbf{B}\mathbf{Q}^T, \tag{4.7}$$

where $\mathbf{B}$ is the regression coefficient matrix [57].

In practical implementations of PLSR, matrix decomposition and regression are performed simultaneously. Unlike PCA, in which the principal components are extracted solely by maximizing the variance of the input data, PLSR determines latent components by iteratively exchanging the score vectors of $\mathbf{X}$ and $\mathbf{Y}$. This procedure ensures that the extracted components of $\mathbf{X}$ are directly related to the target variables. Consequently, the PLSR algorithm seeks a balance between maximizing the variance captured by each latent component and maximizing its correlation with the target matrix. The former allows the extraction of the most relevant information from the spectral data, while the latter ensures optimal exploitation of the linear relationship between spectral variables and the target concentrations [57].

[1] PCA maximizes the explained variance in X, while PLSR maximizes the covariance between X and Y.

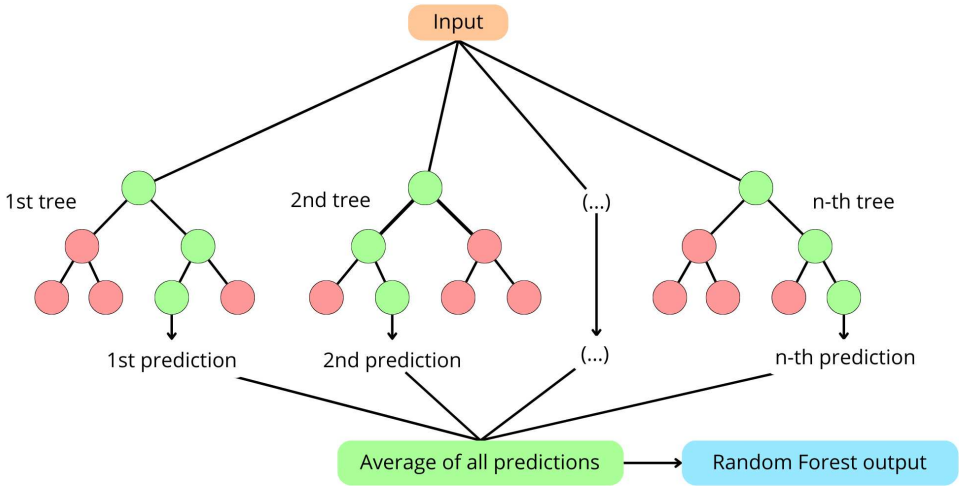
In the context of this study, PLSR was performed computationally using the Python library `scikit-learn` version 1.4.2 [47]. The main argument to be defined when using the `PLSRegression` is the number of components.

4.2.2 Random Forest Regression

Random Forest Regression (RFR) is a non-parametric, ensemble-based machine learning method widely used for regression problems involving complex and potentially non-linear relationships between variables. The method is based on the construction of numerous decision trees, each trained on a different subset of the original dataset generated through bootstrap sampling [58].

In RFR, each decision tree provides an independent prediction of the target variable, and the final model output is obtained by averaging the predictions of all trees in the ensemble, as illustrated in Figure 22. This aggregation process reduces variance and improves generalization, making the method robust against overfitting, particularly in high-dimensional or noisy datasets.

Figure 22 – Schematics of an RFR. Each decision tree in composed by branches containing different nodes. The green-colored nodes indicate step chosen by bootstrapping.



Source: self-elaboration.

The parameters that define a RFR are the number of trees, the maximum depth at which each tree can grow, the minimum splits in each a node, the amount of features to consider while splitting, and the bootstrapping employed in the branches selection.

Unlike linear regression-based methods, such as PLSR, RFR does not assume a predefined functional relationship between the input variables and the target values. As a result, RFR is well suited for modeling non-linear dependencies and complex interactions

among variables, which are commonly encountered in spectrometric and environmental datasets.

Additionally, RFR provides an estimate of variable importance based on the contribution of each input variable to the reduction of prediction error across the ensemble, allowing the identification of the most relevant spectral features influencing the regression results. In the context of this work, RFR was performed computationally using the Python library `scikit-learn` version 1.4.2 [47].

4.3 Calibration Transfer

A common occurrence when performing spectral analysis is that each spectrometer must have its own calibration to provide quantitative results. Calibration in GRS and XRF mostly relies on the availability of Certified Reference Materials (CRMs), which are often scarce, expensive, and limited in matrix compatibility. Moreover, differences in instrument response, detector resolution, or environmental conditions may prevent a calibration built on one instrument from being directly applied to another. Consequently, transferring calibration models between instruments becomes a relevant issue in analytical spectrometry.

To deal with this matter, an increasing field of chemometrics is Calibration Transfer (CT). By definition, these methods consist of analytical strategies or mathematical techniques used to apply the calibration model developed for one instrument (primary) to other instruments (secondary), while maintaining statistical accuracy and precision [59]. CT methods are generally grouped into three main methodologies [57]:

Transfer in spectra: this approach establishes a functional relationship between the target and source instruments by means of a mathematical transformation applied directly to the spectra, so that the converted spectra from the target instrument (secondary) mimic those of the source (primary);

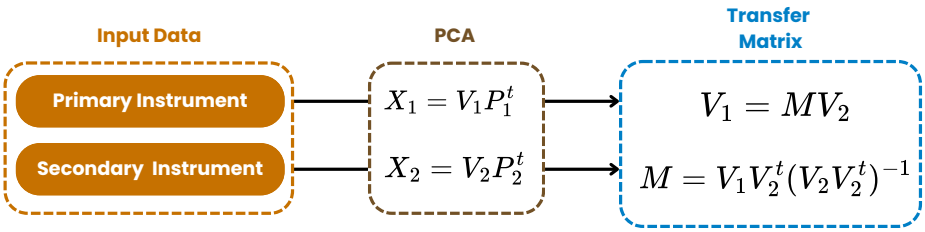
Regression coefficient conversion: in this case, regression models are built independently on both devices, and a transfer function is determined to transform the regression coefficients from the target (secondary) model into those of the source (primary); and

Robust calibration: this strategy involves advanced preprocessing, variable selection or wavelength screening algorithms, or the deliberate expansion of the calibration set by including spectra collected under different conditions or from multiple instruments.

The most commonly employed methodologies are transfer in spectra and robust calibration. In this work, the method adopted is Target Transformation Factor Analysis

(TTFA), which belongs to the first group. TTFA is a PCA-based technique that reduces the dimensionality of the source and target datasets and determines the correlation matrix that maps one latent space onto the other [60]. The main steps of the TTFA procedure are presented in Figure 23.

Figure 23 – Main steps of Target Transformation Factor Analysis.



Source: self-elaboration.

In this method, the data from both instruments are decomposed into score and loading matrices through PCA and then combined to obtain the transfer matrix ($\mathbf{M}$), which relates the scores from both inputs. Once $\mathbf{M}$ is determined, the spectra from the primary instrument can be converted into the secondary domain using Equation (4.8) [60]. In this expression, $\mathbf{V}_i$ and $\mathbf{P}_i$ denote the scores and loadings obtained from the input matrix $\mathbf{X}_i$, while the subscripts 1 and 2 correspond to the primary and secondary instruments, respectively.

$$\mathbf{X}_2^P = \mathbf{X}_1 \mathbf{P}_1 \mathbf{M}^+ \mathbf{P}_2^t \tag{4.8}$$

4.4 Spatial Interpolation

In agricultural and environmental studies, situations often arise where spatially continuous data are required for management and decision-making. Field-obtained data — whether collected for laboratory analysis or measured directly in situ — originate from discrete sampling points and therefore do not provide a clear and continuous spatial distribution [61]. To overcome this limitation, geostatistical and spatial interpolation techniques are commonly employed [61–63].

Spatial interpolation involves predicting the values of an attribute at non-sampled locations based on measurements performed at sampled points within the same area. The point-based results are transformed into continuous surfaces, allowing the spatial distribution of different quantities to be compared, interpreted, and used to support decision-making [62]. The core assumption underlying spatial interpolation is that, on average, the value of a quantity at a given location is more similar to values at nearby locations than to those farther away [61, 62].

The simplest spatial interpolation technique is the Nearest Neighbor (NN) method, in which the value of an un-sampled location is assigned as that of the nearest sampled point within the domain. This approach disregards the influence of surrounding observations, typically resulting in a blocky or pixelated appearance of the interpolated surface. Despite its simplicity, the NN method is still occasionally employed due to its computational efficiency and its suitability for categorical or discrete variables, where intermediate values have no physical meaning [64].

The Inverse Distance Weighting (IDW) method is among the most widely used techniques for spatial interpolation [65]. It assumes that the value of a variable, z, at an un-sampled location is a weighted average of the values at surrounding sampled points, with weights inversely proportional to their distances from the target location [62,65]. The estimated value, $\hat{z}(s_0)$, at the un-sampled point s_0 is obtained according to Equation (4.9).

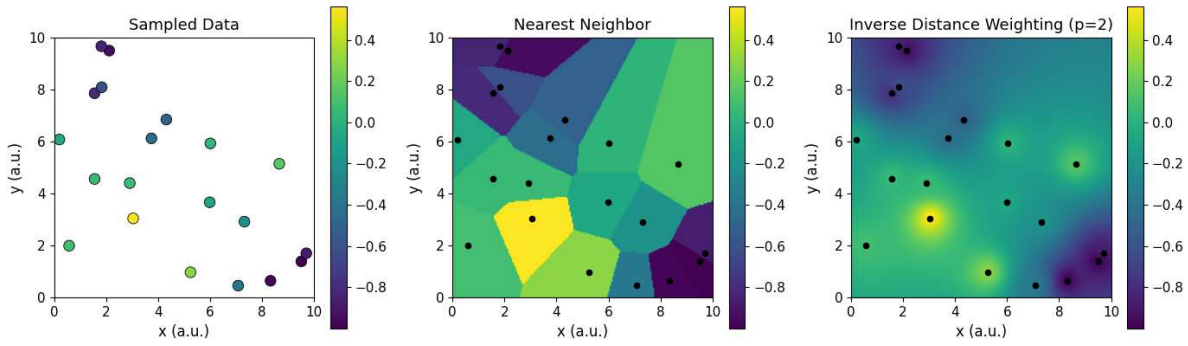
$$\hat{z}(s_0) = \sum_{i=1}^{n} \psi_i z(s_i) \tag{4.9}$$

Here, n denotes the number of sampled points within the interpolation domain, and ψ_i represents the weight assigned to each neighboring observation $z(s_i)$. These weights are determined using Equation (4.10), where d_{i0} is the distance between the sampled location s_i and the prediction point s_0, and p is a user-defined power parameter that controls the influence of distance on the interpolation.

$$\psi_i = \frac{d_{i0}^{-p}}{\sum_{i=1}^{n} d_{i0}^{-p}} \tag{4.10}$$

Larger values of p assign greater importance to closer observations, resulting in smoother yet less generalized surfaces, while smaller p values allow distant points to exert more influence [62]. Figure 24 illustrates the differences between the interpolated surfaces obtained using the NN and IDW methods for a set of randomly generated data points ($n = 20$). While the NN yields a blocky, discontinuous surface due to the absence of smoothing, the IDW method produces a continuous gradient where nearby points exert greater influence on the estimated values. For this reason, IDW was adopted in this work to generate spatially continuous surfaces. In this work, spatial interpolation was performed using the function `interpolate.griddata` from the Python library `scipy`, version 1.10.1 [66].

Figure 24 – Comparison between NN and IDW interpolation methods for a set of randomly generated data points ($n = 20$). The black dots in the interpolated surfaces indicate the sampled locations.



Source: self-elaboration.

5 LITERATURE REVIEW

Soil erosion and sedimentation occur naturally as a result of environmental processes such as rainfall and wind, and they play an important role in landscape formation. However, the intensity of these processes can be substantially amplified by human activities [11, 67, 68]. Soil health and maintenance are fundamental for preserving ecosystem biodiversity, regulating climate, sustaining food production, and ensuring water quality. Recently, the United Nations Educational, Scientific and Cultural Organization (UNESCO) warned that 75% of the Earth's soils are degraded, affecting 3.2 billion people and potentially rising to 90% by 2050 [69]. At the same time, global agricultural production is projected to increase by 50% from 2012 to 2050 in response to growing population demands [70].

In the early 2000s, the global economic costs of soil erosion impacts on agricultural lands - both on-site and off-site - were estimated at approximately US$ 400 billion per year [68, 71]. These economic and social consequences emphasized the need for reliable quantitative data on soil loss worldwide [8, 68]. Since the 1960s, nuclear techniques have been increasingly applied to evaluate, quantify, and model soil degradation over different timescales by tracing gamma-emitting radionuclides naturally present or deposited in soils, including cosmogenic ^{7}Be, fallout ^{137}Cs and $^{239+240}$Pu, and natural ^{210}Pb, among others [4, 68, 72].

Pioneering work conducted in 1965 by Rogowski and Tamura [73] demonstrated the feasibility of using ^{137}Cs as a tracer for soil redistribution. Since then, the ^{137}Cs fallout technique has been widely employed across diverse climatic, geomorphological, and land-use contexts. By the early 2000s, Zapata [68] reported more than 2,500 publications utilizing ^{137}Cs for soil-erosion and sedimentation studies, and subsequent bibliometric assessments indicate that applications of this technique have continued to expand significantly.

In 1999, Blake and Walling [74] demonstrated the potential of using ^{7}Be as a tracer for short-term soil redistribution, highlighting its main advantage relative to ^{137}Cs: its ability to capture processes occurring over timescales of weeks to months. Mabit et al. [4] further emphasized the usefulness of ^{7}Be for evaluating the effectiveness of recent soil-conservation practices [38]. They also noted two practical limitations: the technique is often restricted to bare soil surfaces due to the spatial heterogeneity introduced by vegetation cover, and the characteristic gamma emission of ^{7}Be requires high-resolution detectors for reliable quantification.

Schuller and Walling [75] demonstrated that ^{137}Cs measurements can be used effec-

tively to quantify changes in soil-redistribution rates following a transition from conventional mechanized tillage to no-tillage systems, employing both standard and simplified approaches. Complementary investigations by the same authors showed that ^{7}Be is an appropriate tracer for documenting short-term soil redistribution associated with extreme rainfall events and residue-burning practices in no-till systems, and is also suitable for assessing immediate post-harvest soil loss in forest environments. Together, ^{7}Be and ^{137}Cs provide complementary temporal windows—weeks to months for ^{7}Be and decades for ^{137}Cs—allowing a more integrated characterization of soil-redistribution processes.

In alpine environments, Meusburger et al. [76] demonstrated that ^{137}Cs serves as a robust long-term integrator of soil-erosion processes. Their study further showed that stable carbon and oxygen isotopes provide valuable complementary indicators, capable of distinguishing long- from short-term erosive processes and revealing soil organic carbon (SOC) degradation during detachment and transport.

Findings from Schuller and Walling [75] and Meusburger et al. [76] collectively indicate that combining radionuclide-derived erosion estimates with information obtained from complementary analytical methods yields a more comprehensive assessment of soil condition and management. In this direction, Castilhos et al. [77] combined XRF and GRS measurements for topographic assessments in agricultural soils and demonstrated that the integration of these techniques can provide valuable insights into spatial soil properties and site characteristics.

Hancock et al. [78] examined soil redistribution in Australia between 2006 and 2014 using ^{137}Cs and identified a positive relationship between SOC content and redistribution patterns. Their results indicated that variations in SOC were influenced by both elevation and redistribution processes. The authors also emphasized the need for expanding similar investigations using ^{7}Be and ^{210}Pb to enhance temporal and methodological coverage.

Beyond radionuclide tracing, spectroscopic techniques such as MIRS and XRF have been widely applied to characterize soil fertility parameters, aiming to streamline soil-health evaluation and support agricultural management [50,79–84]. Lelago and Bibiso [85] showed that MIR spectra combined with PLSR models can accurately predict a range of soil properties, including exchangeable Ca, CEC, OC, and pH. More recently, Santos et al. [51] integrated XRF and vis-NIR data using p-ComDim and reported improved predictive performance relative to models developed with single-technique datasets.

In summary, previous studies demonstrate that fallout radionuclides and complementary spectrometric methods provide distinct yet synergistic insights into soil erosion, redistribution dynamics, and soil property variation. Important gaps remain, however, regarding the optimization of GRS-based approaches, the broader application of ^{7}Be using low-resolution detectors, and the integration of distinct spectroscopic datasets. Motivated

by these gaps, the present work focuses on improving GRS-derived erosion estimates through multivariate analysis, establishing a practical methodology for short-term assessments using ^{7}Be, and evaluating soil redistribution across agricultural areas. Additionally, by combining data from GRS, XRF, and MIRS, this study seeks to enhance predictions of soil fertility parameters and to better understand how these complementary techniques can jointly support soil-erosion assessment and land-management strategies.

6 METHODOLOGICAL DEVELOPMENTS IN GAMMA-RAY SPECTROMETRY

6.1 Spectral Deconvolution for Beryllium-7 Detection

As discussed in Section 3.2.1, ^{7}Be is a cosmogenic radionuclide continuously produced in the stratosphere and troposphere, with a half-life of approximately 53 days. It decays to the ground state of ^{7}Li through EC, followed by the emission of a photon with energy of 477.6 keV. Most semiconductor detectors have sufficient energy resolution to readily detect this emission and determine the activity concentration of the radionuclide.

As mentioned in Sections 2.3.1 and 2.3.2, detector resolution is closely associated with its manufacturing technology and cost, which may represent a limitation for routine or large-scale environmental studies. HPGe detectors, despite their excellent resolution and efficiency, require cryogenic cooling and involve elevated acquisition and maintenance costs. In contrast, NaI(Tl) scintillation detectors are considerably more affordable, robust, and easier to operate, yet present a lower energy resolution, which complicates the analysis of spectra with overlapping peaks or high background levels.

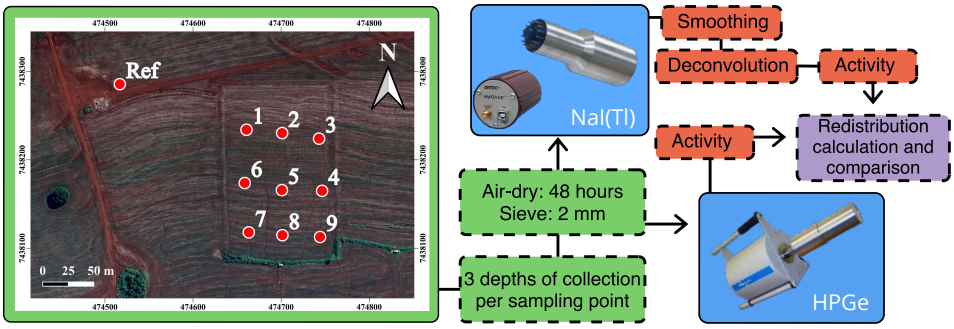
When measuring topsoil samples with low-resolution detectors, the photon emission from ^{7}Be at 477.6 keV may overlap with nearby peaks originating from natural radionuclides and background processes. Among the naturally occurring radionuclides present in soils, ^{228}Ac (463.0 keV) and ^{208}Tl (510.7 keV), both daughters of the ^{232}Th decay chain (Appendix A.1), emit gamma rays with energies close to that of ^{7}Be. In addition, photons with energies above 1.022 MeV can undergo pair production in the detector or surrounding materials, resulting in the emission of annihilation photons at 511.0 keV [19]. The combined presence of these emissions complicates the identification and quantification of ^{7}Be in environmental samples when using low-resolution detectors.

To overcome these limitations, this study proposed a spectral deconvolution methodology for NaI(Tl) spectra aimed at quantifying radionuclides with closely spaced emission lines, and compared the obtained results with those measured using a HPGe detector. In addition, the short-term soil redistribution estimates derived from ^{7}Be inventories obtained with both detector systems were evaluated and compared. The methodology and results of this study were published in the peer-reviewed journal Radiation Physics and Chemistry [86].

6.1.1 Sampling and Preparation

The workflow of this study, from sample collections to data analysis is presented in Figure 25. A detailed description of each step is described in the following subsections.

Figure 25 – Workflow of the employed methodology for spectral deconvolution.



Source: self-elaboration.

The soil samples were collected following the recommendations outlined by IAEA [1, 87]. The study site was situated on a hillside within an agricultural zone located in the Ribeirão Vermelho basin, Cambé municipality, Paraná State, Brazil. This area falls within the latitudinal range of 23°9′59.70″ to 23°9′50.14″ S and the longitudinal range of 51°14′42.27″ to 51°14′56.87″ W. The soil types correspond to "Latossolo Vermelho Eutroférrico Típico" (Rhodic Ferralsol) and "Nitossolo Vermelho Eutroférrico Latossólico" (Rhodic Nitosol) as classified by the Brazilian Agricultural Research Corporation (Embrapa) and the Food and Agriculture Organization (FAO) [88, 89].

Nine sampling points were selected following the multiple transect approach [1]. These points were equally divided into three sections in the parcel: upper, middle, and bottom. At each sampling point, three samples were collected following a depth-incremental technique. The samples ranged from 0 to 3 cm in depth with a 1 cm step. Thus, the method allows the evaluation of soil redistribution effects over a few months on the hill slope. A reference sampling point was also chosen near the mega-parcel, in a plain non-eroded area, where four samples were collected. The fourth sample was collected at approximately 5 cm depth to obtain a sample with no ^{7}Be presence and was used as a blank sample.

After collection, all samples were dried out for a minimum of 48 hours and subsequently sieved through a 2 mm mesh. The sieved portion of the samples was deposited into plastic cups and stored until the measurements could be performed. The measurements were performed for 86400 seconds of live time, employing the detectors described in Subsections 2.3.1 and 2.3.2.

6.1.2 Spectral Deconvolution

The 477.6 keV ^{7}Be peak is contained in a broad peak with contributions of secondary lines from ^{228}Ac (463.0 keV) and ^{208}Tl (510.7 keV). To circumvent this situation, spectral deconvolution was performed to calculate ^{7}Be activity in the spectra obtained with the NaI(Tl) detectors. The concentration activities were calculated using the 911.2

and 968.9 keV emissions from ^{228}Ac, and 2614.5 keV emissions for ^{208}Tl. Afterward, the corresponding area for the emissions of 463.0 and 510.7 keV were estimated using the obtained activity value and isolating the net area in Equation (2.15).

The positron annihilation contribution is caused by every photon emission with energies of 1.022 MeV and above. In soil samples, this is mainly due to ^{214}Bi, ^{40}K, and ^{208}Tl. Since higher energies are more likely to interact with matter by pair production, the intensity of the effect was estimated using ^{208}Tl emissions. In all measurements performed with the HPGe detector, the calculated ^{208}Tl activity using the 510.7 keV peak was 1.93 times greater than the ones calculated by other emissions.

Based on these results, it was assumed that a similar behavior would happen in the NaI(Tl) detector volume. The area corresponding to ^{7}Be emission was then calculated using Equation (6.1) and varying the proportionality constant, ξ, until the best correspondence with the activity obtained using the HPGe detector was achieved. In the Equation, each $\mathcal{A}$ corresponds to a peak's area, indicated as the subindex, and $\xi = 1.70$ is the proportionality constant to account for the effects of pair production.

$$\mathcal{A}_{^{7}Be} = \mathcal{A}_{\text{peak}} - \mathcal{A}_{^{228}Ac} - \xi\mathcal{A}_{^{208}Tl} \tag{6.1}$$

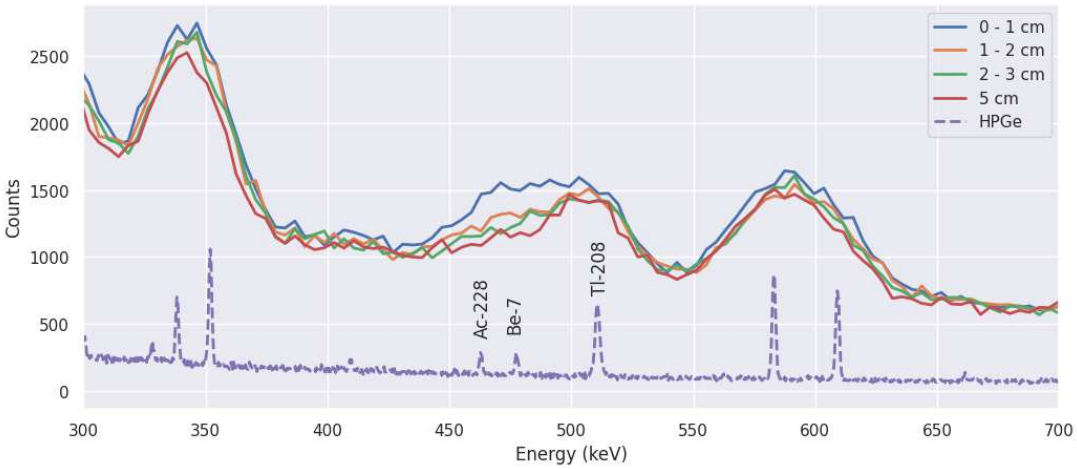
6.1.3 Activity and Soil Redistribution Calculations

The detected isotopes activity concentration and respective MDAs were calculated using Equation (2.18). Given the long measurement times and amount of samples,^{7}Be activity was corrected for the decay using the radioactive decay equation between the collection and measurement dates. The activity values below the MDA were not corrected. Soil redistribution was calculated using Equations (3.3) through (3.8).

6.1.4 Detectors Comparison

When comparing spectra obtained from the same sample using both detectors, it is observed that the ^{7}Be peak is unresolved with signals from ^{228}Ac, ^{208}Tl, and positron annihilation effects. This is depicted in Figure 26. The four continuous lines in the figure represent NaI(Tl)-obtained spectra for samples collected at different depths. The changes in the shape of the peak at the center follow the collection depth. The dashed purple line represents a spectrum obtained for a soil sample using an HPGe detector, and the radionuclides responsible for each emission in the peak of interest are indicated. Through the spectrum comparison, it becomes evident that ^{7}Be plays a significant role in determining the shape of the combined peak. Moreover, ^{7}Be is the only radionuclide among these three that has significant concentration variation for such shallow depths.

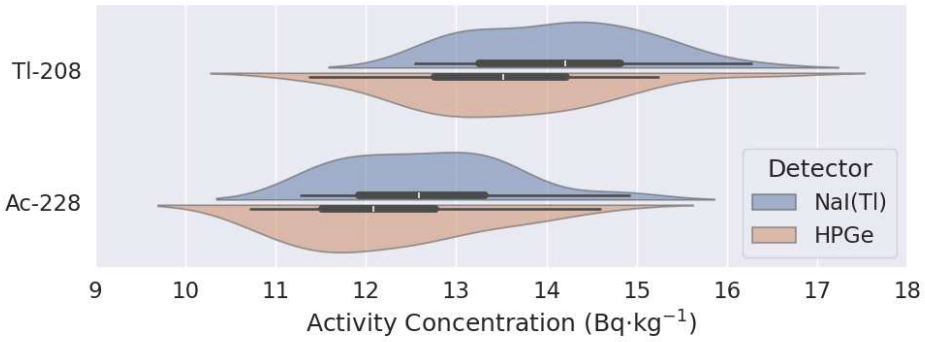
Figure 26 – Comparison of ^{7}Be influence in spectra obtained with HPGe and NaI(Tl) detectors at different depths. The four continuous lines represent the NaI(Tl) spectra obtained from a sample at different depths, while the dashed line is an HPGe spectrum. The ^{228}Ac (463.0 keV), ^{7}Be (477.6 keV) and ^{208}Tl (510.7 keV) emissions are identified.



Source: self-elaboration.

The concentration activities, in Bq·kg^{-1}, of ^{208}Tl and ^{228}Ac were calculated for the 31 samples measured with both detectors and the results are shown in Figure 27. The violin plot allows the examination of the calculated activity distribution. Each row indicates a radionuclide, while the X-axis is scaled according to the measured concentration activities and the different colors correspond to the employed detector. The similarity between concentration ranges and the shape of the distribution indicates that the performed calibration and efficiency calculations for the two detectors are equivalent.

Figure 27 – Violin plots displaying the measured activity concentrations (Bq·kg^{-1}) of ^{208}Tl and ^{228}Ac across the 31 studied soil samples. The congruence between both distributions suggests equivalency in the calibration and efficiency calculations for the two detectors.



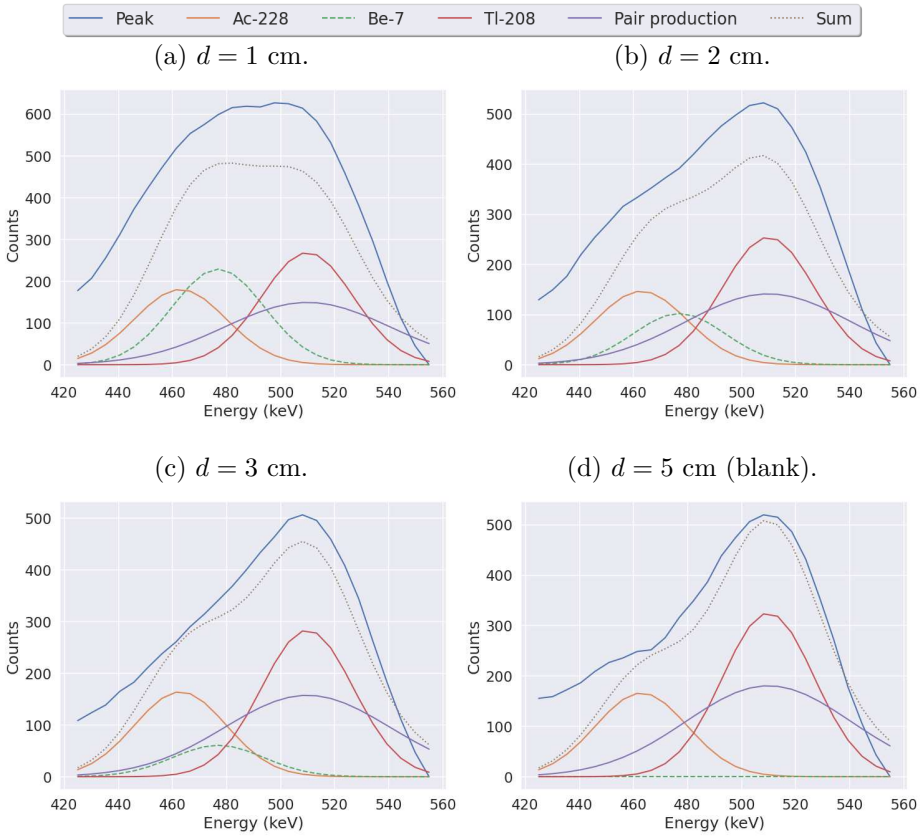
Source: self-elaboration.

The slight changes verified in Figure 27 are caused mainly by the different efficiencies and resolution of the employed detectors. Given its better resolution, the HPGe detector yields more precise measurements. Nonetheless, the equivalence between the detectors ensures the suitability of performing the proposed deconvolution methodology to evaluate the ^{7}Be activity concentrations in soil samples.

6.1.5 Deconvolution and Soil Redistribution

Performing the spectral deconvolution with the described methodology provided good results, as shown in Figure 28. The four selected samples correspond to the reference sampling point. In the four spectra, the blue line represents the original peak after performing smoothing and background subtraction; the yellow line is the ^{228}Ac peak; the green dashed line is the calculated ^{7}Be peak; the red line is the ^{208}Tl peak; the purple line is the estimated positron annihilation contribution; and the gray dotted line is the sum of the four calculated peaks. The peaks were plotted assuming Gaussian distributions and using the FWHM for each emission as verified in the detector resolution curve.

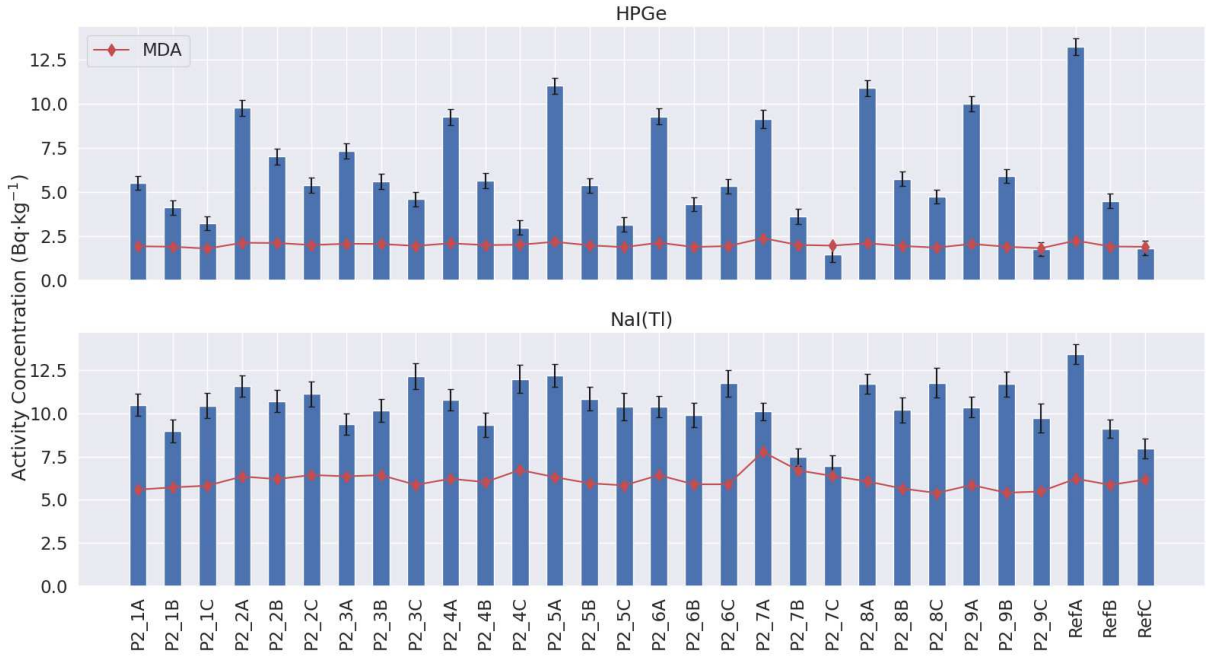
Figure 28 – Resulting peak deconvolution from four depths (1–5 cm) at one sampling point. Blue line: measured peak, yellow/red: estimated ^{228}Ac/^{208}Tl contributions (911.2/968.9 keV, and 2614.5 keV), purple: estimated positron annihilation, green dashed: calculated ^{7}Be, gray dotted: sum of all peaks.



Source: self-elaboration.

The compared ^{7}Be concentration activities for both detectors are in Figure 29, along with their respective uncertainties and MDA values. While there are discrepancies in the calculated values between the detectors, the observed trends among the samples in each dataset remain consistent. These variances primarily stem from disparities in detector resolution and the positron annihilation contribution in the NaI(Tl) detector. The smaller MDA for the HPGe detector evidences its superior resolution.

Figure 29 – Measured ^{7}Be activity concentration values, uncertainties and MDA for all samples. The range of calculated values is the same for both detectors, and the trends between them are similar.



Source: self-elaboration.

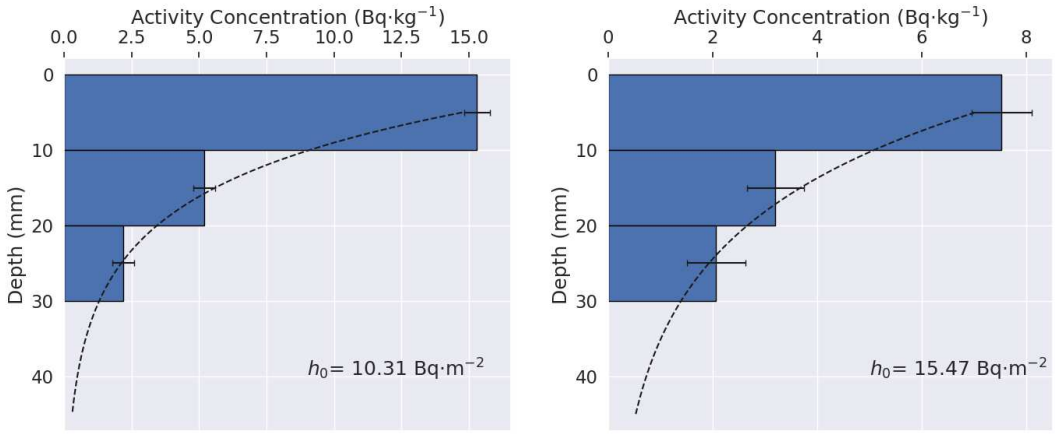
Applying Equation 6.1 to the blank sample (collected at 5 cm depth) yields a non-zero area value. This area was converted into activity concentration and treated as a "virtual" ^{7}Be activity (5.94 Bq·kg^{-1}), which was subtracted from the other calculated activities. This adjustment aimed to mitigate noise and unwanted contributions caused by other nuclear phenomena within the convoluted peak. The decision to employ the virtual activity approach instead of directly subtracting the areas was made to account for various factors influencing the spectra, including variations caused by differing sample masses, spectral overlaps, and background contributions.

Figure 30 illustrates the vertical distribution profile of ^{7}Be at the reference sampling point for both detectors. The obtained z_0 values, calculated using Equation (3.4), for the HPGe and NaI(Tl) detectors were 10.31 ± 0.20 and 15.47 ± 0.10 Bq·m^{-2}, respectively. The z_0 calculated for the NaI(Tl) was estimated after subtracting the virtual activity concentration derived from the blank sample analysis. The discrepancy in the calculated z_0

values is attributed to the differences in the obtained ^{7}Be activity concentrations, which are influenced by the detector resolution and the positron annihilation contribution in the NaI(Tl) spectra.

Figure 30 – Comparison of the ^{7}Be distribution profile obtained with the reference samples for the HPGe and NaI(Tl) detectors. The obtained mass relaxation depths, z_0, are presented.

(a) HPGe. (b) NaI(Tl).



Source: self-elaboration.

With the activity concentrations of ^{7}Be calculated for the two detectors, the soil redistribution calculated in the mega parcel was determined using Equations (3.3) through (3.8). The redistribution results are displayed in Figure 31, and the calculated net deposition was 36.45±18.95 kg·m^{-2} for the NaI(Tl) detector and 26.40±22.35 kg·m^{-2} for the HPGe detector during the evaluated period. The analysis of soil redistribution within the hill slope agricultural mega parcel, employing HPGe and NaI(Tl) detectors, reveals a consistent trend across all sampling points. Mostly, changes in intensity are observed and are due to the differences in resolution and efficiency.

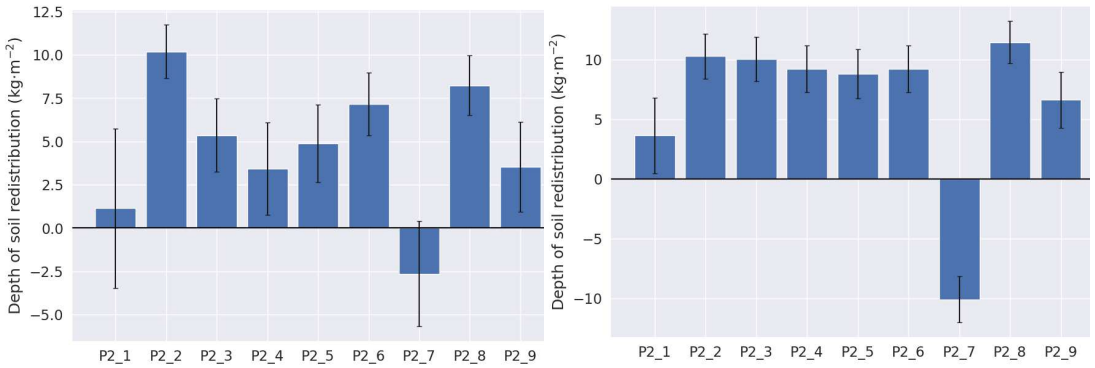
The depth of soil redistribution calculated for the NaI(Tl) results lack precision when compared to the HPGe results, as evidenced by the homogenous results (Figure 31). This is mainly due to the lower resolution of the NaI(Tl) detector, which leads to a less accurate estimation of ^{7}Be activity and consequently, soil redistribution. However, the overall trend of soil deposition across the sampling points is consistent between both detectors, indicating that the NaI(Tl) can still provide valuable insights into soil redistribution patterns, albeit with less precision than the HPGe detector.

Soil deposition was verified for almost the entire length of the area, except at the sampling point P2_7, where erosion was detected. Positive values of soil redistribution depth (z) observed in the uppermost points suggest contributions from regions above the parcel, likely due to sediment deposition or lateral soil movement. As depicted in Figure 25

the parcel is located on a hill slope below a dirt road, this position makes the sampling region prone to sediment deposition originating from the road and upper areas.

Figure 31 – Calculated soil redistribution based on the ^{7}Be activity for the HPGe and NaI(Tl) detectors. Both sets of results show the same behavior in the mega parcel.

(a) HPGe. (b) NaI(Tl).



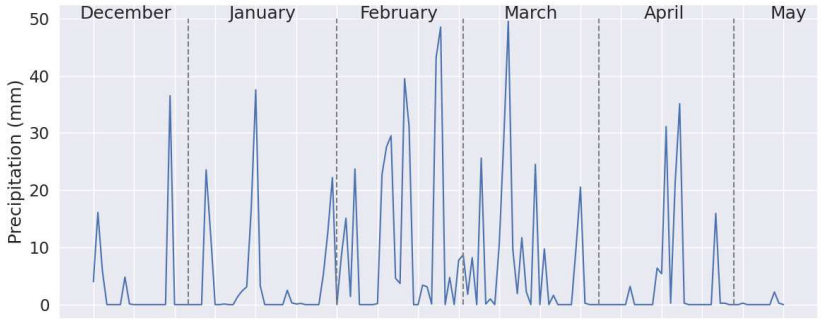
Source: self-elaboration.

The propagated uncertainty from sampling points P2_1 and P2_7 in Figure 31a makes the calculated z for these points statistically unreliable. Such large uncertainties arise from the measured spectra, and as depicted in Figure 29, these two points have one or more sub-samples whose concentration activity value is near or below the MDA.

Overall, these results points toward a net soil deposition in the evaluated area during the six months prior to the sampling date, with the exception of one sampling point where erosion was detected. The observed soil redistribution patterns are consistent with the expected behavior for a hill slope agricultural area, where sediment deposition is likely to occur in lower regions due to runoff, soil movement from upper areas, and from external sources.

The soil movement observed aligns closely with the rainfall patterns recorded in the six months leading up to the sampling date, as depicted in Figure 32. Rainfall data, collected by IDR-Paraná at 15-minute intervals, were aggregated to provide a comprehensive view of daily precipitation trends, as illustrated in the figure. The six-month period was chosen based on the IAEA guidelines to evaluate soil redistribution using ^{7}Be. The semester before the sampling date encompasses the wet season in southern Brazil, characterized by substantial rainfall, particularly notable in February and March. These periods of intense precipitation are likely significant contributors to the soil redistribution dynamics observed through Beryllium-7 detection.

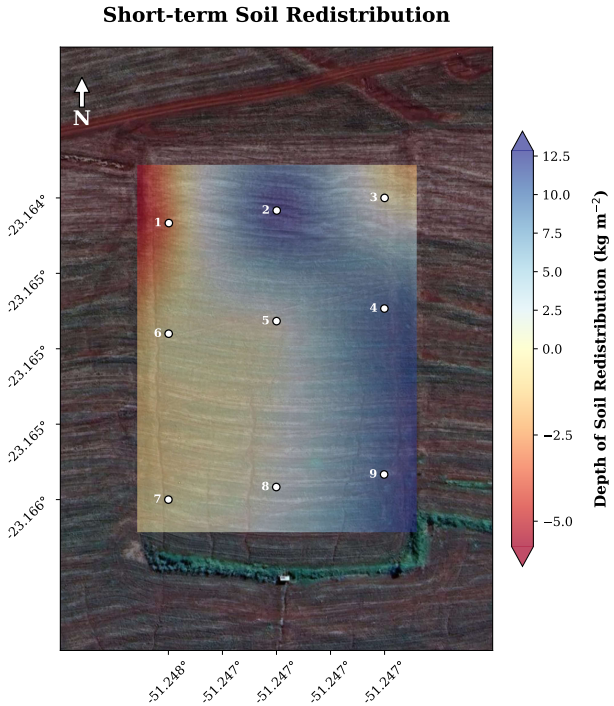
Figure 32 – Rainfall data, in millimeters, obtained for the megaparcel region from December/2022 to May/2023. The sampling was performed on 11th May.



Source: self-elaboration.

To facilitate the visualization of the spatial patterns of soil redistribution within the studied area, a spatial interpolation procedure was applied to the calculated redistribution values. The interpolation was performed using the IDW method, as described in Section 4.4, which estimates values at unsampled locations based on the proximity and magnitude of neighboring measurements. Figure 33 presents the resulting interpolated map for the agricultural plot, where the white markers indicate the sampling points and the color scale represents the intensity of soil redistribution across the area.

Figure 33 – Soil redistribution intensity over the studied agricultural area obtained by IDW interpolation. The color scale represents the magnitude of redistribution.



Source: self-elaboration.

6.1.6 Conclusions

These findings confirm the possibility of utilizing NaI(Tl) detectors as a substitute for HPGe detectors in assessing ^{7}Be in soil samples. Employing mathematical techniques such as Savitzky-Golay smoothing and spectral deconvolution significantly improved the results with the scintillation detector, approaching the quality of those achieved with the more precise HPGe detector. This substitution offers the advantage of greater durability and lower maintenance costs associated with the NaI(Tl) detectors when compared to HPGe detectors. Moreover, this procedure facilitates the dissemination of GRS applied to evaluate short-term soil erosion, which may lead to a more robust and thorough evaluation of soil management techniques.

6.2 Calibration Transfer for Radionuclide Quantification

Despite the differences in detectors construction and overall properties, HPGe and NaI(Tl) detect and measure the same kind of radiation. In the field of chemometrics, several CT techniques have been developed to improve the comparability of spectral data from different instruments [57]. These methods, widely applied in Infrared Spectrometry (IRS), include TTFA, which uses PCA to reduce spectral dimensionality and derive a transfer matrix that correlates the responses between different detectors [90].

To date, no studies have investigated the application of chemometric methodologies of calibration transfer between different types of detectors for GRS measurements. This gap is significant as it presents a challenge for laboratories relying on different detector types to achieve consistent and accurate radionuclide quantification.

In this context, this research hypothesis is that a reliable conversion between the two detector responses can be achieved. Such a conversion would significantly improve the accuracy of measurements obtained with NaI(Tl) detectors, offering a more accessible alternative to HPGe systems for a variety of environmental applications and further promoting the use of GRS as a soil evaluation technique. Accordingly, the present study applies the TTFA method to a soil dataset measured with both HPGe and NaI(Tl) detectors, with the following objectives: (*i*) calculating a transfer matrix that links the spectral response; and (*ii*) comparing the concentration results of some radionuclides calculated from HPGe detector measurements with those obtained with the transformed NaI(Tl) spectra. The methodology and results of this study were published in the peer-reviewed Journal of Environmental Radioactivity [91].

6.2.1 Methodology

The soil samples employed in this study were obtained in two different locations: (*i*) on a hillside within an agricultural area in the Ribeirão Vermelho basin, situated in the

municipality of Cambé, Paraná State, Brazil ($n = 31$); and (ii) at the State University of Central West, at Guarapuava, Paraná State, Brazil ($n = 10$). All samples were superficial, with a maximum depth of 3 cm, to allow ^{7}Be quantification [87]. Frequent rain events occurred at the second site in the weeks prior to the sampling, which lead to higher ^{7}Be concentrations. Both datasets were combined to enhance the variability of the detected radionuclides and, therefore, improve the model robustness.

After sampling, the soil samples were air-dried for at least 48 hours before being sieved through a 2 mm mesh and sealed into plastic cups with a total volume of 300 ml. The mean mass of the sieved samples was of 320±70 g.

6.2.2 Experimental Setup

The measurements were performed employing HPGe (Subsection 2.3.2) and NaI(Tl) (Subsection 2.3.1) detectors, and conducted by positioning the samples directly on top of the detectors for a live time of 86400 seconds.

Both detectors were energy-calibrated using point-like calibrated sources of ^{137}Cs and ^{60}Co manufacture by IPEN. The calibration curve for soil samples was built using the IAEA-327 [92] certified sample and validated using IAEA-312 [93] and IAEA-444 [94] certified samples. The reference and calculated activity concentrations for validation samples are displayed in Table 6.

Table 6 – Reference and calculated activity concentration values, in Bq/kg, for IAEA-312 and IAEA-444 certified samples after using the IAEA-327 certified samples to build de efficiency curve for the HPGe and NaI(Tl) detectors.

	IAEA-312		**IAEA-444**			
	Ra-226 (Bq/kg)		**Co-60* (Bq/kg)**		**Cs-137* (Bq/kg)**	
	Reference Value	Calculated Value	Reference Value	Calculated Value	Reference Value	Calculated Value
HPGe	269 ± 19	275	10.07 ± 3.18	8.62	48.95 ± 6.95	55.74
NaI(Tl)		263		7.60		48.55

*The activity concentration and uncertainties of this radionuclide were corrected for the time elapsed since the reference material was issued.

6.2.3 Data Management

To perform TTFA with the obtained spectra, each result was converted from their original extension ('.CNF' for the HPGe detector and '.CHN' for the NaI(Tl)) into '.TXT' files using Python 3.10, and all measurements from each detector were combined into a '.CSV' file to facilitate data handling.

Both sets of spectra consisted of 41 files containing 2-columns: energy values and the number of photons measured with the corresponding energy. Given the instrument's characteristics, all '.CNF' files contained 8192 rows, while the '.CHN' files had 1024, corresponding to the channels of the detectors. Therefore, the first step in the data treatment was to interpolate the photon counts in the '.CHN' files to match the energy resolution of the '.CNF' files. This was achieved using a linear interpolation approach, implemented through the `interp1d` function from the SciPy library, version 1.10.1 [66]. The energy values from the '.CNF' files were used as the reference points for the interpolation, ensuring that the photon counts from the '.CNF' files were accurately mapped onto the lower resolution energy grid of the '.CHN' files.

After interpolating the NaI(Tl) spectra to match the HPGe's dimensionality, the Kennard-Stone algorithm was applied to separate the target and source spectra into training and testing dataset (24 and 17 samples, respectively). This was made to create the transfer matrix using the training dataset and validating it with the test dataset.

PCA was then performed using the `PCA` function from the scikit-learn library, version 1.4.2 [47]. After splitting the data into training and test sets, PCA was conducted with $n = 24$ PCs. By evaluating the cumulative explained variance for both sets of spectra, it was found that a plateau close to 100% explained variance was reached with $n = 5$ PCs. Therefore, this number of PCs was chosen for the CT calculations, minimizing the influence of spectroscopic noise in the model.

The transformation matrix calculation was performed using the `linalg.lstsq` function from the NumPy library, version 1.23.5 [95]. This function receives as input the scores from the source and the target instruments, respectively. By using the least-squares method, it finds the matrix that maps the scores from the source instrument to the target instrument. The result in a $n \times n$ matrix.

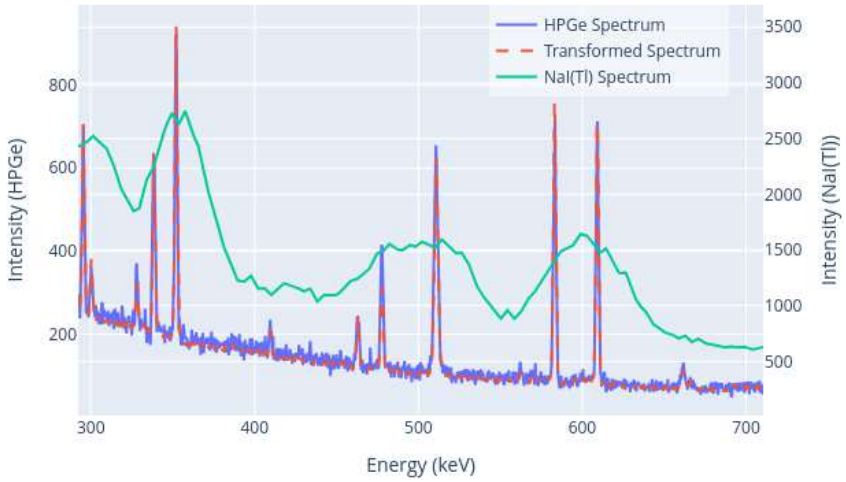
After calculating the transfer matrix $\mathbf{M}$, the transformation was performed by multiplying the PCA scores from the NaI(Tl) detector by $\mathbf{M}$ and performing the inverse PCA transformation (using `PCA.inverse_transform` from scikit-learn) to reconstruct the spectra. The performance of the calibration transfer was evaluated by calculating its root mean-squared error (RMSE), Mean Absolute Error (MAE) and the Pearson correlation coefficient (ρ).

To further study the CT method suitability in GRS analysis, the concentration activities, in Bq/kg, for ^{7}Be, ^{40}K, ^{137}Cs, ^{232}Th, and ^{238}U were calculated using the original HPGe and the NaI(Tl)-transformed spectra and the results were compared.

6.2.4 Calibration Transfer

Figure 34 shows an example of the transformation. Both HPGe and NaI(Tl) spectra obtained for a sample are displayed in blue and green, respectively. The resulting spectrum after performing the transformation is displayed as the red dashed line. In the plot, the region between 300 keV and 700 keV is zoomed in to provide better visualization.

Figure 34 – Spectra for a representative sample. The blue line represents the HPGe spectrum, the green line the NaI(Tl) spectrum, and the red dashed line the reconstructed spectrum. The left y-axis shows counts for the HPGe and transformed spectra, while the right y-axis shows NaI(Tl) counts. The plot is zoomed in as to provide better visualization.



Source: self-elaboration.

To perform the calibration transfer, both spectral datasets were divided into training and test sets to calculate and validate the transfer matrix. Table 7 summarizes the evaluation metrics for the transformation in both datasets. RMSE provides an overall measure of the difference between the original and transformed spectra. The low RMSE values (10.83 for the training set and 15.57 for the test set) indicate strong performance of the transformation, with minimal errors.

MAE provides an average measure of the errors between the HPGe-measured and NaI(Tl)-transformed spectra. These values are reported in the same units as the spectra, with MAE values of 4.44 for the training set and 5.58 for the test set. This metric indicates that the transfer produces minimal deviations from the target spectra.

Pearson's correlation coefficient measures the linear relationship between the original and transformed spectra, indicating how well the transformation captures the structure of the target spectra. The coefficient ranges from 0 to 1 indicating the strength of the correlation. Since both datasets resulted in $\rho = 0.99$, this suggests an strong correlation, implying that the transformation preserved the essential characteristics of the original

spectra, allowing for a highly accurate reconstruction.

Table 7 – Performance metrics for the training and test datasets.

Dataset	RMSE	MAE	ρ
Training Dataset	10.83	4.44	0.99
Test Dataset	15.57	5.58	0.99

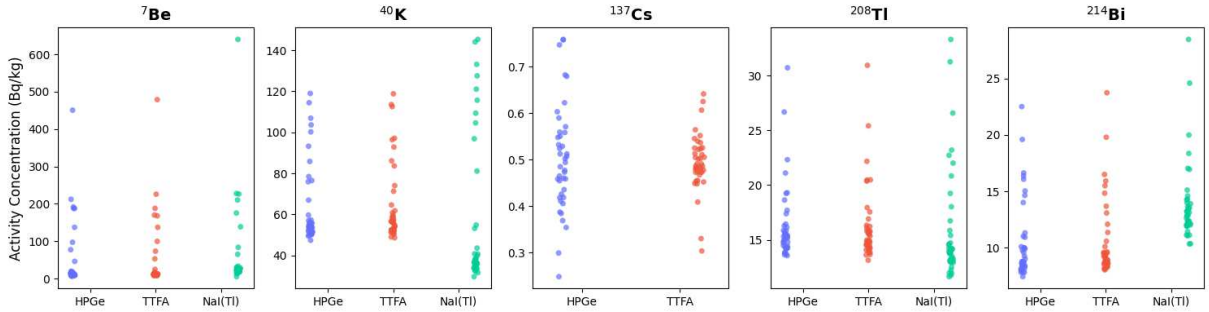
RMSE: Root Mean Squared Error; MAE: Mean Absolute Error; ρ: Pearson correlation coefficient.

6.2.5 Activity Concentration

Prior to performing the transformation, emissions of ^{40}K, ^{232}Th, and ^{238}U were easily identified in both the HPGe and NaI(Tl) spectra. ^{7}Be, albeit identified in the NaI(Tl) spectra, is convoluted with other emissions and its quantification is not so straightforward. ^{137}Cs, on the other hand, is only detectable by the HPGe detector due to the low concentration found in the Southern Hemisphere [1] and the detector's resolution.

The activity concentration of ^{7}Be, ^{40}K, and ^{137}Cs were calculated by their characteristic emissions (477.6 keV, 1460.8 keV, and 661.6 keV, respectively), when possible, and the concentrations of ^{232}Th and ^{238}U are reported as those of their daughters ^{208}Tl (2614.7 keV) and ^{214}Bi (609.31 keV), respectively. Since ^{7}Be is one of the radionuclides of interest, its short half-life (53 days) makes it impossible to wait for the equilibrium of the Th and U series to be achieved while also obtaining precise measurements for ^{7}Be. Figure 35 illustrates the activity concentration distribution of the five radionuclides calculated using the NaI(Tl), HPGe, and transformed spectra.

Figure 35 – Comparison of the activity concentrations of ^{7}Be, ^{40}K, ^{137}Cs, ^{208}Tl, and ^{214}Bi calculated using the NaI(Tl) (green), HPGe (blue) and transformed (red) spectra.



Source: self-elaboration.

The concentrations range for all radionuclides, except ^{137}Cs, is similar for the three types of quantification with NaI(Tl) providing slightly broader ranges than HPGe and TTFA. The descriptive statistics of these results are displayed in Table 8. Results obtained

with the HPGe spectra provides the most precise measurements, as seen by the lowest SD values. On the other hand, NaI(Tl) generally shows the highest variability (largest SD and IQR), indicating that this detector is less reliable for precise quantification. Finally, TTFA consistently yields values very close to HPGe, with slightly higher dispersion but improved accuracy over NaI(Tl).

Table 8 – Mean activity concentration, standard deviation (SD), median, interquartile range (IQR), minimum and maximum values, mean MDA and its SD of the five radionuclides calculated using the HPGe, transformed (TTFA) and NaI(Tl) spectra ($n = 41$ samples). All values are reported in Bq/kg.

	Dataset	**Mean**	**SD**	**Median**	**IQR**	**Min**	**Max**	**MDA**	
								Mean	**SD**
	HPGe	47.52	87.05	11.35	8.64	6.40	450.49	3.33	2.69
^{7}Be	TTFA	47.71	89.49	11.64	7.60	7.59	478.72	3.33	2.72
	NaI(Tl)	60.18	110.12	21.93	10.42	5.09	639.87	6.64	3.11
	HPGe	63.73	19.96	53.97	15.23	47.36	118.99	4.74	1.48
^{40}K	TTFA	64.50	19.30	56.42	12.87	48.62	118.82	4.78	1.50
	NaI(Tl)	56.83	36.74	37.10	20.04	29.64	145.33	9.41	3.60
^{137}Cs	HPGe	0.50	0.12	0.45	0.13	0.25	0.76	0.21	0.05
	TTFA	0.49	0.06	0.30	0.48	0.10	0.64	0.21	0.05
	HPGe	16.36	3.47	15.30	1.87	13.56	30.72	1.53	0.40
^{208}Tl	TTFA	16.26	3.46	14.97	1.80	13.14	30.94	1.53	0.40
	NaI(Tl)	15.93	5.09	13.90	2.75	11.64	33.31	2.05	0.68
	HPGe	10.45	3.55	8.76	2.60	7.45	22.51	0.86	0.25
^{214}Bi	TTFA	10.59	3.45	9.21	2.25	8.05	23.74	0.86	0.25
	NaI(Tl)	13.87	3.54	12.90	1.93	10.34	28.46	1.96	0.64

For ^{7}Be, the results obtained with the HPGe and the transformed spectra produces nearly identical means. Meanwhile, NaI(Tl) provides a higher median value, indicating overestimation, and higher SD and IQR values, which suggests high measurement dispersion. This might be due to the necessity of performing peak deconvolution in the NaI(Tl) spectra to quantify this radionuclide, causing the results to be less accurate.

Overall, NaI(Tl) underestimates the ^{40}K activity concentration values when compared to the HPGe and the transformed spectra. The lower mean and median suggests a systematic bias in this radionuclide detection for the NaI(Tl) system. Similar behavior is observed for ^{214}Bi, but the results obtained with the NaI(Tl) spectra are overestimated in comparison with the other two. In both cases, the application of TTFA brings the results closer to the ones obtained with the HPGe spectra.

Despite producing skewed results, the application of TTFA over NaI(Tl) spectra is capable of producing ^{137}Cs results very similar to the HPGe spectra. For such low concentrations ^{137}Cs can not be detected by NaI(Tl) detectors, indicating a clear advantage

of employing this technique in GRS measurements.

^{208}Tl is the radionuclide whose results are the most consistent between the three spectra types. Nonetheless, NaI(Tl) shows greater measurement variation, as indicated by the larger SD and IQR. Overall, by employing TTFA over the NaI(Tl) spectra to transform it and later on calculate the activity concentrations of the five selected radionuclides provides more accurate results than the direct calculation.

6.2.6 Results Evaluation

To further evaluate the performance of TTFA over the NaI(Tl) spectra, the activity concentration results were compared to the values obtained with the HPGe spectra by calculating the mean difference, p-value, MAE and Mean Bias Error (MBE). The comparisons, in terms of deviation from the HPGe results are summarized in Table 9.

Table 9 – Comparison of the quantification performance for NaI(Tl) and transformed spectra relative to HPGe measurements.

	^{7}Be		^{40}K		^{137}Cs	^{208}Tl		^{214}Bi	
	NaI(Tl)	**TTFA**	**NaI(Tl)**	**TTFA**	**TTFA**	**NaI(Tl)**	**TTFA**	**NaI(Tl)**	**TTFA**
Mean Diff.	12.66	0.19	-6.89	0.78	-0.01	-0.43	-0.10	3.42	0.14
p-value	0.8224	1.0000	0.4712	0.9904	0.6521	0.8806	0.9936	0.0001	0.9827
MAE	42.21	3.75	21.28	3.87	0.07	2.14	0.42	3.75	0.45
MBE	12.65	0.19	-6.89	0.78	-0.01	-0.43	-0.09	3.42	0.14

The mean differences between the NaI(Tl) and TTFA-transformed spectra were generally low, with values close to zero for most radionuclides, indicating minimal deviation between the original and transformed spectra. For instance, the mean difference for ^{7}Be was 12.66 for NaI(Tl) and 0.19 for TTFA, while ^{137}Cs showed a negligible difference of -0.01 for TTFA.

P-values for most of the comparisons were above the significance threshold (0.05), suggesting that the differences between the NaI(Tl) and TTFA spectra were not statistically significant. The only exception is ^{214}Bi, who presented a p-value of 0.0001, indicating a strong statistical difference between NaI(Tl) and TTFA. But, as seen in Table 8, the quantification of this radionuclide using the NaI(Tl) detector was significantly overestimated in comparison with the HPGe.

The MAE and MBE values for each detector also support the effectiveness of the CT method. While NaI(Tl) spectra showed larger MAE values, particularly for ^{7}Be, the TTFA-transformed spectra demonstrated significantly lower MAE values across all radionuclides. Notably, the MAE for ^{7}Be decreased from 42.21 in NaI(Tl) to 3.75 in the TTFA-transformed spectra. The MBE values, which indicate systematic bias, were also

reduced after transformation. These improvements suggest that TTFA provides a robust transformation method for improving GRS results employing NaI(Tl) detectors.

Overall, the TTFA method shows strong potential for improving the quantification accuracy of radionuclides in gamma-ray spectrometry, particularly when using NaI(Tl) detectors. The comparison of results between the two methods demonstrates that TTFA offers a feasible solution for enhancing NaI(Tl) measurements.

6.2.7 Conclusions

The study demonstrated that TTFA is an effective methodology for performing calibration transfer between HPGe and NaI(Tl) detectors for GRS. The low root-mean squared error and mean absolute error indicate that the transformed spectra from the NaI(Tl) detector closely match the original spectra obtained from the HPGe detector. In conclusion the calibration transfer was successful, capturing the essential characteristics of the spectra.

The activity concentration of ^{7}Be, ^{40}K, ^{137}Cs, ^{208}Tl and ^{214}Bi were calculated using the HPGe, NaI(Tl) and the transformed spectra for comparison. In all cases the mean, standard deviation and median values calculated for the transformed spectra are more alike the results obtained with the HPGe spectra than the ones obtained by directly employing the NaI(Tl) spectra. Overall, the use of TTFA also provided results with lower MDA's, indicating that the use of TTFA may enhance the detection capacity of NaI(Tl) measurements. Which is proven by the satisfactory quantification of ^{137}Cs.

Most comparisons presented a p-value > 0.05, indicating that the differences between the values were not statistically significant. Moreover, the Pearson correlation coefficient showed very strong positive correlation for all the measured radionuclides, with the exception of ^{137}Cs.

These results imply that target transformation factor analysis provide accurate and reliable results for radionuclide quantification using NaI(Tl) detectors. Consequently, this method could enhance the accessibility and accuracy of gamma-ray spectrometry measurements, particularly in settings where HPGe detectors are not available, thereby broadening the applicability of this technique.

7 MULTITECHNIQUE APPROACHES FOR SOIL REDISTRIBUTION STUDIES

7.1 Gamma and X-Ray Spectrometries Applied to Evaluate Soil Redistribution

As discussed in Section 3.2.2, the ^{137}Cs methodology is commonly employed for soil erosion evaluation. Soil erosion is characterized by loss of fertile topsoil containing nutrients that are necessary for plants to grow and ensure agricultural productivity. A technique often applied as a quick alternative to the traditional soil analysis is EDXRF [79–82, 84].

In this work, the novelty relies on the combination of ^{137}Cs inventories, measured using GRS, with elemental composition results obtained with EDXRF. Combining data from EDXRF and GRS offers a more detailed perspective on soil physical and chemical characterization [77]. While EDXRF provides information on the soil's elemental composition (such as mineralogy, geochemical and anthropogenic signatures, etc.), GRS complements this knowledge by analyzing and quantifying radioactive isotopes. This integrated approach enables a deeper assessment of erosion dynamics. The multivariate statistical tool employed to perform a simultaneous analysis of EDXRF and GRS was PCA.

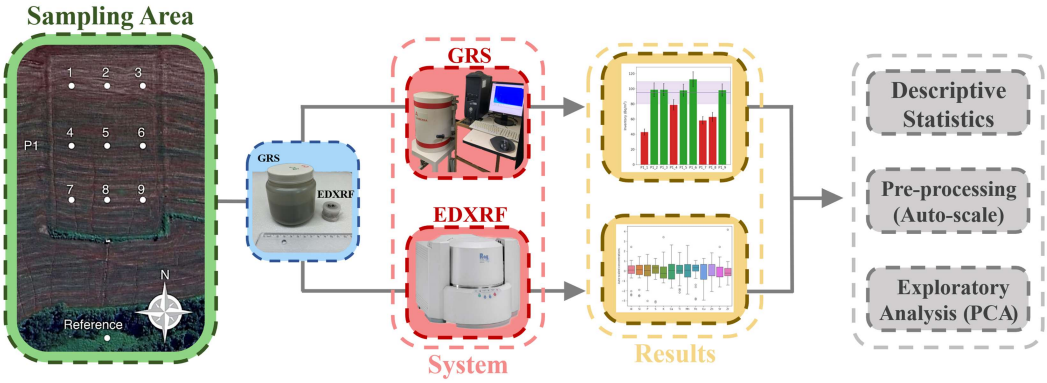
The complementary nature of these datasets also introduces a challenge, as both techniques generate multiple correlated variables that cannot be readily interpreted using conventional univariate approaches. Relationships between radionuclide inventories and elemental composition are often indirect and may involve several interacting factors, making multivariate statistical methods particularly suitable for their joint evaluation. Among these methods, PCA provides an efficient framework for identifying the dominant sources of variability and revealing correlations between the variables measured by both spectroscopic techniques.

Considering the GRS and EDXRF properties, and the PCA capability to facilitate the analysis of several variables, the hypothesis elaborated in this study were: (*i*) the combination of the two spectroscopic techniques could be performed by PCA to better understand the correlations among ^{137}Cs and the macro and micro-nutrients found in soil; and (*ii*) by studying the correlations between GRS and EDXRF results using PCA it would be possible to determine if EDXRF can be used as a complementary technique to better understand soil redistribution. The methodology and results of this study were published in the peer-reviewed Journal of Environmental Radioactivity [96].

7.1.1 Methodology

Figure 36 summarizes the study process, from sample collection to data analysis. A detailed description of each step of the workflow is given in the following subsections.

Figure 36 – Workflow of the applied methodology for the study of complementarity of GRS and XRF for assessing soil redistribution.



Source: self-elaboration.

7.1.1.1 Sampling and Preparation

The soil samples employed in this work were collected in the region described in Section 6.1. Nine sampling points were selected within the agricultural plot area, as shown in Figure 36, distributed in a way that the soil redistribution inside the plot could be evaluated. The agricultural plot was occupied by two rotary cultures for the past 50 years, soybean (*Glycine max*) and corn (*Zea mays*). A reference sample was also collected nearby in an undisturbed area for inventory comparison.

A depth-incremental technique was employed in the sampling. At each point, a motorized cylinder auger was used to collect three samples, each with a 10 cm increment. Consequently, every sampling point yielded three samples: one from 0 to 10 cm in depth, another from 10 to 20 cm, and a third from 20 to 30 cm. This method enables the evaluation of soil redistribution effects over time on the hillslope.

Following the collections, the samples were allowed to air dry for a minimum of 48 hours and subsequently sieved through a 2 mm mesh to separate the finer soil particles from the larger rock fragments. The sieving process was conducted to achieve the optimal ratio between sample mass and the ^{137}Cs concentration.

7.1.2 Experimental Setup

The measurements were performed employing the detectors described in Subsections 2.3.2 and 2.6.1.

For the measurements, approximately 300 grams of sieved soil samples were deposited inside the plastic recipients shown in Figure 36 and positioned directly on top of the detector for 86400 seconds of live time. ^{137}Cs activity and concentration were calculated with the spectra net peak area and the Proportional Model for soil loss was employed to estimate soil redistribution, as described in Subsection 3.2.2.

After completing the GRS measurements, approximately 10 g of each soil sample were dedicated to subsequent EDXRF analysis. The soil samples were placed in polyethylene XRF cups (No. 1530, Chemplex Industries Inc., USA) and sealed with Mylar film. The EDXRF measurements were conducted using the Shimadzu EDX 720 equipment.

XRF calibration was carried out by measuring various CRMs for soils, resulting in empirically derived calibration curves. These CRMs were manufactured by the International Atomic Energy Agency and are constituted of clay soils (PTXRF-IAEA04 and PTXRF-IAEA13), river clay sediment (PTXRF-IAEA09), plastic clay (IPT32) and São Simão clay (IPT42).

7.1.3 Quantified Elements

The descriptive statistics with data from GRS and EDXRF spectrometry are shown in Table 10. Higher Fe, Al and Si concentrations were identified, what may be attributed to soil characteristics, including the soil profile and mineral fraction. On the other hand, soils that are in an advanced state of weathering show intense loss of silica (composed mainly of SiO_2), with a resulting accumulation of insoluble iron and aluminum oxides [97], the most frequent being goethite (FeOOH), hematite (Fe_2O_3) and gibbsite ($Al(OH)_3$) [97, 98]. In lower orders, Ti, Mn, Ca and K contents were also quantified in larger quantities than the remaining elements. Although K, Ca and Mn may be naturally found in soil, fertilizers are rich in these minerals and their application increases their concentration in agricultural soil [99]. It is worth noting that fertilizers are commonly applied up to 20 cm in depth.

Beyond the quantification of ^{137}Cs, GRS allowed for the simultaneous quantification of ^{40}K, ^{214}Bi and ^{228}Ac in the soil samples. Regarding CV values, most elements were below 30%. Nonetheless, the ^{137}Cs concentration had a 62% variation, which is attributed to the natural vertical distribution of the radionuclide and the soil redistribution experienced in the agricultural plot, as described in Section 3.2.2. The MDA of ^{137}Cs in the soil samples was 0.043 ± 0.016 Bq·kg^{-1}.

Table 10 – Descriptive statistics of concentration data for all the samples ($n = 27$).

Element	Mean ± SD	Med	CV	Max	Min	Q25	Q75	Kurt	Skew
Al (g kg^{-1})	110±17	103	16%	141	61	97	110	1.22	-0.66
Si (g kg^{-1})	107±14	109	13%	127	71	102	115	0.60	-0.95
P (g kg^{-1})	0.90±0.17	0.91	19%	1.19	0.38	0.81	1.01	1.18	-0.84
S (mg kg^{-1})	434±85	460	20%	540	170	415	482	3.64	-1.96
K (mg kg^{-1})	543±165	503	30%	1097	350	425	613	2.50	1.50
Ca (g kg^{-1})	2.6±0.5	2.7	21%	4.0	1.7	2.2	3.0	-0.35	0.52
Ti (g kg^{-1})	25.3±1.7	25.3	7%	28.5	21.6	24.9	26.3	-0.28	-0.39
Mn (g kg^{-1})	2.4±0.3	2.4	12%	3.1	1.7	2.2	2.6	0.20	0.13
Fe (g kg^{-1})	161±7	163	5%	169	140	161	165	1.83	-1.67
Cu (mg kg^{-1})	66±6	70	9%	73	50	62	70	0.26	-0.88
Zn (mg kg^{-1})	82±4	80	5%	90	73	80	85	-0.71	0.28
Sr (mg kg^{-1})	12±3	10	28%	20	7	10	13	-0.50	0.40
Zr (mg kg^{-1})	411±20	407	5%	493	383	402	417	7.70	2.33
^{40}K (kBq m^{-2})	3.67±0.50	3.65	13%	4.65	1.90	3.50	3.95	4.61	-1.38
^{137}Cs (Bq m^{-2})	27±17	29	62%	56	<LOQ	12	56	-1.13	-0.22
^{214}Bi (kBq m^{-2})	14.2±0.2	15.1	14%	17.5	0.8	14.2	15.7	2.01	-1.44
^{228}Ac (kBq m^{-2})	25.3±0.3	25.7	12%	29.4	13.8	24.4	27.1	4.56	-1.80

SD = Standard Deviation, Med = Median, CV = Coefficient of Variation, Max = Maximum, Min = Minimum, Q25 = First Quartile, Q75 = Third Quartile, Kurt = Kurtosis, Skew = Skewness, LOQ = Limit of Quantification.

7.1.4 Soil Redistribution

The measured ^{137}Cs inventory values are displayed in Figure 37. In Figure 37a, individual inventory values for each sample collected are shown. Figure 37b illustrates the summed inventory values for the increments in each sampling point. It also displays the mean reference value calculated in a set of undisturbed soil samples, which was 95 Bq m^{-2}, calculated from three different reference samples and is depicted as the purple line. The shaded area represents its standard deviation.

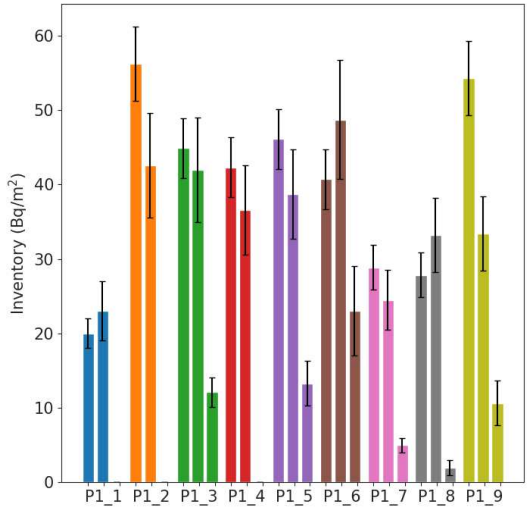
Overall, the points from P1_1 to P1_8 show the expected behavior for mechanized soils. By comparing the inventory for each sample with the mean undisturbed calculated inventory, it becomes clear that point P1_6 experienced deposition, points P1_1, P1_4, P1_7, and P1_8 experienced erosion, and points P1_2, P1_3, P1_5, and P1_9 are seemed to be mechanically stabilized.

The points whose third increment showed ^{137}Cs activities below the detection limit, P1_1, P1_2, and P1_4, suffered so much erosion that the soil profile shifted from organic to mineral, given that the uppermost soil layer has been moved completely. Even so, the point labeled P1_2 still shows an inventory higher than the reference one, which could indicate that this point experienced soil deposition in the past, but the current

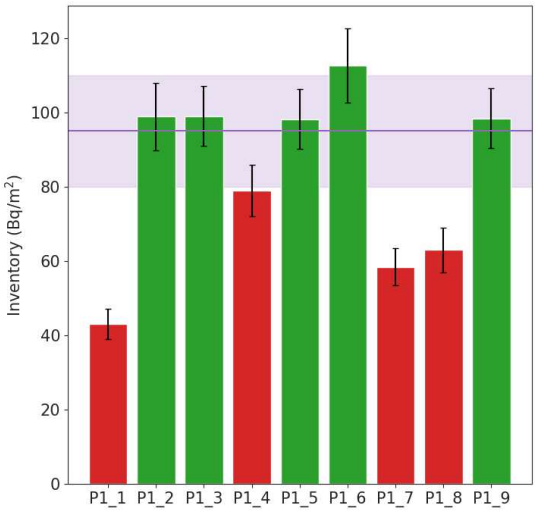
soil redistribution is intense enough to cause a net soil loss, or might indicate heavy soil mechanization in the past.

Figure 37 – Barplot of the calculated inventories for the soil samples collected in the agricultural plot. In 37a, the different colors indicate the sampling points, while in 37b they indicate soil erosion (red) or deposition (green). The purple line indicate the mean reference inventory, and the shaded area its standard deviation.

(a) Individual inventory per increment.



(b) Total inventory per sampling point.

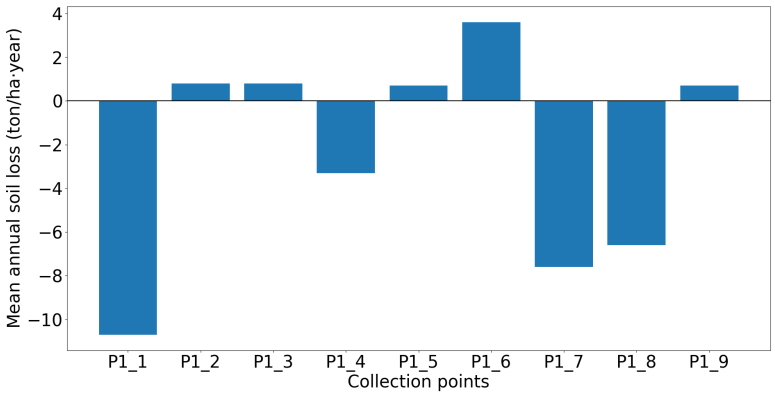


Source: self-elaboration.

Point P1_9 does not show the normal behavior of homogenized soil. Instead, its profile is similar to undisturbed soil. Given this behavior and the fact that the total inventory is similar to the reference one, it is possible to assume that point P1_9 experienced soil accumulation of topsoil originated from within the agricultural plot. Both sampling points P1_7 and P1_8 experienced topsoil loss and are aligned with point P1_9, following the cultivation direction. This row of the agricultural plot seems to have experienced topsoil loss and, currently, the moved soil from the first two points is accumulated in point P1_9.

Equation 3.10 was applied to the results in Figure 37b in order to calculate the mean annual soil loss for each sampling site. The results are displayed in Figure 38, which shows the regions of the agricultural plot that have suffered the most intense soil redistribution in the past 60 years. The proportional model estimated a gross erosion rate of 28.2 ton ha^{-1} $year^{-1}$ and a net soil deposition of 6.6 ton ha^{-1} $year^{-1}$, hence, a net soil loss of 21.6 ton ha^{-1} $year^{-1}$.

Figure 38 – Mean annual soil loss for the nine sampling points.



Source: self-elaboration.

7.1.4.1 Multivariate Analysis

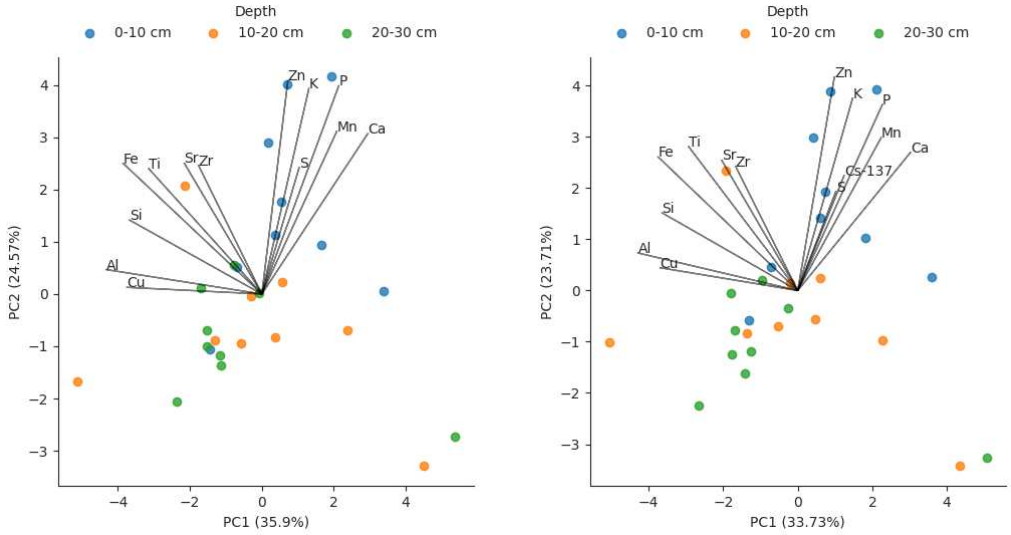
Once all the data regarding elemental concentration in the soil samples was gathered, they were organized into a single matrix to study the correlations between the detected elements and ^{137}Cs. With the aim of understanding how the quantified elements influence the soil redistribution, the PCA with the data from both techniques was performed and the results are shown in Figure 39. Two PCAs were performed in this case, the first one with all the detected elements and the second one without the influence of ^{137}Cs. Upon inspecting the plots, two conclusions can be immediately drawn. The first one is that the PCA shows a tendency to separate the samples according to their sampling depth. And, the second one is that ^{137}Cs influence is not enough to, on its own, cause too much change in the visualization of the samples in the scores plot.

The PC2 positive axis is responsible for making the depth separation. According to the loadings in Figure 39, the elements associated with this direction are, in order of magnitude, Zn, K, P, Mn, Ca, ^{137}Cs and S. Most of these elements can be associated with the use of fertilizers [99] and, according to Table 10, all of these elements, except Zn, have a coefficient of variation equal or greater than 20%. Therefore, the higher concentration of these elements in topsoil can be attributed to the use of fertilizers.

The vertical distribution of ^{137}Cs coincides with the distribution of these elements, which is expected since they are more concentrated in the topsoil and deposited from above. The fact that ^{137}Cs is associated with the same direction as these elements in the PCA plot indicates that this radionuclide behaves similarly to fertilizers, which supports the hypothesis that combining the two spectroscopic techniques provides a better understanding of soil dynamics.

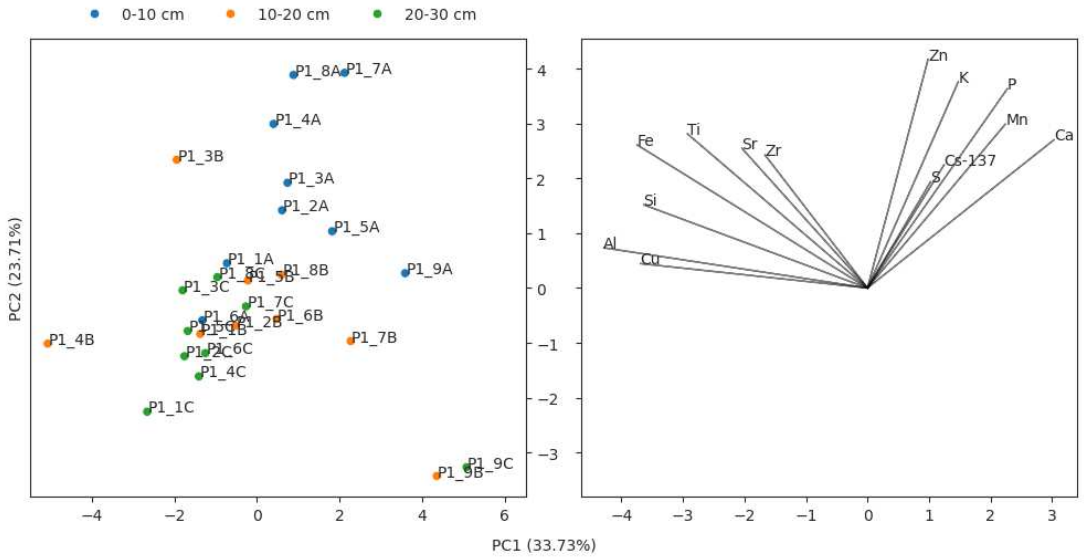
Figure 39 – PCA of all the studied samples. The scores and loadings are plotted together in both cases.

(a) PCA without ^{137}Cs. (b) PCA with ^{137}Cs.



Source: self-elaboration.

Figure 40 – Scores and loadings plots of all elements with the names of the samples.



Source: self-elaboration.

Upon inspection of the PCA plots, it is possible to attest that some samples from the depth range of $0 - 10$ cm are mixed with the deeper ones. Figure 40 shows the PCA of all elements with their respective names added to the scores plot, to become evident which samples are behaving differently.

As stated before, the GRS analysis provided the information that point P1_9 experienced intense soil redistribution in the past 60 years. The effects of this fact can

also be verified in Figure 40 since the three samples from this particular point are further apart from the others. Samples P1_1C, P1_2C and P1_4C, that were shown to have ^{137}Cs concentrations below the detection limit in Figure 37a are displayed at the bottom of the distribution, in the opposite direction of this element's contribution.

Another result that can be compared to what has been seen in Subsection 7.1.4 is the score of sample P1_3B in Figure 40. The fact that this sample is closer related to samples from profile $0 - 10$ cm than to $10 - 20$ cm indicates that this particular region of the agricultural plot experienced enough redistribution to change the soil's original profile, shifting this increment from mineral to organic horizon.

In counterpart, the effects of soil redistribution over samples P1_1A and P1_6A shifted their profiles from organic to mineral. This behavior can be noted in Figure 40 since these samples are displayed among the deeper ones. These results are in accordance with the conclusions gathered from Figure 37a. The points P1_1 and P1_6 have a similar behavior between the first two increments that differ only by the magnitude of the measured inventories. This behavior can be attributed to the loss of topsoil in the former, while the latter experienced soil accumulation originating from deeper horizons.

Overall, Figures 39 and 40 corroborate the two previously made hypotheses, that the combination of both techniques allows for a better understanding regarding the correlations between ^{137}Cs and the macro and micro-nutrients present in the soil; and the feasibility of using EDXRF to complement GRS measurements.

7.1.5 Conclusions

The GRS performance in soil samples from an agricultural plot allowed the estimation of soil redistribution within such an agricultural plot in the past six decades. A gross erosion rate of 28.2 ton ha^{-1} year^{-1} and a net soil deposition of 6.6 ton ha^{-1} year^{-1} were calculated. These results may be used to evaluate the quality of soil management practices applied in the agricultural area.

The results indicate that EDXRF may be used as a complementary technique to GRS since it was possible to detect shifts in soil profiles ranging from 0 to 30 cm in depth and correlate ^{137}Cs vertical distribution with elements introduced into the soil due to the application of fertilizers. Moreover, the use of PCA combining the results obtained with EDXRF and ^{137}Cs quantification corroborated that this radionuclide behaves similarly to fertilizers. However, further investigation is required.

7.2 Data Fusion of Gamma and X-ray Spectra via ComDim for Soil Characterization

This study aimed to integrate pXRF and GRS data through the ComDim approach (Subsection 4.1.2) using soil samples collected from an agricultural field, and to compare the performance of this methodology with that of more conventional single-block exploratory analyses. This integrated strategy seeks to improve the assessment of soil redistribution processes in agricultural environments, as well as the monitoring of sediment-associated nutrient transport. Specifically, the ComDim results were compared with short-term soil redistribution estimates derived from ^{7}Be inventories in soil.

The samples employed in this work are the same as those described in Section 6.1. Gamma-ray spectrometry measurements were carried out using the detector described in Subsection 2.3.1, with a live time of 86400 s. Subsequently, approximately 5 g of each sample was repurposed for the pXRF measurements. The powdered material was transferred to polyethylene XRF sample cups (No. 153, Chemplex Industries Inc., USA), sealed with Ultralene film, and manually compacted into packed-powder form, resulting in sample heights of approximately 10 mm, which is considered infinitely thick for the X-ray energy range employed [15]. The pXRF measurements were performed, in duplicate, using a Bruker Tracer 5i (Subsection 2.6.2), with a live time of 30 s. The spectra used for ComDim were those obtained with the 15 kV condition, since this excitation corresponds to the optimal condition for modelling spectral data obtained with this spectrometer [50]. The methodology and results of this study were published in the peer-reviewed Soil Science Society of America Journal [100].

7.2.1 Data Fusion

Prior to the ComDim analysis, both spectral datasets were trimmed to reduce the influence of noise and electronic fluctuations. The pXRF spectra were restricted to the energy range between 0.98 and 15.00 keV (750 channels), while the final GRS spectra covered the interval from 238 to 2800 keV (640 channels).

The fused dataset was constructed using the raw spectra obtained for each sample from both analytical techniques after applying the previously described spectral trimming procedures. Prior to fusion, each block was normalized by its maximum intensity variable, namely the Fe Kα emission at 6.4 keV for pXRF and the ^{212}Pb emission at 238.6 keV for GRS. The number of Common Dimensions (CDs) retained in the model was determined based on the cumulative explained variance, excluding components dominated primarily by statistical fluctuations and spectral noise. All analyses were carried out in the Julia programming language using the `Jchemo.jl` package [56].

The use of raw spectra aimed to preserve the intrinsic spectral structure while

avoiding excessive preprocessing assumptions. By operating directly on spectral channels, information related to peak morphology, spectral overlap, continuum contributions, and covariance between neighboring variables is retained, allowing multiblock methods such as ComDim to more effectively capture shared latent structures between the GRS and pXRF datasets. Conversely, transforming spectra into peak areas or elemental concentrations introduces additional processing stages, such as background correction, peak deconvolution, and calibration, each of which depends on modelling assumptions capable of altering the original error structure and propagating uncertainty [101, 102]. Since the objective of this study was the exploratory investigation of coupled spectral patterns associated with soil redistribution processes, rather than the construction of predictive calibrated models, maintaining the spectral-domain representation was considered more suitable. To reduce the influence of noise-related artifacts while preserving relative spectral contrasts, long GRS acquisition times, replicate pXRF measurements followed by averaging, exclusion of noisy spectral regions, and block normalization procedures were employed.

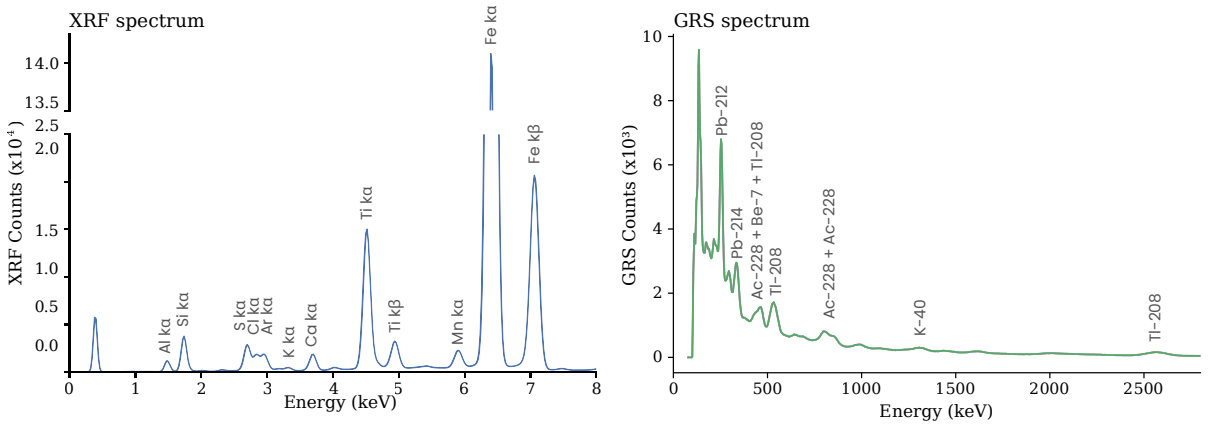
The projections of the samples onto the CD score space were subsequently compared with the short-term soil redistribution patterns inferred from the presence of ^{7}Be in the GRS spectra. In addition, the loadings and saliences obtained from the ComDim model were used to evaluate the relative contribution of variables from each analytical technique, providing an estimate of the degree of synergism between pXRF and GRS in the characterization of soil redistribution processes.

7.2.2 XRF and GRS results

Figure 41 presents the average pXRF and GRS spectra obtained for the analyzed dataset. Due to the 15 kV operating voltage of the X-ray tube, the excitation efficiency for elements heavier than Fe ($Z > 26$) is considerably reduced, restricting the detectable XRF signal predominantly to the K-lines of lighter elements [15]. For this reason, the XRF spectrum is displayed only within the 0.0–8.0 keV interval to improve visualization.

From the fluorescent emissions observed in the 15 kV pXRF spectra, the presence of Al, Si, P, S, K, Ca, Ti, Mn, Fe, Cu, and Zn could be confirmed in the samples. In addition to the influence of the pXRF operating conditions, the comparatively higher intensities of the Fe and Ti peaks are likely associated with the mineralogical composition of the Rhodic Ferralsol and Nitisol soils. These soils are typically characterized by elevated concentrations of hematite (Fe_2O_3), goethite (FeOOH), and gibbsite ($Al(OH)_3$), while Ti may occur within iron oxide structures as an isomorphic substitute for Fe [103].

Figure 41 – Mean spectra of the measured samples (n=31) for pXRF (left) and GRS (right). Detected peaks are indicated in both plots, and the pXRF spectra y-axis is broken between 2.5 and 13.5 keV to facilitate the visualization of low-intensity emissions.



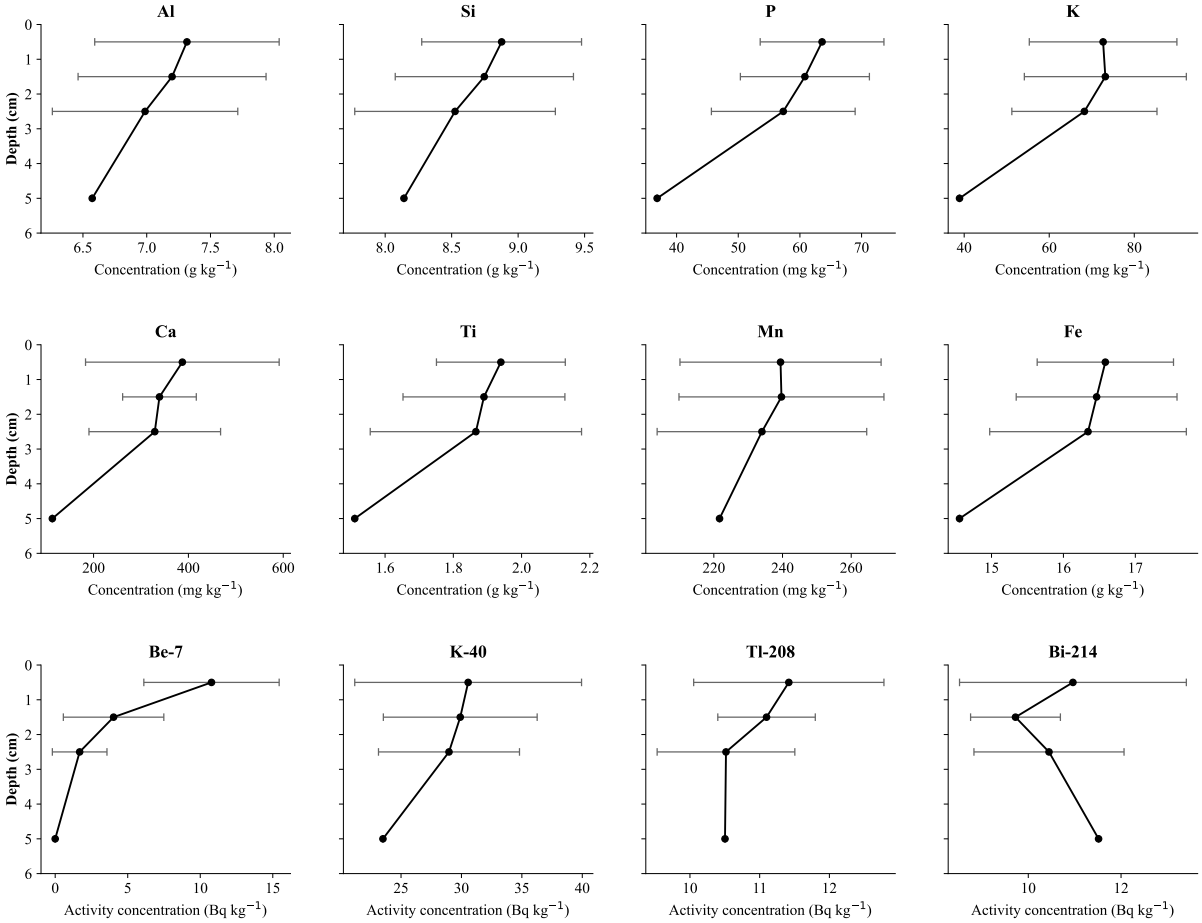
Source: self-elaboration.

The mean GRS spectrum was restricted to the 0.0–2800 keV energy range to emphasize the naturally occurring radionuclide emissions. Spectral regions below 238 keV are more difficult to interpret because part of the detected signal may originate from electronic noise and instrumental fluctuations. Characteristic emissions associated with the ^{238}U decay series (^{214}Pb and ^{214}Bi) and the ^{232}Th series (^{208}Tl, ^{212}Pb, and ^{228}Ac) were identified, together with emissions from ^{40}K and ^{7}Be. Although the ^{7}Be signal is less pronounced than those of the naturally occurring radionuclides, variations in activity can still be observed in the region around 500 keV.

After the qualitative identification of the spectral features, the quantified elemental concentrations and radionuclide activities were used to construct vertical distribution profiles (Figure 42). These profiles provide an overview of the geochemical stability of the soil matrix across the analyzed depth intervals, as well as the characteristic depth-dependent behavior of ^{7}Be.

The elemental profiles shown in Figure 42 remain relatively stable throughout the 0–3 cm depth interval, exhibiting only minor fluctuations that may reflect natural heterogeneity and sampling variability. In contrast, the ^{7}Be profile displays a pronounced exponential decrease with depth, consistent with the expected behavior of a cosmogenic radionuclide characterized by rapid adsorption onto soil particles and a relatively short half-life [4]. The reference sample collected at 5 cm depth presented ^{7}Be activity below the detection limit, confirming its suitability as an undisturbed reference for soil redistribution assessment. While Figure 42 shows the mean trends (n = 10 per depth increment), a complete summary of the descriptive statistics for each variable, including quartiles, skewness, and kurtosis, is presented in Table 11.

Figure 42 – Vertical distribution profiles of elemental concentrations (g kg^{-1} and mg kg^{-1}) and radionuclide activities (Bq kg^{-1}) for the 0–3 cm soil layers. Data points represent the mean (n=10 per depth increment) and error bars indicate the standard deviation. The reference sample at 5 cm is included to illustrate the baseline for redistribution assessment.



Source: self-elaboration.

Table 11 – Complete descriptive statistics of soil elemental concentrations and radionuclide activities separated by depth ($n = 10$ per depth increment). Elements are ordered by atomic number; minor constituents are reported in mg kg^{-1} for clarity.

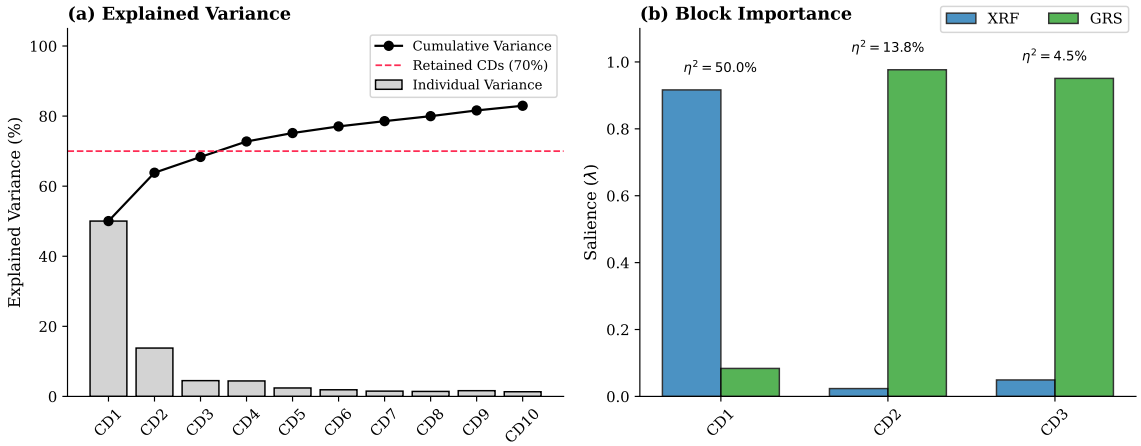
Depth	Var.	Unit	Mean	SD	Min	Q25	Q50	Q75	Max	Kurt	Skew
0–1 cm	Al	g kg^{-1}	7.3	0.7	6.5	6.8	7.3	7.7	8.8	-0.03	0.82
	Si	g kg^{-1}	8.9	0.6	8.1	8.4	8.8	9.3	10.0	-0.82	0.33
	P	mg kg^{-1}	63.6	10.0	47.5	56.3	63.7	71.4	80.0	-0.91	0.12
	K	mg kg^{-1}	72.7	17.4	55.8	63.0	67.5	75.0	114.4	1.42	1.45
	Ca	mg kg^{-1}	387.2	204.2	169.4	283.1	308.4	453.1	845.1	0.52	1.21
	Ti	g kg^{-1}	1.9	0.2	1.7	1.9	1.9	2.1	2.2	-0.91	0.07
	Mn	mg kg^{-1}	239.4	29.3	206.5	214.6	235.5	255.3	296.6	-0.63	0.64
	Fe	g kg^{-1}	16.6	0.9	15.2	15.8	16.6	17.1	18.0	-1.13	0.13
	^{7}Be	Bq kg^{-1}	16.4	4.7	8.2	13.7	16.3	19.3	23.5	-0.76	-0.13
	^{40}K	Bq kg^{-1}	30.6	9.4	21.1	27.8	29.1	30.6	55.3	3.15	1.93
	^{208}Tl	Bq kg^{-1}	11.4	1.4	9.1	10.7	11.2	12.1	13.7	-0.58	0.11
	^{214}Bi	Bq kg^{-1}	11.0	2.4	8.1	8.8	11.0	13.4	13.7	-1.79	-0.01
1–2 cm	Al	g kg^{-1}	7.2	0.7	6.1	6.9	7.3	7.5	8.6	-0.14	0.14
	Si	g kg^{-1}	8.7	0.7	7.7	8.3	8.9	9.1	9.8	-0.96	-0.34
	P	mg kg^{-1}	60.8	10.5	37.7	57.4	60.8	67.0	75.0	0.55	-0.84
	K	mg kg^{-1}	73.2	19.0	50.5	62.6	68.4	81.4	117.5	0.98	1.20
	Ca	mg kg^{-1}	339.0	77.7	205.9	317.6	321.2	404.9	437.0	-0.98	-0.32
	Ti	g kg^{-1}	1.9	0.2	1.6	1.7	1.8	2.0	2.4	-0.07	0.65
	Mn	mg kg^{-1}	239.7	29.8	206.9	218.6	227.9	257.8	302.9	-0.12	0.92
	Fe	g kg^{-1}	16.5	1.1	14.6	16.1	16.4	17.4	18.1	-0.87	-0.19
	^{7}Be	Bq kg^{-1}	9.6	3.5	5.5	7.8	8.9	10.3	15.9	-0.49	0.78
	^{40}K	Bq kg^{-1}	29.9	6.4	23.9	26.9	28.0	30.2	45.5	1.73	1.60
	^{208}Tl	Bq kg^{-1}	11.1	0.7	9.9	10.6	11.3	11.6	11.9	-0.96	-0.71
	^{214}Bi	Bq kg^{-1}	9.7	1.0	8.7	9.1	9.4	9.9	12.1	1.80	1.55
2–3 cm	Al	g kg^{-1}	7.0	0.7	6.0	6.7	6.8	7.7	8.0	-1.20	0.11
	Si	g kg^{-1}	8.5	0.8	7.6	7.9	8.4	9.0	9.9	-0.94	0.40
	P	mg kg^{-1}	57.3	11.7	34.9	51.1	57.7	61.6	76.8	-0.11	-0.21
	K	mg kg^{-1}	68.3	17.1	46.7	58.4	62.6	79.7	99.0	-0.93	0.44
	Ca	mg kg^{-1}	329.1	138.8	156.3	260.0	294.4	367.7	625.7	0.18	0.95
	Ti	g kg^{-1}	1.9	0.3	1.3	1.7	1.9	2.1	2.3	-0.81	-0.04
	Mn	mg kg^{-1}	234.0	30.5	206.4	209.4	221.9	257.4	287.2	-1.09	0.70
	Fe	g kg^{-1}	16.3	1.4	14.0	15.6	16.2	17.3	18.3	-1.00	-0.07
	^{7}Be	Bq kg^{-1}	6.4	3.1	0.0	4.9	7.2	8.1	11.4	0.17	-0.56
	^{40}K	Bq kg^{-1}	29.0	5.8	21.0	25.3	28.9	30.8	41.3	0.15	0.64
	^{208}Tl	Bq kg^{-1}	10.5	1.0	9.3	9.7	10.5	10.8	12.6	-0.02	0.77
	^{214}Bi	Bq kg^{-1}	10.4	1.6	8.0	9.3	10.5	11.6	13.0	-1.07	-0.01

SD = Standard deviation, Min = Minimum, Q25 = First quartile, Q50 = Second quartile, Q75 = Third quartile, Max = Maximum, Kurt = Kurtosis, Skew = Skewness.

7.2.3 ComDim Analysis

To evaluate the synergistic information contained in pXRF and GRS datasets for the assessment of soil redistribution processes, ten common dimensions (CDs) were computed using the ComDim framework. The individual and cumulative explained variance associated with each component are shown in Figure 43(a). The first common dimension alone accounts for approximately 50% of the total variance, indicating a strong shared structure between the two analytical blocks.

Figure 43 – (a) Individual and cumulative explained variance for the first ten common dimensions. The dashed line indicates the cumulative variance threshold used to define the retained dimensionality. (b) Normalized saliences (λ) of the pXRF and GRS blocks for the first three CDs. The values of η^2 indicate the proportion of variance explained by each CD.



Source: self-elaboration.

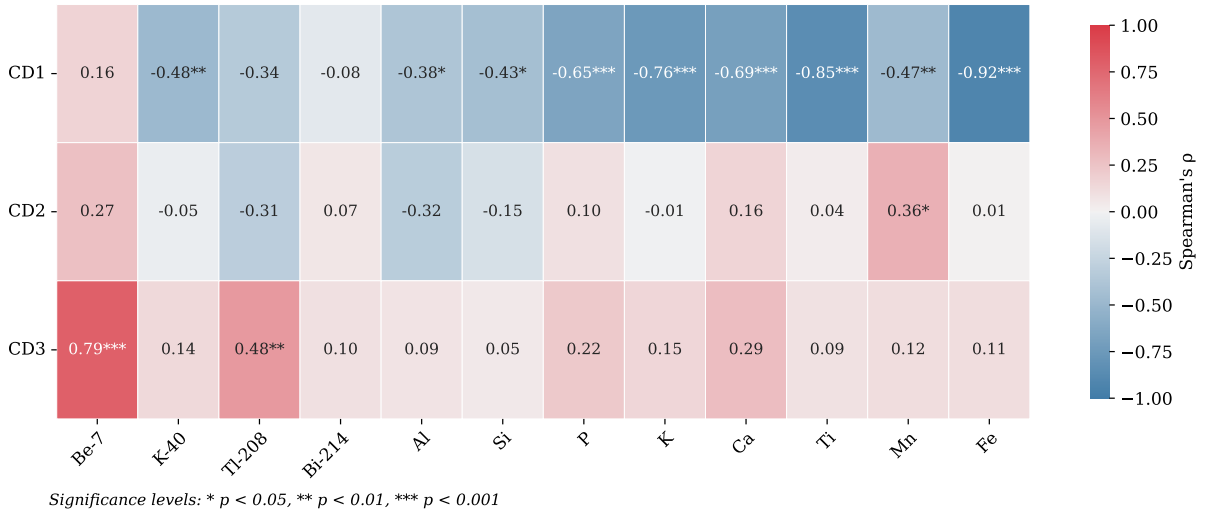
Based on the cumulative explained variance, the first three CDs were retained for further analysis. Together, they explain approximately 68% of the total variance, while higher-order components contribute only marginal additional information and are therefore not considered in the interpretation.

The relative contribution of each analytical technique to the retained dimensions is shown through the saliences in Figure 43(b). CD1 is strongly dominated by the pXRF block, whereas CD2 and CD3 are primarily associated with the GRS block. This pattern indicates that CD1 captures variability mainly driven by elemental composition, while CD2 and CD3 reflect variability more closely related to radionuclide activity. This separation highlights the complementary nature of the two datasets within a shared multiblock representation.

To further interpret these latent structures, Spearman's rank correlation coefficients (ρ) were computed between CD scores and the independently quantified elemental concentrations and radionuclide activities (Figure 44). CD1 shows strong and statistically

significant correlations with major elemental variables measured by pXRF, particularly Fe, Ti, K, Ca, and P ($\rho > 0.65$, $p < 0.01$ and $p < 0.001$), confirming its association with geochemical variability of the soil matrix.

Figure 44 – Spearman's rank correlation coefficients (ρ) between the scores of the first three common dimensions and the analytical variables from GRS and pXRF. Statistical significance is indicated by asterisks.



Source: self-elaboration.

In contrast, CD3 is primarily correlated with ^{7}Be activity ($\rho = 0.79$, $p < 0.001$), with a secondary significant contribution from ^{208}Tl ($\rho = 0.48$, $p < 0.01$). This result indicates that CD3 captures variability related to radionuclide redistribution processes, which are typically associated with recent surface dynamics.

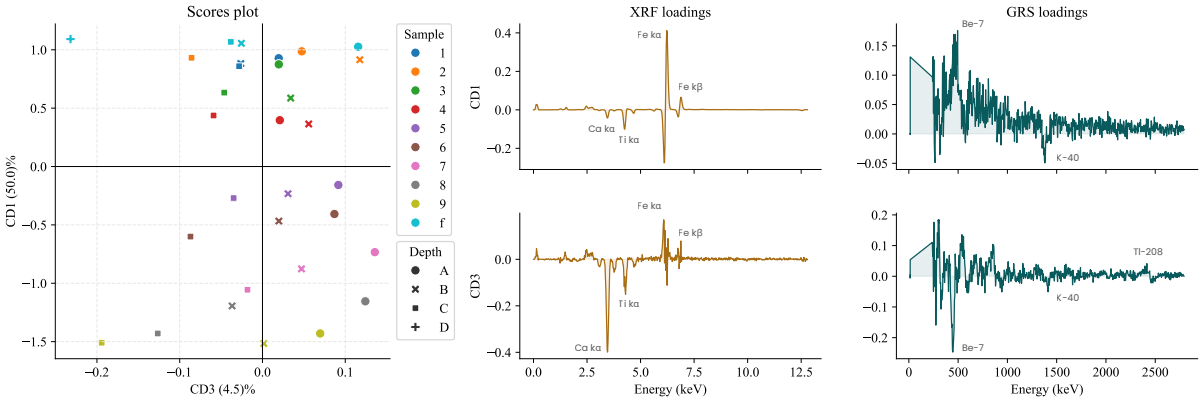
The strong influence of XRF variables on CD1 reflects the dominant role of the soil mineral matrix in structuring the multiblock space. The studied Rhodic Ferralsols and Nitisols are characterized by high clay contents and abundant Fe- and Al-oxyhydroxides, which provide a relatively stable geochemical background. Variability along CD1 is therefore interpreted as reflecting differences in mineralogical composition and spatial redistribution of these phases across the landscape.

In contrast, CD3 is strongly linked to ^{7}Be activity, which is consistent with its rapid adsorption onto fine soil particles after deposition. As a consequence, CD3 scores can be interpreted as a proxy for recent surface processes affecting soil particle redistribution. This suggests that the ComDim framework is sensitive to processes involving the transport and reworking of fine particles, which simultaneously carry radionuclides and geochemical tracers.

The interpretation of the common dimensions is further supported by the score and loading structures (Figure 45). CD1 separates samples primarily according to sampling location, consistent with its dominance by the pXRF block ($\lambda \approx 90\%$), while CD3

separates samples according to depth, consistent with its dominance by the GRS block ($\lambda \approx 95\%$). Accordingly, CD1 is interpreted as a descriptor of spatial geochemical variability, while CD3 reflects depth-dependent radionuclide behavior, particularly associated with ^{7}Be.

Figure 45 – Scores and loadings representation of the ComDim results. On the left, the scores plot shows the distribution of samples in the CD1xCD3 space, where colors represent sampling locations and marker shapes indicate sampling depths. On the right, the loadings plots for the pXRF and GRS spectra illustrate the contribution of spectral energy variables to each common dimension, highlighting the variables driving the observed sample separation.



Source: self-elaboration.

Variability along CD1 is primarily associated with differences in soil composition, whereas CD3 captures depth-dependent variations related to radionuclide redistribution. Samples from the same location show limited dispersion along CD1, indicating relatively consistent elemental composition within soil profiles, while variations along CD3 follow the expected vertical behavior of ^{7}Be.

It is important to emphasize that proximity between samples in the ComDim space reflects similarity in multivariate geochemical and radiometric signatures, rather than a direct measure of soil stability. Therefore, interpretation of soil redistribution must be made relative to a reference condition. In this study, the reference sample represents an undisturbed baseline, and deviations from it are interpreted as potential indicators of erosion or deposition processes.

Spatially, samples from the upper part of the study area, including the reference condition, exhibit relatively limited compositional variability, whereas samples from other locations show greater dispersion along CD1. However, no clear monotonic trend is observed along the hillslope, suggesting that soil redistribution processes are spatially heterogeneous and likely influenced by local factors such as microtopography, vegetation cover, and land management practices.

These findings are consistent with the relatively low and spatially variable soil

redistribution rates estimated for the area (Figure 33), which indicate a combination of minor erosion and deposition rather than systematic downslope transport.

7.2.4 Conclusions

ComDim demonstrated its effectiveness as a powerful multiblock algorithm for evaluating the underlying influences of variables originating from complementary spectroscopic techniques. The fusion of pXRF and GRS raw spectra allowed for the identification of joint patterns that distinguished sampling locations and depths. Specifically, CD1 (strongly correlated with pXRF) separated samples by geochemical signature, while CD3 (strongly correlated with GRS) successfully described samples according to their sampling depth.

Statistical validation through Spearman's correlation confirmed that the common dimension scores serve as reliable proxies for ^{7}Be activity and elemental concentrations, providing a quantitative link between the multiblock output and established soil redistribution tracers. However, the current study is limited by a small sample size ($n = 31$) and focus on specific soil classes (Ferralsols and Nitisols). Further research with larger, more diverse datasets is required to assess the method's transferability and its sensitivity to environmental factors such as soil moisture.

While these results indicate that raw spectral signatures can effectively capture redistribution-related variability, the use of these signatures to classify regions as disturbed or undisturbed remains unclear. This integrated approach enhances the monitoring of soil redistribution and sediment-bound nutrient transport by studying the synergy between nuclear and atomic spectroscopic data.

7.3 Exploratory Analysis of GRS, XRF, and Wet Chemistry Quantitative Data for Soil Analysis

The continuous monitoring of soil health and fertility is of the utmost importance to ensure quality and quantity of food production [104]. Access to high-density information on the soil fertility attributes spatial variability is vital in the decision-making process to promote better nutrition conditions for plants [50]. The determination of soil fertility parameters is commonly performed through wet chemistry processes, which are expensive, time demanding, and waste-producing [105, 106]. In recent years, XRF has emerged as a rapid alternative to wet chemistry for soil fertility analysis, targeting key nutrients like P, K, Ca, Mg, and Fe [50, 80, 107–110].

Moreover, some studies have combined data from GRS and XRF to model and predict soil fertility parameters [111,112], perform topographic studies [113], and study soil erosion [96]. Motivated by the synergy between GRS and XRF analytical results for soil

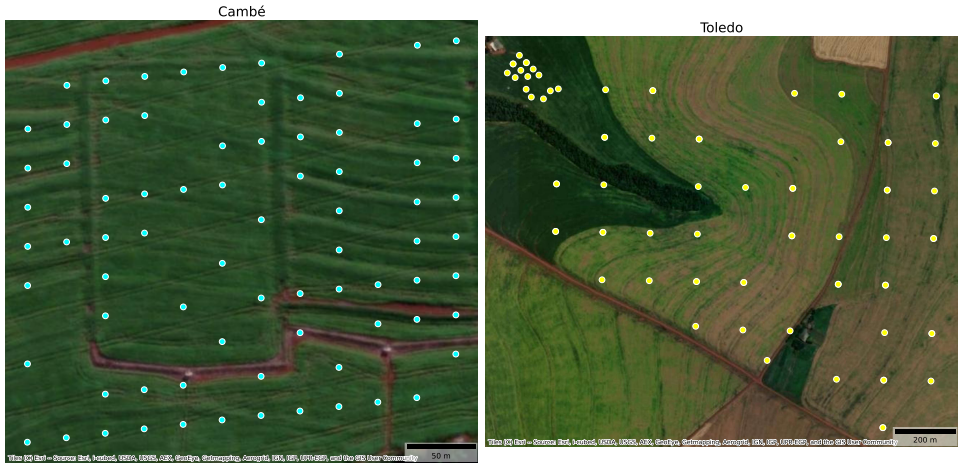
studies, this part of the present work focused on the exploratory analysis of quantitative data obtained from GRS, XRF, and wet chemistry measurements, using a multiblock approach to jointly evaluate the information provided by each analytical technique.

7.3.1 Sampling and Measurements

For this part of the work, soil samples originated from two distinct sites and were collected by the Instituto de Desenvolvimento Rural do Paraná (IDR-Paraná). One of the sampling regions corresponds to the area described in Section 6.1, while the second was located within an agricultural area in the municipality of Toledo, Paraná State, Brazil, approximately between the latitudes 24°41′ and 24°46′ S and the longitudes 53°41′ and 53°48′ W. The inclusion of two independent agricultural sites allowed the proposed methodology to be evaluated under different environmental and pedological conditions, increasing the representativeness of the analyzed dataset.

At each study site, a total of 160 soil samples were collected from two depth intervals, namely 0–10 cm and 10–20 cm, hereafter referred to as horizons A and B, respectively. This sampling strategy was designed to capture both the spatial variability across the agricultural fields and the vertical variability associated with soil redistribution processes and nutrient distribution. Although all samples were available for analysis, only a subset corresponding to half of the samples from each region was selected for GRS measurements. This reduction was primarily motivated by the long acquisition times required for gamma-ray spectrometry, together with the relatively large sample mass necessary to ensure adequate counting statistics, making the analysis of the complete dataset impractical. Figure 46 presents a satellite view of both study areas together with the corresponding sampling locations.

Figure 46 – Satellite view of Cambé (left) and Toledo (right) sampling sites. Sampling points and scale are indicated.



Source: self-elaboration.

7.3.1.1 Wet Chemistry

Soil chemical analysis is based on the determination of the levels at which nutrients are available to plants. This procedure is usually composed of two main steps: extraction and quantification. The extraction step is performed by adding chemical solutions to the analyzed sample in order to simulate root activity in the soil. Subsequently, the extracted nutrients are quantified using analytical techniques such as atomic absorption spectrophotometry and UV-Vis spectrophotometry [114–116].

The soil fertility parameters for the dataset were determined at the Soils Department of IDR-Paraná, and the results are summarized in Table 12.

Table 12 – Descriptive statistics of the fertility parameters in the measured samples ($n = 158$).

Parameter	**Unit**	**Mean**	**SD**	**Median**	**IQR**	**Min**	**Max**
Soil chemical properties							
pH	–	5.3	0.6	5.2	0.9	4.1	6.6
OC	$g\,kg^{-1}$	20.5	5.5	20.5	7.5	6.5	36.8
H+Al	$cmol_c\,kg^{-1}$	5.53	1.72	5.34	2.41	2.35	10.40
CEC	$cmol_c\,kg^{-1}$	13.9	1.7	13.9	2.1	9.4	20.5
BS	%	59.3	14.5	60.0	20.3	19.7	86.7
Exchangeable elements and fertility parameters							
exCa	$cmol_c\,kg^{-1}$	5.60	1.80	5.60	2.70	1.60	11.40
exMg	$cmol_c\,kg^{-1}$	2.36	1.04	2.24	1.12	0.57	6.66
exK	$cmol_c\,kg^{-1}$	0.44	0.27	0.41	0.38	0.05	1.15
exAl	$cmol_c\,kg^{-1}$	0.12	0.27	0.00	0.12	0.00	1.30
exP	$mg\,kg^{-1}$	24.3	22.1	19.1	22.1	0.7	114.4
SB	$cmol_c\,kg^{-1}$	8.43	2.77	8.31	4.04	2.39	17.78
SAL	$cmol_c\,kg^{-1}$	2.49	5.89	0.00	2.11	0.00	34.69
Particle size distribution							
Clay	$g\,kg^{-1}$	706	23	706	3.33	600	770
Silt	$g\,kg^{-1}$	133	14	133	3.20	100	190
Sand	$g\,kg^{-1}$	160	17	160	0.13	110	220

SD = Standard deviation, IQR = Interquartile range, Min = Minimum, Max = Maximum.

The parameters presented in Table 12 describe the chemical and textural attributes of the soil that directly influence nutrient availability and plant development. Exchangeable bases, such as Ca, Mg, and K (exCa, exMg, and exK), contribute to the soil base sum (SB), which reflects the availability of essential cations. Exchangeable aluminum (exAl) and potential acidity (H+Al) are indicators of soil acidity and aluminum toxicity, particularly relevant in highly weathered tropical soils. The cation exchange capacity (CEC) represents the soil's ability to retain and exchange nutrients, while base saturation (BS) expresses the proportion of exchange sites occupied by basic cations, serving as an im-

portant fertility index. Available phosphorus (exP) is a key macronutrient often limiting in tropical soils due to strong fixation processes. Soil organic carbon (OC) plays a fundamental role in nutrient cycling, aggregation, and chemical buffering. Finally, the textural fractions (sand, silt, and clay) provide information on soil physical structure and strongly influence water retention, nutrient dynamics, and sorption processes, while pH integrates several chemical properties affecting nutrient solubility and biological activity.

7.3.1.2 GRS

The GRS samples were prepared by sieving the soil through a 2 mm mesh and storing the material in sealed Petri dishes until secular equilibrium was achieved. This preparation ensured sample homogeneity while allowing the uranium and thorium decay series to reach radioactive equilibrium before the measurements. The samples were measured by placing the Petri dishes directly on top of an HPGe detector (Section 2.5.1) with a live time of $14,400$ s. Following the measurements, the activity concentrations of ^{40}K, ^{137}Cs, ^{232}Th, and ^{238}U were determined and are summarized in Table 13. As expected for soils from the Southern Hemisphere, ^{137}Cs was not detected in all samples, as several activity concentrations were below the MDA for this radionuclide.

Table 13 – Descriptive statistics of the radionuclides found in the measured samples. All values, except Counts, are reported in Bq/kg.

Radionuclide	**Counts**	**Mean**	**SD**	**Mean**	**IQR**	**Min**	**Max**
^{40}K	158	60	28.89	53.40	16.84	21.53	287.01
^{137}Cs	33	0.5	0.11	0.50	0.0	0.31	0.80
^{232}Th	158	29.93	17.65	25.53	16.01	8.68	125.81
^{238}U	158	26.79	16.12	21.13	13.13	9.66	116.97

SD = Standard deviation, IQR = Interquartile range, Min = Minimum, Max = Maximum.

7.3.1.3 XRF

The XRF measurements were performed using pelletized samples. Approximately 5 g of soil from each sample were milled in a ball mill to obtain a particle size smaller than 10 μm. An aliquot of 2.5 g of the milled material was then mixed with 0.25 g (10%) of a binder (Cereox) and pressed into pellets under a load of 10 tons for 90 s. Pelletization provides a flat and mechanically stable sample geometry, reducing the influence of particle-size effects and improving measurement repeatability. The measurements were carried out using the Panalytical Epsilon 5 spectrometer (Section 2.6.3). The resulting elemental concentrations obtained by quantitative XRF analysis are summarized in Table 14.

Table 14 – Descriptive statistics of the elements quantified in the measured samples (n = 158). Oxides are reported in %, while elemental concentrations are reported in ppm.

Analyte	**Mean**	**SD**	**Mean**	**IQR**	**Min**	**Max**
Na_2O	0.38	0.01	0.37	0.01	0.37	0.42
Al_2O_3	24.18	1.37	23.96	2.13	21.15	27.34
SiO_2	29.15	1.20	29.50	1.40	24.90	32.60
P_2O_5	0.14	0.04	0.14	0.04	0.05	0.24
SO_3	0.06	0.01	0.06	0.01	0.03	0.11
K_2O	0.11	0.03	0.11	0.05	0.06	0.21
Ti	28650	2585	28770	3670	20590	34690
V	760	130	760	210	550	1010
Cr	217	38	224	59	138	320
Mn	1690	645	1690	1255	810	2835
Fe_2O_3	30.50	2.05	31.10	3.90	27.00	34.30
Ni	45	9	49	14	23	59
Cu	397	77	414	127	228	524
Zn	175	36	175	64	102	237

SD = Standard deviation, IQR = Interquartile range, Min = Minimum, Max = Maximum.

7.3.2 MB-PCA

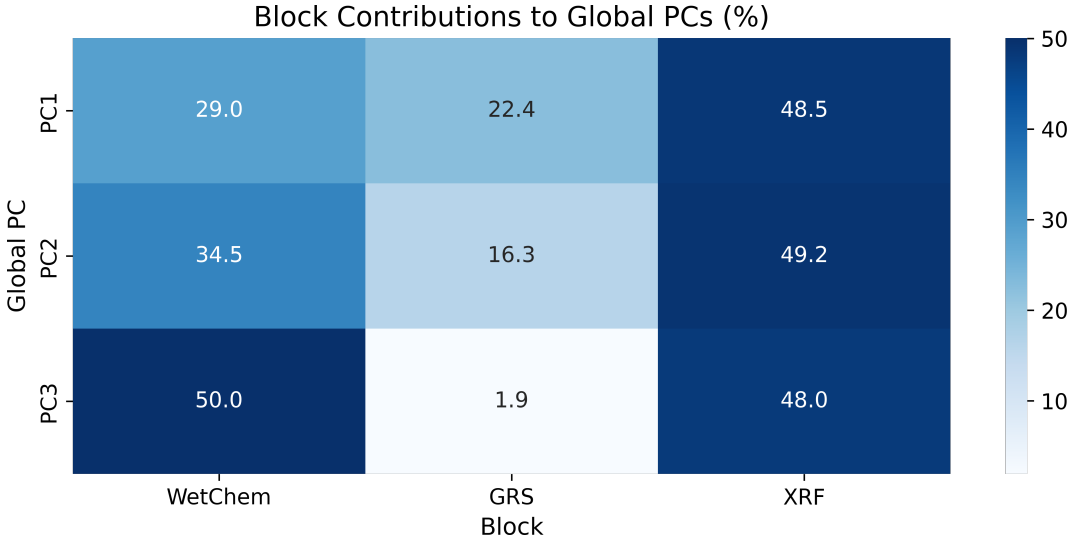
While each analytical technique individually provides valuable information, their combined interpretation requires multivariate approaches capable of handling heterogeneous yet complementary datasets. In this work, the three analytical blocks comprise measurements obtained from distinct physical principles and are expressed in different units and numerical ranges. Consequently, direct comparison between variables is not straightforward, and conventional univariate analyses are unable to fully exploit the relationships among the datasets. MB-PCA (Subsection 4.1.1.1) was therefore employed as an exploratory framework to investigate the shared and block-specific sources of variability while preserving the individual structure of each analytical block.

Within this framework, MB-PCA was used to explore the relationships between elemental information obtained from GRS and XRF measurements and fertility parameters derived from conventional wet chemistry analyses. The objective was to identify common trends, complementarities, and potential redundancies among the three analytical techniques, thereby evaluating the extent to which spectroscopic measurements reproduce or complement the information provided by laboratory-based soil fertility analyses.

Each individual block (GRS, XRF, and WetChem) was preprocessed using autoscaling, which centers each variable by subtracting its mean and scales it to unit variance. This preprocessing step minimizes the influence of differences in variable magnitude and measurement units, ensuring that all variables contribute equally to the analysis re-

gardless of their absolute concentration ranges. Autoscaling is particularly useful for modeling low-concentration components because it assigns the same weight to all variables in the dataset [57]. Subsequently, PCA was applied independently to each block ($k_b = 3$), and the resulting score matrices were concatenated into a global matrix on which a second PCA ($k_g = 3$) was performed to obtain the final multiblock representation.

Figure 47 – Heatmap of block contributions (%) to the global principal components obtained by MB-PCA. The color intensity reflects the relative importance of each data block in the formation of the global components.

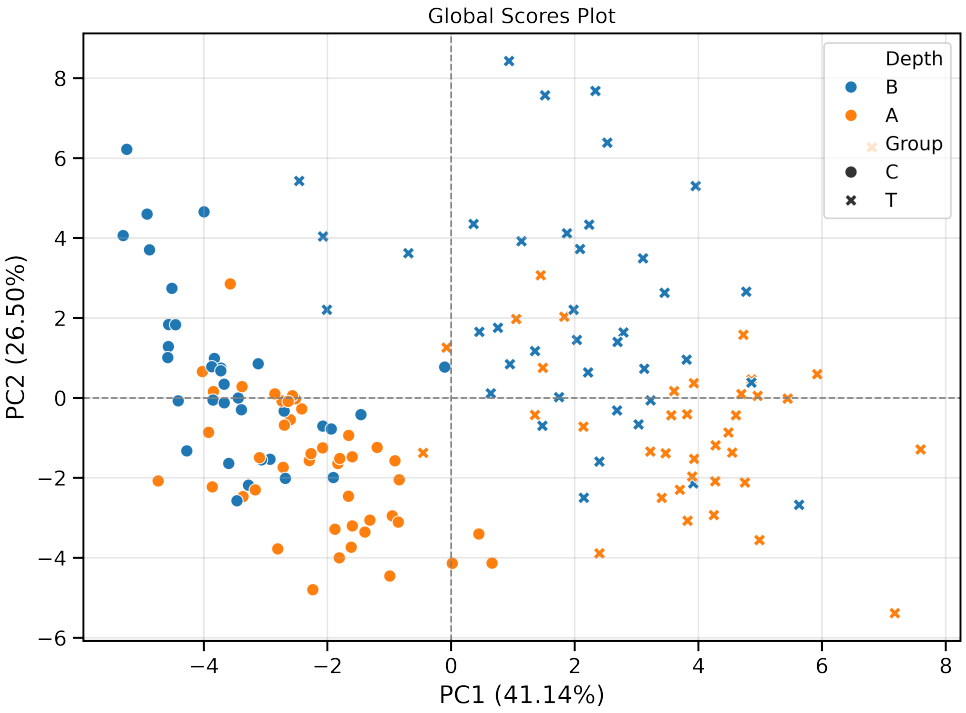


Source: self-elaboration.

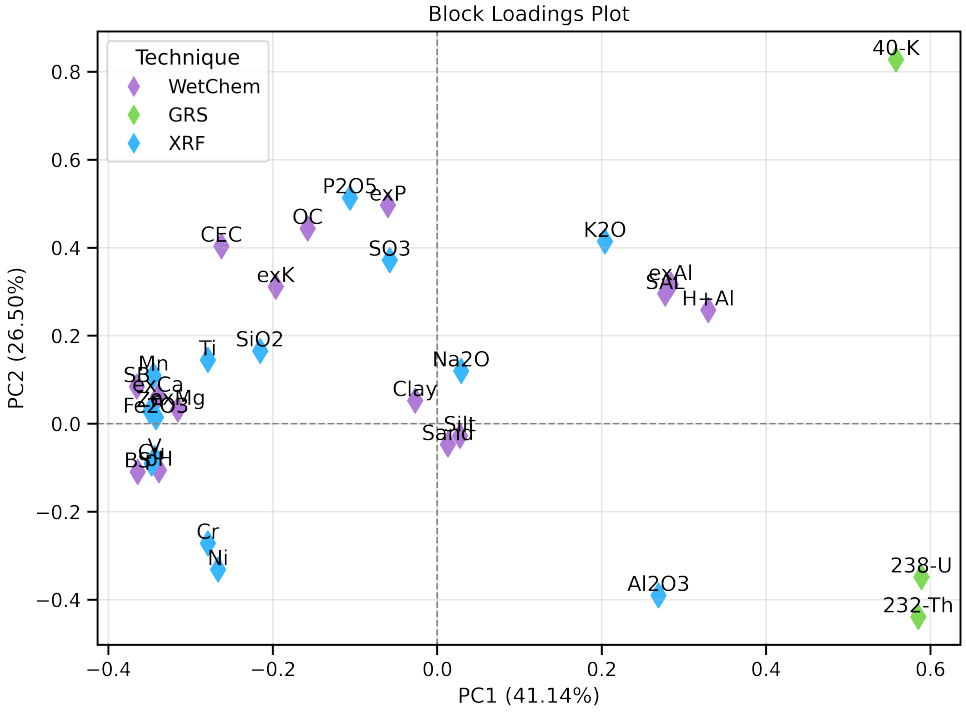
The heatmap in Figure 47 summarizes the relative contribution of each data block to the global principal components obtained by MB-PCA. The first global component receives contributions from all three datasets, although XRF represents the dominant source of variability, followed by WetChem and GRS. A similar trend is observed for the second component, where XRF remains the main contributor while the relative contribution of GRS decreases to approximately 16%. In contrast, the third global component is almost entirely explained by the WetChem and XRF blocks, each contributing close to 50%, whereas GRS plays only a marginal role. Overall, these results suggest that elemental composition measured by XRF consistently explains a substantial fraction of the shared variance, while GRS primarily contributes to the dominant component and has a more limited influence on higher-order variability. The distribution of the samples in the reduced multivariate space is presented in Figures 48 and 49, which show the PC1 versus PC2 and PC2 versus PC3 score plots, respectively.

Figure 48 – Scores (samples) and loadings (variables) plots for the first two principal components (PC1 and PC2) obtained from the MB-PCA of Wet Chemistry, GRS, and XRF datasets.

(a) Scores plot of PC1 versus PC2. Colors indicate sampling depth, while marker shapes denote the sampling site (C: Cambé; T: Toledo).



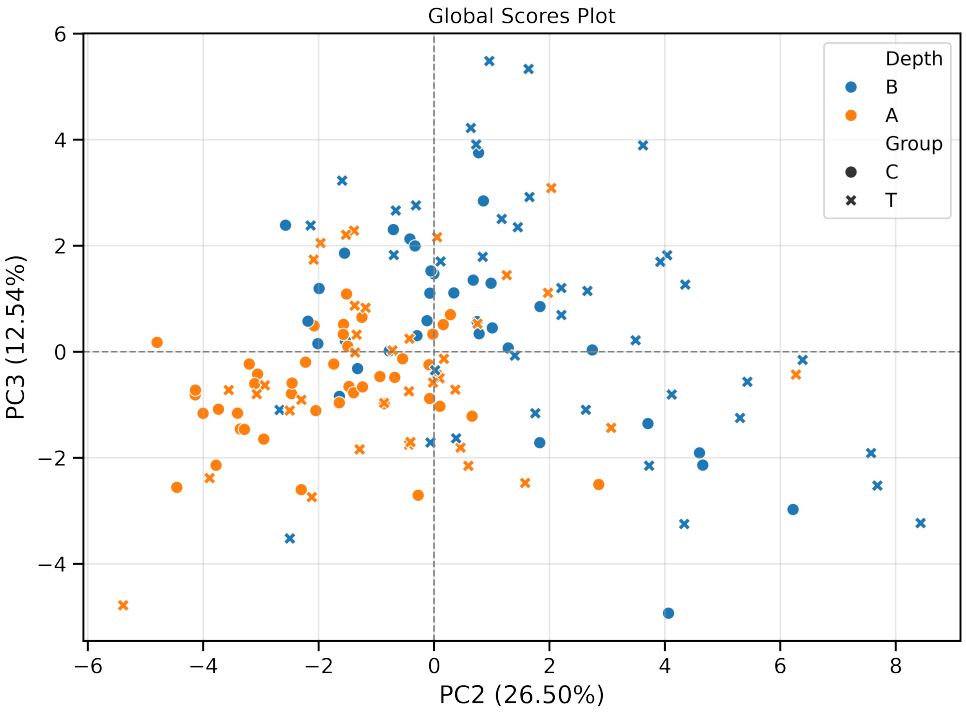
(b) Loadings plot of PC1 versus PC2. Marker colors indicate the data block from which each variable originates (Wet Chemistry, GRS, or XRF).



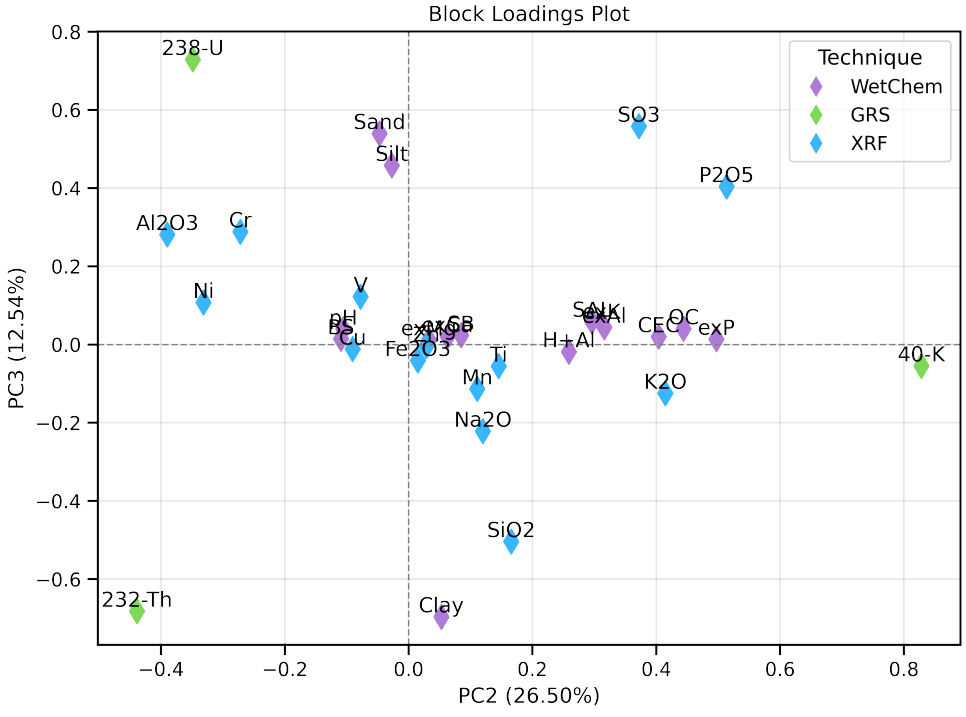
Source: self-elaboration.

Figure 49 – Scores (samples) and loadings (variables) plots for the second and the third principal components (PC2 and PC3) obtained from the MB-PCA of Wet Chemistry, GRS, and XRF datasets.

(a) Scores plot of PC2 versus PC3. Colors indicate sampling depth, while marker shapes denote the sampling site (C: Cambé; T: Toledo).



(b) Loadings plot of PC2 versus PC3. Marker colors indicate the data block from which each variable originates (Wet Chemistry, GRS, or XRF).



Source: self-elaboration.

The first two principal components capture the largest fraction of the shared variance across the data blocks and provide a suitable low-dimensional representation for evaluating sample relationships. Since it was possible to quantify ^{137}Cs in approximately 21% of the samples, it's contribution was not considered in the MB-PCA.

Upon inspection of Figure 48a, it is noticeable that two clusters are formed, distributed along PC1, separating the samples based on their sampling site. By comparing the distribution in the scores plot with the influence of each variable in the loadings plot, Figure 48b, the samples from Toledo (T), on the positive region of PC1, are characterized by high values of radioisotopes (^{40}K, ^{232}Th, and ^{238}U), and minerals like Al_2O_3 and CaO. In counterpart, the samples from Cambé (C) are characterized by higher concentration of heavy metals and transition elements (Cr, Ni, Mn, Zn, Fe, and V), from both WetChem and XRF blocks.

Moreover, a depth separation is noticeable along the PC2 axis. The most superficial samples (A), are associated with higher amounts of Al_2O_3 and heavier radionuclides, ^{232}Th, and ^{238}U. On the other hand, deeper samples (B) are associated with higher contents of Organic Carbon, MgO, and ^{40}K.

The relationships between soil fertility parameters (WetChem) and the elements from the GRS and XRF blocks may be inferred from the proximity of variables in both loading plots. Variables that cluster together are positively correlated, while those on opposite sides are negatively related. PC1 shows a strong correlation between basic fertility indicators, and the transition metals quantified by XRF. Indicators of soil base status, such as pH, BS, SB, exCa, and exMg, are directly associated with higher concentrations of Mn, Zn, Fe_2O_3, Cu, V, Ti, Cr, and Ni. This correlation indicates that the more neutral-pH soils are also the ones with the highest concentration of naturally occurring heavy metals.

PC2 indicates that soils with higher organic matter also show higher MgO and SiO_2 content, likely reflecting the soil mineralogical composition. Figure 48b also indicates that samples with high acidity-related parameters are typically those with higher amounts of Al_2O_3 and higher presence of radioisotopes.

The PC2 versus PC3 scores plot, illustrated in Figure 49a, also tends to separate the samples based on their sampling depth. Samples from depth A are clustered in the negative region of PC3, while depth B is clustered in the positive region. The loadings plot, in Figure 49b, indicates that this separation is driven by textural properties and specific elements signatures. Depth B is mostly associated with coarser particles, as indicated by the contributions from Sand and Silt, and higher concentrations of ^{238}U. In counterpart, depth A is characterized by the higher presence of Clay, P_2O_5, SO_3, and ^{232}Th. This way, PC3 represents mostly the textural gradient of the samples. Table 15 summarizes the relationships between the variables found with the MB-PCA.

Table 15 – Summary of the relationships found between WetChem parameters and elements quantified via XRF and GRS for the analyzed samples through MB-PCA.

WetChem Parameter	**Highly Correlated With**	**Characteristic**
pH, BS, SB, exCa, exMg	$Fe_2O_3, MnO, TiO_2, V, Cr, Ni$	Soil fertility / Basicity
OC, CEC, exP, exK	P_2O_5, SO_3	Organic matter & Nutrients
Sand, Silt	SiO_2, $^{238}U, Na_2O, MgO$	Coarse fraction
Clay	^{232}Th	Fine fraction / Weathering
exAl, H + Al, SAL	Al_2O_3, K_2O, ^{40}K	Acidity / Aluminosilicates

7.3.3 Conclusions

The joint analysis of wet chemistry, GRS, and XRF analytical data shed light on underlying relationships between soil fertility parameters and spectra-derived quantities. By performing MB-PCA with $k_g = 3$, the resulting principal components were able to differentiate the samples according to their source (PC1), depth of collection (PC2), and texture (PC3). The synergy between the different analytical techniques was assessed through the correlation of variable weights in the sample characterization process. In this context, soil fertility parameters exhibited strong correlations with the concentrations of transition metals observed in the analyzed samples, while organic matter content was highly correlated with magnesium, silica, and ^{40}K.

Regarding soil texture, sand and silt contents showed strong associations with ^{238}U and silica, whereas clay content was correlated with ^{232}Th, phosphorus, and sulfur. These relationships are particularly relevant in the context of soil erosion, as erosive processes tend to preferentially remove finer particles and organic matter, leading to measurable changes in both elemental composition and natural radionuclide signatures. Consequently, the observed multivariate patterns reflect not only intrinsic soil properties but also the effects of particle-size redistribution and nutrient depletion associated with erosion. Data-fusion approaches such as MB-PCA therefore provide a powerful framework for identifying erosion-related signatures by integrating complementary chemical, mineralogical, and radiometric information.

7.4 Data Fusion for Predicting Soil Fertility Parameters

Soil erosion is characterized by the loss of fertile soil that plants require to grow, leading to the irretrievable loss of nutrients [3, 5]. Therefore, the continuous monitoring of erosion and nutrient availability is essential for ensuring plant and soil health [78]. The determination of plant-available nutrients in soil is often performed by wet chemistry methods, as described in Subsection 7.3.1.1; however, spectra-based techniques coupled with machine learning models may also be employed to build predictive models of such

nutrients, yielding fast, accurate, and reliable results [50, 107]. As discussed in Chapter 5, several studies have explored the use of IR-based techniques, XRF, or their combination to predict soil fertility attributes. Based on this premise, the present work investigates the predictive power of MIRS and XRF for modelling the soil fertility parameters from the soil samples employed in Section 7.3.

In cases where standalone models did not provide satisfactory results, a fusion strategy was employed to combine atomic and molecular fingerprints and, through the synergy between XRF and MIRS measurements, yield robust models. Samples were analyzed using a Panalytical Epsilon 5 for XRF (Subsection 2.6.3), a Bruker Alpha II (Subsection 2.7.1) for MIRS, and the fertility parameters were measured by the IDR-Paraná using different wet chemistry techniques (Subsection 7.3.1.1). The ML models employed were PLSR (Subsection 4.2.1) and RFR (Subsection 4.2.2.)

7.4.1 Standalone Models

Standalone models for the thirteen soil fertility parameters listed in Table 12 were built and their performances were evaluated through their R^2, generalization gap (ΔR^2), MAE, RMSE, and ratio of performance deviation (RPD). The test results of the best model obtained for nine[1] parameters are listed in Table 16.

On both cases, the modelling steps consisted of splitting the dataset into training (70%) and testing (30%) sets using the Kennard-Stone algorithm, which ensures that both datasets are representative of the overall variability of the samples. The training set was used for model calibration and hyperparameter optimization, while the test set was reserved for evaluating the predictive performance of the final model on unseen data.

The preprocessing of the spectra consisted of: (*i*) trimming and feature selection (for XRF, this consisted in choosing the measurement condition that provided optimal quantification of the target element, as well as on the correlated parameters identified in Table 15); (*ii*) application of a Savitzky-Golay (SG) filter, with optimized parameters for each case; and (*iii*) autoscaling. The regression models were optimized by iterating through different hyperparameter configurations and selecting the one that minimized the RMSE[2]. For PLSR, this involved varying the number of latent components, while for RFR, the number of trees ($50 \leq n_{\text{trees}} \leq 300$), maximum tree depth ($10, 15, 20, 25, 30, \infty$[3]), minimum number of samples required for a split ($2, 5, 11$), and the number of features considered when searching for the best split ($n, \sqrt{n}, \log_2 n$).

[1] The results for clay, silt, sand, exAl, H+Al, and SAL are not listed because: (*i*) The physical fractions exhibited a highly homogeneous distribution; and (*ii*) the datasets exhibited a high frequency of null or near-zero values for Al-related parameters.

[2] Mean computational time per model: 20 s for PLSR, and 5 min for RFR on an Intel Core i5–10300H CPU @ 2.3 GHz with 16 GB RAM.

[3] Unconstrained, i.e. no predefined maximum depth.

To quantify the model's predictive utility, the RPD was calculated. The RPD is a statistical parameter often evaluated in soil sensor calibration, as it relates the standard deviation of the measured values to the standard error of the prediction. According to the classification proposed by Rossel et al. [117], RPD values below 1.4 indicate models that are very poor for prediction, values below 1.8 indicates models that are suitable primarily for screening purposes, whereas values above 2.0 reflect reliable quantitative predictions, and values above 2.5 indicates excellent modelling performance.

Model robustness was assessed through the generalization gap, defined here as the difference between calibration and test R^2 values (ΔR^2). This metric provides insight into the stability of the regression model and its susceptibility to overfitting. In chemometric applications, ΔR^2 values below approximately 0.10 are generally considered indicative of acceptable generalization, whereas gaps exceeding 0.15 suggest that the model may be capturing dataset-specific noise rather than physically meaningful relationships. Larger discrepancies ($\Delta R^2 > 0.20$) are commonly associated with severe overfitting and reduced predictive transferability to independent datasets [118].

Table 16 – Standalone XRF and MIRS regression results corresponding to the test datasets of the nine modelled fertility parameters.

Parameter	**Data**	**Reg.**	**R^2**	**MAE**	**RMSE**	**RPD**	**ΔR^2**	**Fusion?**
OC	MIRS	PLS	0.94	0.93	1.26	4.02	0.02	No
	XRF	RF	0.37	2.62	3.35	1.26	0.46	
exCa	MIRS	PLS	0.91	0.38	0.49	3.42	0.07	No
	XRF	RF	0.89	0.43	0.53	3.11	0.04	
exK	MIRS	PLS	0.26	0.17	0.19	1.16	0.08	**Yes**
	XRF	RF	0.49	0.13	0.16	1.41	0.30	
exMg	MIRS	PLS	0.84	0.30	0.38	2.49	0.05	No
	XRF	RF	0.45	0.54	0.69	1.34	0.35	
exP	MIRS	RF	0.14	9.44	12.10	1.08	0.61	**Yes**
	XRF	RF	0.46	0.22	0.27	1.35	0.42	
SB	MIRS	PLS	0.94	0.48	0.60	4.31	0.02	No
	XRF	RF	0.83	0.67	0.84	2.42	0.13	
CEC	MIRS	PLS	0.74	0.67	0.79	1.98	0.14	No
	XRF	RF	0.41	0.90	1.06	1.29	0.45	
BS	MIRS	PLS	0.85	3.68	4.60	2.06	0.05	No
	XRF	RF	0.81	4.19	5.27	2.33	0.13	
pH	MIRS	RF	0.62	0.29	0.36	1.62	0.27	**Yes**
	XRF	PLS	0.65	0.22	0.27	1.69	0.22	

RPD = Ratio of Performance Deviation, ΔR^2 = Generalization gap. These results are further detailed in Appendix B.

MIRS yielded strong predictive accuracy for parameters such as OC, SB, BS, and exchangeable Mg, with high R^2 values and RPD values exceeding 2, indicating satisfactory to excellent model performance. In contrast, parameters including exchangeable Ca, K, and P, as well as pH and CEC, exhibited limited predictive capability, characterized by low R^2 values (< 0.6), RPD values below 1.8, and large generalization gaps (ΔR^2). The elevated ΔR^2 values indicate reduced generalization performance and suggest that the spectral information provided by MIRS alone is insufficient to robustly capture the variability of these attributes, which motivates the use of data fusion strategies to improve both predictive accuracy and model stability for the affected parameters. Data fusion was considered justified for parameters exhibiting either insufficient predictive accuracy (RPD < 1.8) or unstable generalization performance ($\Delta R^2 > 0.15$), as these conditions indicate limited reliability of the standalone model.

The standalone XRF models exhibited strong predictive performance for inorganic soil fertility parameters, particularly exchangeable Ca, SB, and BS, with R^2 values above 0.80 and satisfactory generalization behavior, as indicated by comparatively low ΔR^2 values (Table 16). Exchangeable Ca and SB were accurately predicted, likely due to the strong contribution of Ca-bearing minerals and related geochemical associations to the XRF signal. Exchangeable P, in particular, showed improvement when modelled using XRF, which reflects the direct sensitivity of XRF to elemental composition but is heavily overfitted, as indicated by the high ΔR^2 value.

Nevertheless, XRF performance remained limited for attributes more strongly associated with organic composition or indirect physicochemical processes, such as OC, exchangeable Mg, CEC, and pH. These parameters presented either low predictive accuracy (RPD < 1.8) or elevated generalization gaps ($\Delta R^2 > 0.15$), indicating unstable model behavior and reduced robustness when relying solely on XRF-derived information. Organic carbon, for example, yielded poor predictive performance ($R^2 = 0.37$), which is consistent with the inability of XRF to directly detect low atomic number elements such as C under typical soil measurement conditions. Similarly, pH and CEC are indirectly related to elemental composition and depend on complex interactions involving mineralogy, organic matter, and surface chemistry, which may not be fully represented by XRF spectra alone.

Data fusion was therefore considered for exK, exP, and pH, as these parameters exhibited either insufficient standalone predictive capability or unstable generalization performance in both of the individual spectroscopic approaches. In these cases, combining molecular information from MIRS with elemental information from XRF was expected to improve both predictive accuracy and model robustness by integrating complementary sources of soil chemical and mineralogical information.

7.4.2 Data Fusion Strategy

The adopted data fusion approach consisted of a low-level fusion, by concatenating the preprocessed spectral data from MIRS and XRF. The last step prior to data fusion was scaling spectral blocks by their respective maximum values to ensure comparable signal amplitudes and prevent dominance of one technique over the other. In both strategies, the resulting fused matrix was split into training (70%) and testing (30%) datasets using the Kennard-Stone algorithm to ensure homogeneity between both datasets, based on the values obtained for the soil fertility parameters (Table 12). The regression models were optimized following the same procedure described for the standalone models, with the same hyperparameter search space.

7.4.3 Soil Macronutrients Prediction

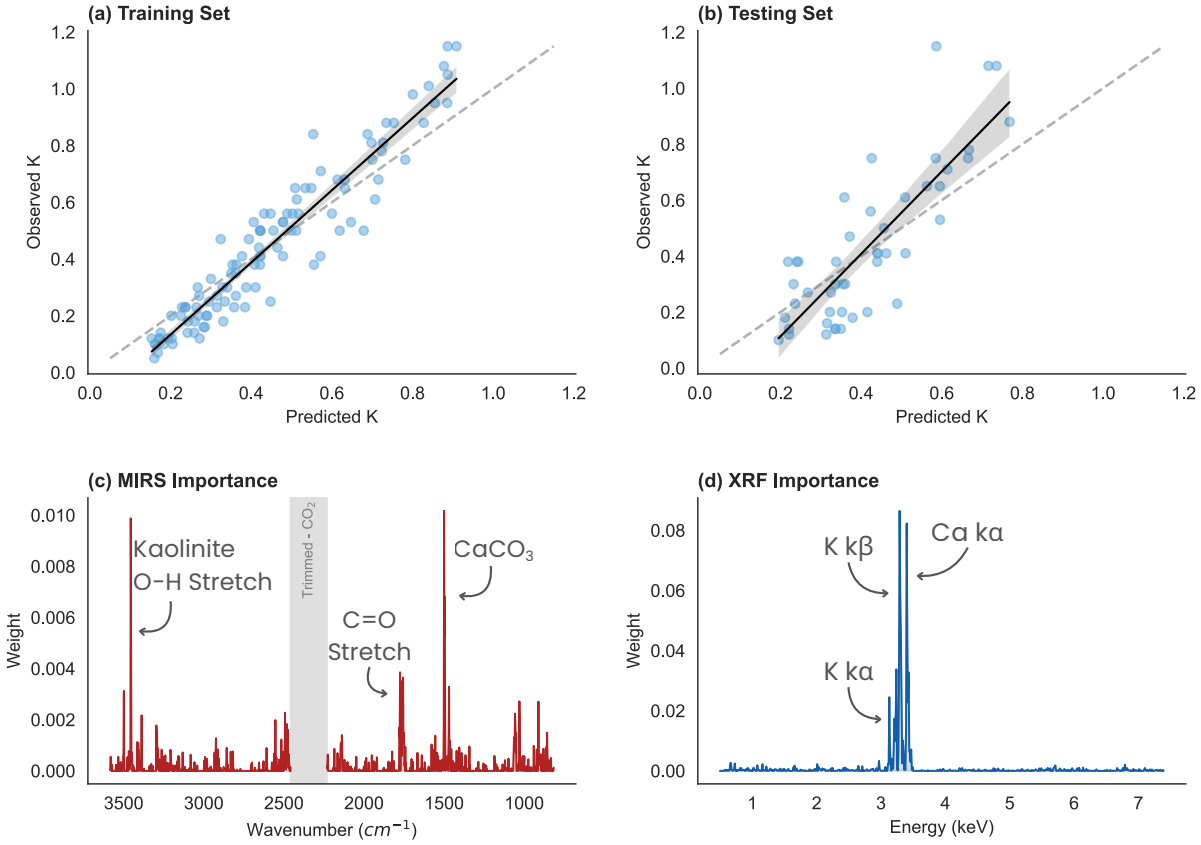
The determination of soil macronutrient levels is important for increased crop and plant production. The measurement of such parameters allow decision-makers to optimize fertilizer input in soil, leading to lower production costs [119, 120]. The mean R^2 values reported in the literature for exK and exP employing MIRS measurements are 0.37, and 0.35, respectively [121]. Comparing these values with the ones obtained during for standalone MIRS, detailed in Table 16, it is noticeable that exP modeling results are in general accordance with the literature, while the ones obtained for exK are well below.

7.4.3.1 exK

Potassium is an essential macronutrient for plants nutrition, whose modelling based on MIRS measurements usually yields poor results [121]. Despite the fact that K depends on soil mineralogical composition, the exchangeable fraction of K is influenced by factors that are not visible through in the mid-infrared range. Moreover, K is often externally added to agricultural sites as fertilizer, and is leached with more ease than Ca [121, 122]. These factors contribute to the limited predictive performance of MIRS for exK, as reflected by the low R^2 value and high generalization gap (Table 16). The fusion strategy for exK consisted of integrating: (*i*) XRF-CaF_2 spectra, trimmed at 4.3 keV and preprocessed with a first-derivative SG filter; and (*ii*) MIRS spectra, trimmed to the 1100–3500 cm^{-1} range and preprocessed with a first-derivative SG filter. The regression results, employing RFR, are displayed in Figure 50.

These features and preprocessing choices were designed to incorporate geochemically meaningful elemental information from XRF and exclude uninformative regions of the MIRS spectra while preserving diagnostically relevant O−H stretching bands and silicate lattice vibrations to which K^+ is structurally or chemically associated.

Figure 50 – RFR results for soil exK prediction based on XRF and MIRS. Predicted versus measured values are shown for (a) training and (b) test datasets. Individual variable importance for (c) MIRS and (d) XRF. Performance metrics for both datasets are summarized below.



	R2	MAE	RMSE	RPD
Train	0.862	0.083	0.102	2.695
Test	0.608	0.133	0.17	1.598

Method: RF | *Parameters: n_est=230, depth=10, leaf=5, max_feat=0.31, max_samp=0.84*

Source: self-elaboration.

Compared to the standalone models (Table 16), the fused model achieved a substantial improvement in both accuracy and robustness. The RFR model yielded R^2 values of 0.862 and 0.608 for the training and testing sets, respectively, accompanied by reduced prediction errors and increased RPD values, indicating a transition from unreliable to screening-level performance.

Feature importance analysis indicated that the fused model leveraged complementary information from both techniques. The MIRS importance plot revealed contributions from spectral regions associated with kaolinite C–O stretching (3600–3200 cm^{-1}) and $CaCO_3$ bending (1650–1500 cm^{-1}) vibrations, which are consistent with the known association of exchangeable K with clay minerals. The XRF importance plot highlighted the relevance of the K Kα emission line at approximately 3.31 keV, confirming that total K content contributes significantly to the predictive model. The intrinsic correlation between K and Ca in the studied soils, as indicated by the correlation analysis in Table 15, may also explain the importance of the Ca Kα line at approximately 3.69 keV, as it likely provides indirect information related to K availability through geochemical associations. The integration of these complementary sources of information enabled the model to capture both the elemental abundance and the mineralogical context influencing exchangeable K levels, overcoming the limitations observed when standalone MIRS- and XRF-based models were employed 16.

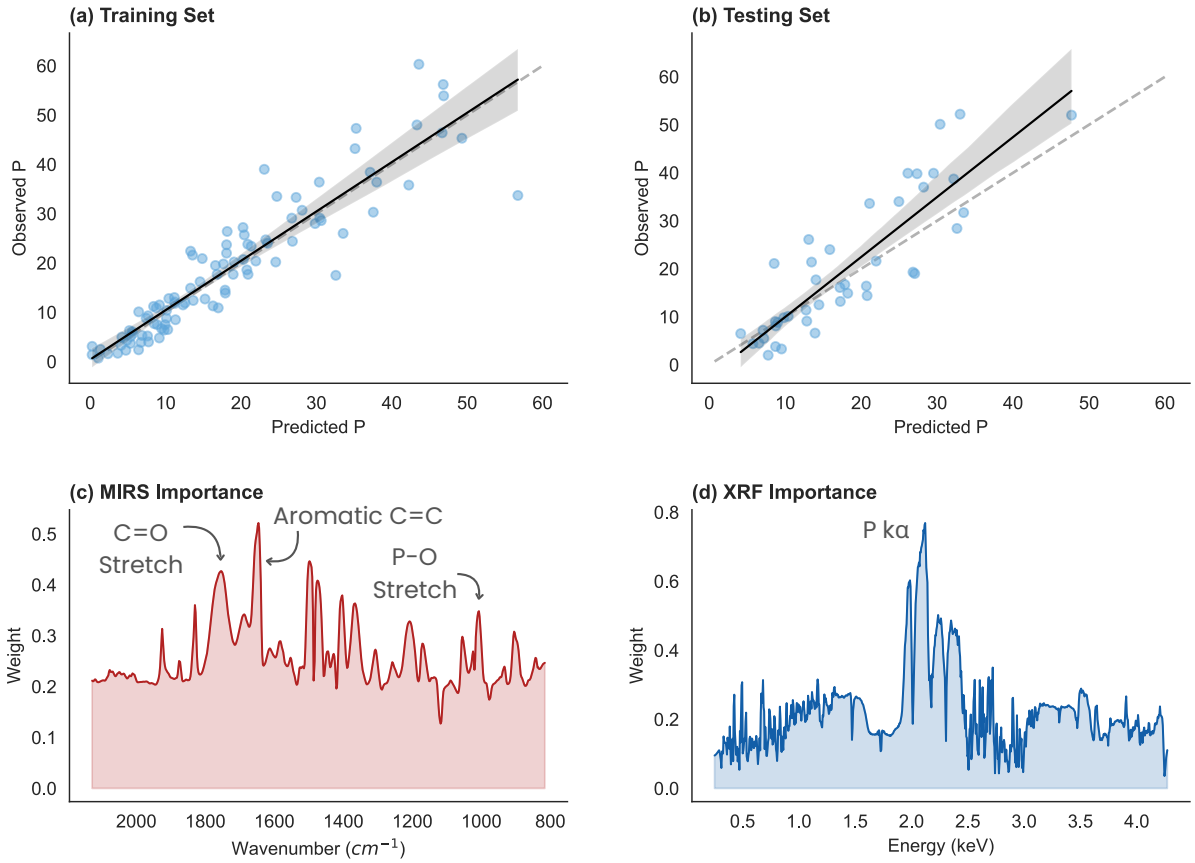
7.4.3.2 exP

Phosphorus is required in substantial quantities to sustain adequate plant growth and development [121, 123, 124]. However, the predictive performance of spectra-based models for exchangeable P was consistently poor, as reflected by low or unreliable R^2 values. This limitation is likely associated with the weak and non-specific mid-infrared absorption features of phosphorus-bearing species, whose vibrational bands often overlap with dominant signals from other soil components [125]. Such spectral interference reduces the sensitivity of MIRS to variations in exchangeable P content.

The fusion approach for modelling exP consisted of: (*i*) XRF-CaF_2 spectra, trimmed at 4.3 keV and a first-derivative SG filter; and (*ii*) MIRS spectra, trimmed to the fingerprint region (2200-800 cm^{-1}), with a first-derivative SG filter. The regression results, employing PLSR, are displayed in Figure 51.

While the standalone MIRS approach yielded a highly overfitted model ($\Delta R^2 = 0.61$), the fused PLSR model achieved substantially improved performance, with a test R^2 of 0.70 and an RPD of 1.83, which is considered suitable for screening purposes. The reduced generalization gap ($\Delta R^2 = 0.14$) further indicates improved model stability, suggesting that the integration of complementary sensor information contributes to a more robust prediction of exP.

Figure 51 – PLSR results for exP prediction fusing MIRS and XRF spectral datasets. Predicted versus measured values are shown for the (a) training and (b) test datasets. Individual variable importance for (c) MIRS and (d) XRF. Performance metrics for both datasets are summarized below.



	R2	MAE	RMSE	RPD
Train	0.851	3.655	5.317	2.592
Test	0.71	5.684	7.578	1.857

Method: PLS | Parameters: n_components=5

Source: self-elaboration.

The MIRS importance plot indicates strong contributions from spectral regions associated with C=O, C=C, and P–O vibrational modes, particularly around 1720–1600 cm^{-1} and 1200–900 cm^{-1}. The bands in the 1720–1600 cm^{-1} region are commonly associated with carbonyl and aromatic functional groups linked to soil organic matter, whereas the 1200–900 cm^{-1} region includes phosphate-related vibrations and overlapping Si–O stretching modes from clay minerals. These regions may therefore reflect interactions between phosphorus and soil organic constituents, as well as adsorption processes involving organo-mineral complexes. Although the direct spectral response of exchangeable P is gen-

erally weak, these associated soil components likely provide indirect information linked to P availability.

In contrast, the XRF importance plot indicates the relevance of the P Kα emission line at approximately 2.01 keV, indicating that total phosphorus content contributes significantly to the predictive model. This behavior suggests that, although XRF primarily reflects total elemental concentrations, the total P pool still retains some relationship with the exchangeable fraction within the studied soils. Consequently, the fused approach appears to benefit from the complementary nature of the techniques: while XRF captures elemental abundance, MIRS provides information related to the molecular and mineralogical environment influencing phosphorus retention and availability.

7.4.4 Soil pH Prediction

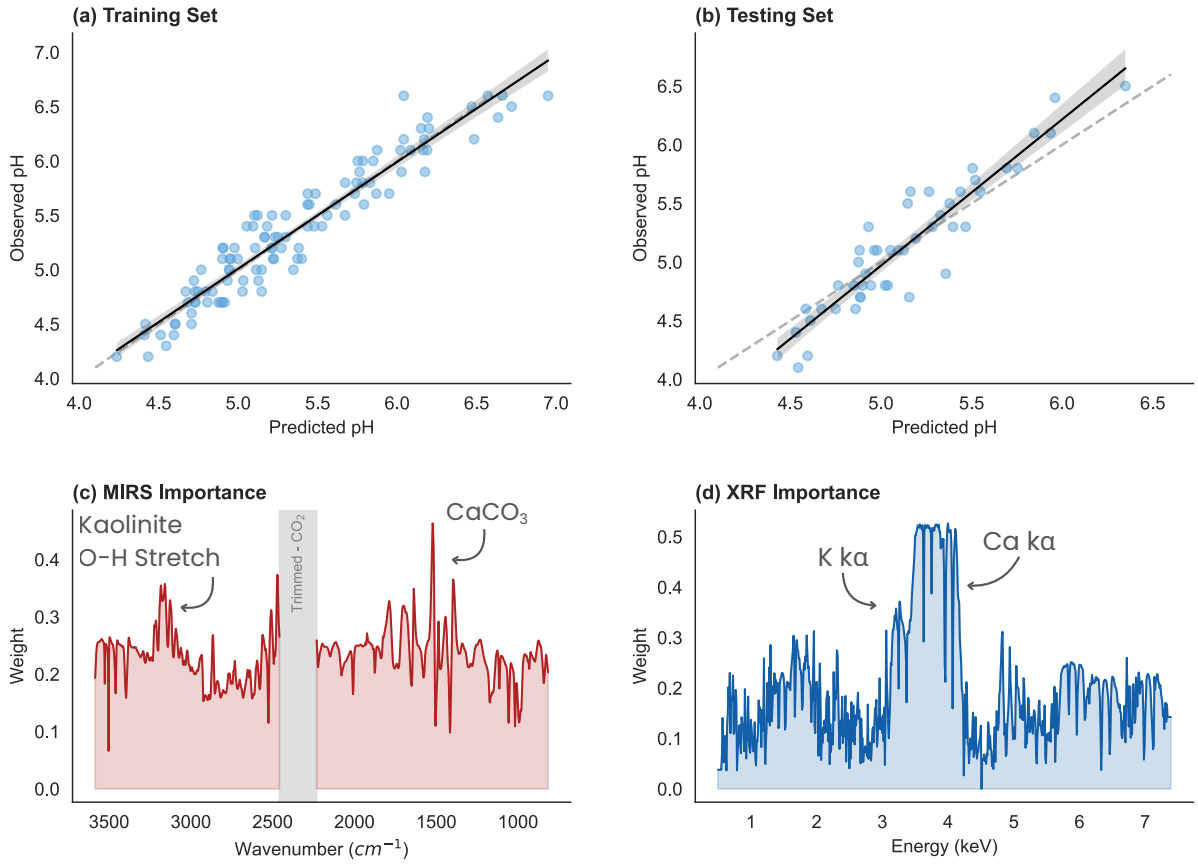
The estimation of soil pH through infrared spectroscopy is a recognized challenge in soil science, often yielding R^2 values below 0.70 in the literature. Despite these values, models are frequently considered acceptable for practical soil assessment if the RMSE remains below 0.31 [121]. The standalone MIRS model developed in this work, utilizing RFR, achieved performance metrics consistent with these established studies, as described in Table 16.

Since soil pH is intrinsically linked to the presence and mobility of transition metals and oxide phases, as evidenced by the correlation analysis in Table 15, the XRF data obtained using the Fe secondary target were selected for the fusion process. The resulting integrated model (Figure 52) was designed to jointly evaluate the organic and mineralogical signatures captured by MIRS and the elemental sensitivity of XRF, thereby integrating complementary physical perspectives of soil acidity.

The fused data blocks and preprocessing steps were as follows: (*i*) XRF-Fe spectra, trimmed at 7.2 keV and preprocessed using a first-derivative SG filter; and (*ii*) MIRS spectra, also preprocessed with a first-derivative SG filter. This consistent derivative treatment enhances subtle spectral variations associated with mineralogical differences while minimizing baseline effects prior to modelling.

Compared to the standalone MIRS model (Table 16), the fused model exhibited an improvement in predictive performance. The PLSR model yielded R^2 values of 0.908 and 0.839 for the training and testing sets, respectively, accompanied by low prediction errors (RMSE of 0.185 and 0.223) and RPD values of 3.299 and 2.494. These results indicate excellent calibration performance and robust predictive capability in the independent test set, reaching analytical-quality standards. The low generalization gap further suggests stable model behavior and limited overfitting.

Figure 52 – PLSR results for soil pH prediction based on XRF and MIRS spectra. Predicted versus measured values are shown for the training (left) and test (right) datasets. Variable importance and performance metrics for both datasets are summarized below.



	R2	MAE	RMSE	RPD
Train	0.908	0.152	0.185	3.299
Test	0.839	0.179	0.223	2.494

Method: PLS | Parameters: n_components=5

Source: self-elaboration.

The feature contribution profile indicates that both XRF-Fe and MIRS variables contributed substantially to model performance, reinforcing the complementary nature of elemental and spectroscopic information. While MIRS captures vibrational signatures associated with clay minerals, hydroxyl groups, and organic matter, XRF provides direct sensitivity to the elemental composition of oxides. The fusion strategy therefore enhances the model's capacity to represent both mineralogical structure and geochemical controls on soil acidity.

Overall, these results demonstrate that data fusion significantly improves pH prediction compared to standalone approaches, yielding a model with high accuracy, strong generalization capacity, and physicochemical interpretability.

7.4.5 Conclusions

The results presented in this section demonstrate that data fusion effectively overcomes the limitations observed in standalone spectroscopic modelling. Parameters such as exP, exK, and pH showed substantial improvements under the fused framework. In the case of exP, the validation R^2 increased from 0.14 to 0.70, representing a transition from non-predictive to screening-level performance. For soil pH, the integrated model achieved analytical-quality performance, with a validation R^2 of 0.839.

The observed gains were not simply a consequence of increasing the volume of input variables, but rather resulted from the integration of geochemically meaningful information. The inclusion of raw spectral data from atomic and molecular sensors provided the contextual framework necessary to capture nutrient dynamics in highly weathered soils by giving context to the elemental composition through mineralogical and organic matter associations.

Fusion strategies also led to improvements in model robustness. While standalone XRF and MIRS models for exP exhibited severe overfitting ($\Delta R^2 = 0.61$ and $\Delta R^2 = 0.42$, respectively), the fused PLSR model reduced the generalization gap to 0.144, indicating improved stability and transferability to independent datasets. This reduction highlights the role of complementary data sources in constraining model variance and mitigating noise-driven calibration. The feature importance analysis further supports the notion that the fused models leverage synergistic information from both techniques, with MIRS capturing mineralogical and organic matter signatures and XRF providing elemental sensitivity.

Overall, these findings highlight the value of data fusion in soil spectroscopy, demonstrating that the integration of complementary sensor information can yield substantial improvements in both predictive accuracy and model robustness for key soil fertility parameters.

8 FINAL CONCLUSIONS

This thesis demonstrated that the integration of Gamma-Ray Spectrometry (GRS), X-ray Fluorescence (XRF), and Mid-Infrared Spectroscopy (MIRS), supported by multivariate analysis, provides a comprehensive framework for soil erosion assessment and soil fertility characterization. By combining nuclear, atomic, and molecular-level techniques, this work progressed from methodological optimization in radiation detection to the development of predictive models capable of assessing the agronomic consequences of soil redistribution.

The first stage of this research focused on improving the accessibility and practicality of radionuclide-based soil erosion monitoring. Although HPGe detectors offer superior spectral resolution, their high cost and operational complexity limit their broader application. It was demonstrated that mathematical signal-processing approaches, particularly Savitzky-Golay smoothing and spectral deconvolution, enable NaI(Tl) scintillation detectors to quantify ^{7}Be with performance comparable to HPGe systems [86]. Furthermore, the application of Target Transformation Factor Analysis (TTFA) significantly improved calibration transfer between detectors and reduced the Minimum Detectable Activity (MDA) for radionuclides such as ^{137}Cs [91]. These findings establish that reliable radionuclide quantification can be achieved using more affordable and operationally accessible instrumentation, expanding the feasibility of large-scale erosion monitoring. While these results are specific to the employed detector configurations and soil profiles, they provide a foundation for future research to evaluate the generalizability of these mathematical approaches across different environmental conditions.

The second stage of the thesis explored the integration of multiple spectrometric techniques to enhance the physical interpretation of soil redistribution processes. By applying multivariate analysis to fused datasets, it was possible to identify patterns and associations that were not evident when analyzing GRS and XRF data separately. Three approaches were employed: (*i*) Principal Component Analysis (PCA) was performed on analytical datasets from GRS and XRF to identify common patterns in radionuclide and elemental distributions, which indicated similar behavior between the vertical distribution of ^{137}Cs and elements commonly introduced through fertilization, such as P, K, and Ca; (*ii*) Common Dimensions (ComDim) analysis was used to extract shared components across fused pXRF and GRS raw spectra, allowing for the identification of joint patterns in the spectrometric signatures that can effectively capture redistribution-related variability; and (*iii*) Multi-Block PCA (MB-PCA) was applied to fused datasets of GRS, XRF, and wet chemistry data to explore the relationship between radionuclide inventories, elemental composition, and soil properties, which revealed strong relationships between soil

texture, fertility indicators, and radionuclide distribution. However, because these specific associations are tied to the soil types and management practices of the study area, further research is required to evaluate their stability across different contexts and to uncover the mechanisms driving these underlying signatures.

The final stage of this research addressed the consequences of erosion by focusing on soil fertility prediction. Through the fusion of MIRS and XRF raw spectra, predictive models were developed that substantially outperformed standalone approaches. Parameters that originally exhibited limited performance, displayed significant improvement within the fused framework. For exchangeable phosphorus, the validation R^2 increased from 0.14 and 0.46 for standalone XRF and MIRS models, respectively, to 0.701, representing a transition from non-predictive behavior to screening-level applicability. Soil pH achieved analytical-quality performance, with a validation R^2 of 0.839 and strong model stability. These improvements were accompanied by a significant reduction in generalization gaps, demonstrating enhanced robustness and transferability. Despite these promising results, the models were developed and validated using a specific dataset from a highly weathered tropical soil. Future research should evaluate the performance of these fused models across different soil types and environmental conditions to confirm their generalizability and to further refine the modeling approaches for broader applicability.

Overall, this thesis demonstrates that the strategic integration of complementary spectrometric techniques, combined with multivariate analysis, enhances both the physical interpretation and predictive capability of soil studies. By reducing reliance on high-cost instrumentation and labor-intensive wet chemistry, this work demonstrates that integrated spectrometric approaches can provide reliable and scalable alternatives for assessing soil erosion and nutrient redistribution. The progression from detector optimization to multisensor data fusion and fertility prediction highlights the potential of integrated spectrometric approaches to support more sustainable soil management in highly weathered tropical landscapes.

To build upon the conclusions drafted from this thesis, the next step involves shifting from localized point-source characterization to continuous landscape-scale assessment. Future research should leverage the nature of the optimized GRS-XRF-MIRS models to explore the spatial structure and directional dependencies of soil redistribution and fertility parameters. By coupling these spectrometric datasets with geographic information systems, topographic attributes, and geostatistical interpolation, it will be possible to model the spatial vulnerability of agricultural areas. Classifying erosion-prone regions and identifying areas susceptible to severe nutrient loss will provide land managers with the precise spatial intelligence required to implement targeted, site-specific conservation practices.

BIBLIOGRAPHY

[1] MABIT, L. et al. 137 cs: A widely used and validated medium term soil tracer. *IAEA TECDOC SERIES*, n. 1741, p. 27, 2014.

[2] ASTORGA, R. T. et al. Exploring innovative techniques for identifying geochemical elements as fingerprints of sediment sources in an agricultural catchment of argentina affected by soil erosion. *Environmental Science and Pollution Research*, Springer, v. 25, p. 20868–20879, 2018.

[3] LAL, R. Soil management in the developing countries. *Soil Science*, LWW, v. 165, n. 1, p. 57–72, 2000.

[4] MABIT, L. et al. The use of 7 be as a short term soil redistribution tracer. *IAEA TECDOC SERIES*, n. 1741, p. 105, 2014.

[5] TARAKANOV, V. *What is Soil Erosion? How Can Nuclear Techniques Help to Identify and Mitigate it?* [S.l.]: International Atomic Energy Agency, 2022.

[6] Food and Agriculture Organization. *Global Symposium on Soil Erosion: Key Messages*. 2019. Disponível em: <https://www.fao.org/about/meetings/soil-erosion-symposium/key-messages/en/>.

[7] ARATA, L. et al. Modelling deposition and erosion rates with radionuclides (modern)–part 1: A new conversion model to derive soil redistribution rates from inventories of fallout radionuclides. *Journal of environmental radioactivity*, Elsevier, v. 162, p. 45–55, 2016.

[8] RITCHIE, J. C.; MCHENRY, J. R. Application of radioactive fallout cesium-137 for measuring soil erosion and sediment accumulation rates and patterns: A review. *Journal of environmental quality*, Wiley Online Library, v. 19, n. 2, p. 215–233, 1990.

[9] WALLING, D.; HE, Q. Use of fallout 137cs in investigations of overbank sediment deposition on river floodplains. *Catena*, Elsevier, v. 29, n. 3-4, p. 263–282, 1997.

[10] ANDRELLO, A. C.; APPOLONI, C. R.; FILHO, V. F. d. N. Assesment of soil erosion by 137cs technique in native forests in londrina city, parana, brazil. *Brazilian archives of biology and technology*, SciELO Brasil, v. 50, p. 1051–1060, 2007.

[11] MABIT, L. et al. Assessment of soil erosion and sedimentation: the role of fallout radionuclides. *IAEA TECDOC SERIES*, n. 1741, 2014.

[12] BENMANSOUR, M. et al. The use of excess 210 pb (210 pbex) as a soil and sediment tracer. *IAEA TECDOC SERIES*, v. 79, n. 1741, 2014.

[13] COHEN, B. L. *Concepts of Nuclear Physics*. [S.l.]: McGraw-Hill, 1971.

[14] GARCÍA-LEÓN, M. Environmental radioactivity. In: *Detecting Environmental Radioactivity*. [S.l.]: Springer, 2022. p. 51–103.

[15] GRIEKEN, R. V.; MARKOWICZ, A. *Handbook of X-ray Spectrometry*. [S.l.]: CRC press, 2001.

[16] BERGER, M. et al. *XCOM: Photon Cross Sections Database*. [S.l.]: National Institute of Standards and Technology (NIST), 2010. <https://physics.nist.gov/xcom>. NIST Standard Reference Database 8 (XGAM).

[17] LEO, W. R. *Techniques for nuclear and particle physics experiments: a how-to approach*. [S.l.]: Springer Science & Business Media, 1994.

[18] ANDERSON, D. W. *Absorption of ionizing radiation*. [S.l.]: University Park Press, 1984.

[19] GILMORE, G. *Practical gamma-ray spectroscopy*. [S.l.]: John Wiley & Sons, 2008.

[20] SEGRÈ, E. *Nuclei and Particles: An Introduction to Nuclear and Subnuclear Physics*. [S.l.]: W.A. Benjamin, 1965.

[21] KAPLAN, I. *Nuclear Physics*. [S.l.]: Addison-Wesley, 1963. (Addison-Wesley series in nuclear science and engineering).

[22] CURRIE, L. A. Limits for qualitative detection and quantitative determination. application to radiochemistry. *Analytical chemistry*, ACS Publications, v. 40, n. 3, p. 586–593, 1968.

[23] SITKO, R.; ZAWISZA, B. Quantification in x-ray fluorescence spectrometry. *X-ray spectroscopy*, IntechOpen London, p. 137–162, 2012.

[24] JENKINS, R. *Quantitative X-ray spectrometry*. [S.l.]: CRC Press, 1995.

[25] SHIMADZU. *Energy Dispersive X-Ray Fluorescence Spectrometer - EDX Series*. [S.l.].

[26] BRUKER. *PXRF Tracer 5 - Portal XRF Analyzer for Curtting-Edge Researchers*. [S.l.].

[27] Malvern Panalytical. *Epsilon 5 User Guide*. [S.l.].

[28] STUART, B. H. *Infrared spectroscopy: fundamentals and applications*. [S.l.]: John Wiley & Sons, 2004.

[29] GRIFFITHS, P. R. Fourier transform infrared spectrometry. *Science*, American Association for the Advancement of Science, v. 222, n. 4621, p. 297–302, 1983.

[30] MARK, H. Chemometrics in near-infrared spectroscopy. *Analytica chimica acta*, Elsevier, v. 223, p. 75–93, 1989.

[31] Bruker Optics GmbH. *FT-IR Spectroscopy: Alpha II*. [S.l.].

[32] CABRERA, M. et al. Using 137cs and 210pbex to assess soil redistribution at different temporal scales along with lithogenic radionuclides to evaluate contrasted watersheds in the uruguayan pampa grassland. *Geoderma*, Elsevier, v. 435, p. 116502, 2023.

[33] WHITE, W. M. *Isotope geochemistry*. [S.l.]: John Wiley & Sons, 2023.

[34] DICKIN, A. P. *Radiogenic isotope geology*. [S.l.]: Cambridge university press, 2018.

[35] KASTE, J. M.; NORTON, S. A.; HESS, C. T. Environmental chemistry of beryllium-7. *Reviews in Mineralogy and Geochemistry*, Mineralogical Society of America, v. 50, n. 1, p. 271–289, 2002.

[36] PAPASTEFANOU, C.; IOANNIDOU, A. Aerodynamic size association of 7be in ambient aerosols. *Journal of Environmental Radioactivity*, Elsevier, v. 26, n. 3, p. 273–282, 1995.

[37] SEPULVEDA, A. et al. Use of 7be to document soil erosion associated with a short period of extreme rainfall. *Journal of Environmental Radioactivity*, Elsevier, v. 99, n. 1, p. 35–49, 2008.

[38] TAYLOR, A. et al. The use of be-7 as a soil and sediment tracer. *Assessing Recent Soil Erosion Rates through the Use of Beryllium-7 (Be-7)*, Springer International Publishing, p. 1–13, 2019.

[39] CARTER, M. W.; MOGHISSI, A. A. Three decades of nuclear testing. *Health Physics*, LWW, v. 33, n. 1, p. 55–71, 1977.

[40] BERNARD, C. et al. Césium-137 et érosion des sols. *Cahiers Agricultures*, v. 7, n. 3, p. 179–186, 1998.

[41] WALLING, D.; HE, Q.; ZHANG, Y. Conversion models and related software. *IAEA TECDOC SERIES*, n. 1741, 2014.

[42] ANDRELLO, A. C. Aplicabilidade do 137cs para medir erosão do solo: modelos teóricos e empíricos. 2004.

[43] FERREIRA, M. M. C. *Quimiometria: conceitos, métodos e aplicações*. [S.l.]: Editora da UNICAMP, 2015.

[44] FERREIRA, M. M. C. Quimiometria iii-revisitando a análise exploratória dos dados multivariados. *Química Nova*, SciELO Brasil, v. 45, n. 10, p. 1251–1264, 2022.

[45] PEARSON, K. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, Taylor & Francis, v. 2, n. 11, p. 559–572, 1901.

[46] HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, Warwick & York, v. 24, n. 6, p. 417, 1933.

[47] PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

[48] MAZEROLLES, G. et al. Common components and specific weights analysis: A chemometric method for dealing with complexity of food products. *Chemometrics and Intelligent Laboratory Systems*, Elsevier, v. 81, n. 1, p. 41–49, 2006.

[49] HANAFI, M. Nouvelles propriétés de l'analyse en composantes communes et poids spécifiques. *Journal de la Société française de statistique & Revue de statistique appliquée*, v. 149, n. 2, p. 75–97, 2008.

[50] RIBEIRO, J. V. et al. Optimization of pxrf instrumentation conditions and multivariate modeling in soil fertility attributes determination. *Spectrochimica Acta Part B: Atomic Spectroscopy*, Elsevier, p. 106835, 2023.

[51] SANTOS, F. R. dos et al. Data fusion of xrf and vis-nir using p-comdim to predict some fertility attributes in tropical soils derived from basalt. *Microchemical Journal*, Elsevier, v. 191, p. 108813, 2023.

[52] GHAZIRI, A. E. et al. Analysis of multiblock datasets using comdim: Overview and extension to the analysis of (k+ 1) datasets. *Journal of Chemometrics*, Wiley Online Library, v. 30, n. 8, p. 420–429, 2016.

[53] GALVAN, D. et al. Dd-comdim: A data-driven multiblock approach for one-class classifiers. *Chemometrics and Intelligent Laboratory Systems*, Elsevier, v. 233, p. 104748, 2023.

[54] QANNARI, E. M. et al. Defining the underlying sensory dimensions. *Food Quality and Preference*, Elsevier, v. 11, n. 1-2, p. 151–154, 2000.

[55] MISHRA, P. et al. Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC Trends in Analytical Chemistry*, Elsevier, v. 137, p. 116206, 2021.

[56] LESNOFF, M. *Jchemo: Julia package for machine learning and chemometrics on high-dimensional data.* 2021. Disponível em: <https://github.com/mlesnoff/Jchemo>.

[57] HUANG, Y. Chemometric methods in analytical spectroscopy technology. In: *Chemometric methods in analytical spectroscopy technology*. [S.l.]: Springer, 2022. p. 1–29.

[58] BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.

[59] JR, J. W.; MARK, H. Calibration transfer chemometrics, part i: Review of the subject. MJH Life Sciences, 2017.

[60] WANG, Y.-B.; YUAN, H.-F.; LU, W.-z. A new calibration transfer method based on target factor analysis. *Spectroscopy and Spectral Analysis*, v. 25, n. 3, p. 398–401, 2005.

[61] LI, J.; HEAP, A. D. A review of spatial interpolation methods for environmental scientists. Geoscience Australia Canberra, 2008.

[62] KARYDAS, C. G. et al. Evaluation of spatial interpolation techniques for mapping agricultural topsoil properties in crete. *EARSeL eProceedings*, v. 8, n. 1, p. 26–39, 2009.

[63] SHAHINZADEH, N. et al. Spatial variability of soil properties determined by the interpolation methods in the agricultural lands. *Modeling Earth Systems and Environment*, Springer, v. 8, n. 4, p. 4897–4907, 2022.

[64] THÉVENAZ, P.; BLU, T.; UNSER, M. Image interpolation and resampling. *Handbook of medical imaging, processing and analysis*, v. 1, n. 1, p. 393–420, 2000.

[65] BHUNIA, G. S.; SHIT, P. K.; MAITI, R. Comparison of gis-based interpolation methods for spatial distribution of soil organic carbon (soc). *Journal of the Saudi Society of Agricultural Sciences*, Elsevier, v. 17, n. 2, p. 114–126, 2018.

[66] VIRTANEN, P. et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, Nature Publishing Group, v. 17, n. 3, p. 261–272, 2020.

[67] TOY, T. J.; FOSTER, G. R.; RENARD, K. G. *Soil erosion: processes, prediction, measurement, and control*. [S.l.]: John Wiley & Sons, 2002.

[68] ZAPATA, F. The use of environmental radionuclides as tracers in soil erosion and sedimentation investigations: recent advances and future developments. *Soil and Tillage Research*, Elsevier, v. 69, p. 3–13, 2002.

[69] United Nations Educational, Scientific and Cultural Organization. *UNESCO raises global alarm on the rapid degradation of soils*. 2024. Disponível em: <https://www.unesco.org/en/articles/unesco-raises-global-alarm-rapid-degradation-soils>.

[70] FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. *The future of food and agriculture: Trends and challenges*. [S.l.]: Fao, 2017.

[71] BERNARD, J.; IIVARI, T. Sediment damages and recent trends in the united states. *Int. J. Sediment Res*, v. 15, n. 2, p. 135–148, 2000.

[72] PARWADA, C.; CHIPOMHO, J.; TIBUGARI, H. Comparison of conventional and artificial fallout radionuclide (frns) methods in assessing soil erosion. *Sustainable Environment*, Taylor & Francis, v. 9, n. 1, p. 2236406, 2023.

[73] ROGOWSKI, A. S.; TAMURA, T. Movement of 137cs by runoff, erosion and infiltration on the alluvial captina silt loam. *Health Physics*, LWW, v. 11, n. 12, p. 1333–1340, 1965.

[74] BLAKE, W.; WALLING, D.; HE, Q. Fallout beryllium-7 as a tracer in soil erosion investigations. *Applied radiation and Isotopes*, Elsevier, v. 51, n. 5, p. 599–605, 1999.

[75] SCHULLER, P.; WALLING, D. The use of 7 be and 137 cs in soil redistribution investigations in chile. *IAEA TECDOC SERIES*, p. 161.

[76] MEUSBURGER, K. et al. Combined use of 137cs and stable isotopes to assess soil degradation in mountainous grasslands of switzerland. 2011.

[77] CASTILHOS, N. D. de et al. X-ray fluorescence and gamma-ray spectrometry combined with multivariate analysis for topographic studies in agricultural soil. *Applied Radiation and Isotopes*, Elsevier, v. 95, p. 63–71, 2015.

[78] HANCOCK, G. et al. Soil organic carbon and soil erosion–understanding change at the large catchment scale. *Geoderma*, Elsevier, v. 343, p. 60–71, 2019.

[79] BÜCHELE, D. et al. Multivariate chemometrics as a key tool for prediction of k and fe in a diverse german agricultural soil-set using edxrf. *Scientific reports*, Springer, v. 9, n. 1, p. 1–11, 2019.

[80] SANTOS, F. R. dos et al. Edxrf spectral data combined with plsr to determine some soil fertility indicators. *Microchemical Journal*, Elsevier, v. 152, p. 104275, 2020.

[81] FERNÁNDEZ, Z. H. et al. Edxrf as an alternative method for multielement analysis of tropical soils and sediments. *Environmental monitoring and assessment*, Springer, v. 189, p. 1–9, 2017.

[82] KANIU, M. et al. Feasibility for chemometric energy dispersive x-ray fluorescence and scattering (edxrfs) spectroscopy method for rapid soil quality assessment. *X-Ray Spectrometry*, Wiley Online Library, v. 40, n. 6, p. 432–440, 2011.

[83] SILVA, F. M. et al. Tropical soil toposequence characterization via pxrf spectrometry. *Soil Science Society of America Journal*, Wiley Online Library, v. 83, n. 4, p. 1153–1166, 2019.

[84] SINGH, V. et al. Elemental profile of agricultural soil by the edxrf technique and use of the principal component analysis (pca) method to interpret the complex data. *Applied Radiation and Isotopes*, Elsevier, v. 69, n. 7, p. 969–974, 2011.

[85] LELAGO, A.; BIBISO, M. Performance of mid infrared spectroscopy to predict nutrients for agricultural soils in selected areas of ethiopia. *Heliyon*, Elsevier, v. 8, n. 3, 2022.

[86] LOPES, J. M. et al. Spectral deconvolution approach for beryllium-7 detection: Evaluating soil redistribution with nai(tl) detectors. *Radiation Physics and Chemistry*, v. 225, p. 112153, 2024.

[87] IAEA. *Guidelines for using fallout radionuclides to assess erosion and effectiveness of soil conservation strategies*. Vienna, Austria, 2014. ISBN: 978-92-0-145610-9.

[88] EMBRAPA SOLOS. Sistema brasileiro de classificação de solos. *Centro Nacional de Pesquisa de Solos: Rio de Janeiro*, v. 5, 2018.

[89] SCHAD, P. The international soil classification system wrb, 2014. *Novel Methods for Monitoring and Managing Land and Water Resources in Siberia*, Springer, p. 563–571, 2016.

[90] JR, J. J. W. A review of calibration transfer practices and instrument differences in spectroscopy. *Applied Spectroscopy*, Sage Publications, v. 72, n. 3, p. 340–365, 2018.

[91] LOPES, J. M. et al. Calibration transfer from HPGe to NaI(Tl) detectors for radionuclides quantification. *Journal of Environmental Radioactivity*, v. 293, p. 107885, fev. 2026. ISSN 0265931X.

[92] BOJANOWSKI, R. et al. *Report on the intercomparison run for the determination of radionuclides in soils. IAEA-326 and IAEA-327*. [S.l.], 2001.

[93] STRACHNOV, V. et al. *Report on the intercomparison run IAEA-312 Ra-226, Th and U in soil*. Vienna, Austria, 1991.

[94] SANSONE, U.; SHAKHASHIRO, A. *Certified Reference Materials for Radioactivity Measurements in Environmental Samples of Soil and Water: IAEA-444 and IAEA-445*. Vienna, Austria, 2011.

[95] HARRIS, C. R. et al. Array programming with NumPy. *Nature*, Springer Science and Business Media LLC, v. 585, n. 7825, p. 357–362, set. 2020. Disponível em: <https://doi.org/10.1038/s41586-020-2649-2>.

[96] LOPES, J. M. F. et al. Gamma-rays and x-rays spectrometries applied to evaluate soil redistribution. *Journal of Environmental Radioactivity*, v. 278, p. 107479, 2024.

[97] UEHARA, G. Acric properties and their significance to soil classification. In: EMBRAPA, SNLCS RIO DE JANEIRO. *International Soil Classification Workshop, 8*. [S.l.], 1988.

[98] ALLEONI, L. R. F.; CAMARGO, O. A. d. Óxidos de ferro e de alumínio e a mineralogia da fração argila deferrificada de latossolos ácricos. *Scientia Agricola*, SciELO Brasil, v. 52, p. 416–421, 1995.

[99] ALCARDE, J. C. Fertilizantes. In: *Fertilidade do Solo*. [S.l.]: Sociedade Brasileira de Ciencia do Solo, 2007. cap. 12, p. 737–768.

[100] LOPES, J. M. et al. Data fusion of gamma and x-ray spectra via comdim for soil characterization. *Soil Science Society of America Journal*, 2026.

[101] SANTOS, F. R. dos et al. Comparison between energy dispersive x-ray fluorescence spectral data and elemental data for soil attributes modelling. *Spectrochimica Acta Part B: Atomic Spectroscopy*, Elsevier, v. 185, p. 106303, 2021.

[102] GUTKNECHT, W. et al. Harmonization of uncertainties of x-ray fluorescence data for $pm_{2.5}$ air filter analysis. *Journal of the Air & Waste Management Association*, Taylor & Francis Ltd., v. 60, n. 2, p. 184, 2010.

[103] SINGH, B.; GILKES, R. Properties and distribution of iron oxides and their association with minor elements in the soils of south-western australia. *Journal of soil Science*, Wiley Online Library, v. 43, n. 1, p. 77–98, 1992.

[104] SILVA, S. H. G. et al. Soil texture prediction in tropical soils: A portable x-ray fluorescence spectrometry approach. *Geoderma*, Elsevier, v. 362, p. 114136, 2020.

[105] TAVARES, T. R. et al. Multi-sensor approach for tropical soil fertility analysis: comparison of individual and combined performance of vnir, xrf, and libs spectroscopies. *Agronomy*, MDPI, v. 11, n. 6, p. 1028, 2021.

[106] GREDILLA, A. et al. Non-destructive spectroscopy combined with chemometrics as a tool for green chemical analysis of environmental samples: A review. *TrAC Trends in Analytical Chemistry*, Elsevier, v. 76, p. 30–39, 2016.

[107] TAVARES, T. R. et al. Assessing soil key fertility attributes using a portable x-ray fluorescence: A simple method to overcome matrix effect. *Agronomy*, MDPI, v. 10, n. 6, p. 787, 2020.

[108] TAVARES, T. R. et al. Analysis of total soil nutrient content with x-ray fluorescence spectroscopy (xrf): Assessing different predictive modeling strategies and auxiliary variables. *AgriEngineering*, MDPI, v. 5, n. 2, p. 680–697, 2023.

[109] TAVARES, T. R. et al. Estimating plant-available nutrients with xrf sensors: Towards a versatile analysis tool for soil condition assessment. *Geoderma*, Elsevier, v. 439, p. 116701, 2023.

[110]ANTONANGELO, J.; ZHANG, H. Assessment of portable x-ray fluorescence (pxrf) for plant-available nutrient prediction in biochar-amended soils. *Scientific Reports*, Nature Publishing Group UK London, v. 14, n. 1, p. 20377, 2024.

[111]RIBEIRO, J. V. et al. Xrf and gamma-ray data fusion for predicting key soil fertility attributes. *Radiation Physics and Chemistry*, Elsevier, v. 234, p. 112750, 2025.

[112]NAWAR, S. et al. Fusion of gamma-rays and portable x-ray fluorescence spectral data to measure extractable potassium in soils. *Soil and Tillage Research*, Elsevier, v. 223, p. 105472, 2022.

[113]CASTILHOS, N. D. de et al. X-ray fluorescence and gamma-ray spectrometry combined with multivariate analysis for topographic studies in agricultural soil. *Applied Radiation and Isotopes*, Elsevier, v. 95, p. 63–71, 2015.

[114]PREZOTTI, L. C.; GUARÇONI, A. Guia de interpretações de análise de solo e foliar. Vitória, ES: Incaper, 2013., 2013.

[115]SOBRAL, L. F. et al. Guia prático para interpretação de resultados de análises de solos. 2015.

[116]RIBEIRO, J. V. *Analysis of fertility attributes in Eutroferric Red Latosol and Red Nitosol by portable EDXRF*. Dissertação (Mestrado) — Londrina State University, 2024.

[117]ROSSEL, R. V.; MCGLYNN, R.; MCBRATNEY, A. Determining the composition of mineral-organic mixes using uv–vis–nir diffuse reflectance spectroscopy. *Geoderma*, Elsevier, v. 137, n. 1-2, p. 70–82, 2006.

[118]HASTIE, T. et al. *The elements of statistical learning*. [S.l.]: Springer series in statistics New-York, 2009.

[119]KIM, H.-J.; SUDDUTH, K. A.; HUMMEL, J. W. Soil macronutrient sensing for precision agriculture. *Journal of Environmental Monitoring*, Royal Society of Chemistry, v. 11, n. 10, p. 1810–1824, 2009.

[120]DU, C.; ZHOU, J. Evaluation of soil fertility using infrared spectroscopy–a review. *Climate Change, Intercropping, Pest Control and Beneficial Microorganisms: Climate change, intercropping, pest control and beneficial microorganisms*, Springer, p. 453–483, 2009.

[121]SORIANO-DISLA, J. M. et al. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied spectroscopy reviews*, Taylor & Francis, v. 49, n. 2, p. 139–186, 2014.

[122]WHITEHEAD, D. C. *Nutrient elements in grassland: soil-plant-animal relationships*. [S.l.]: Cabi, 2000.

[123]SINFIELD, J. V.; FAGERMAN, D.; COLIC, O. Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils. *Computers and Electronics in Agriculture*, Elsevier, v. 70, n. 1, p. 1–18, 2010.

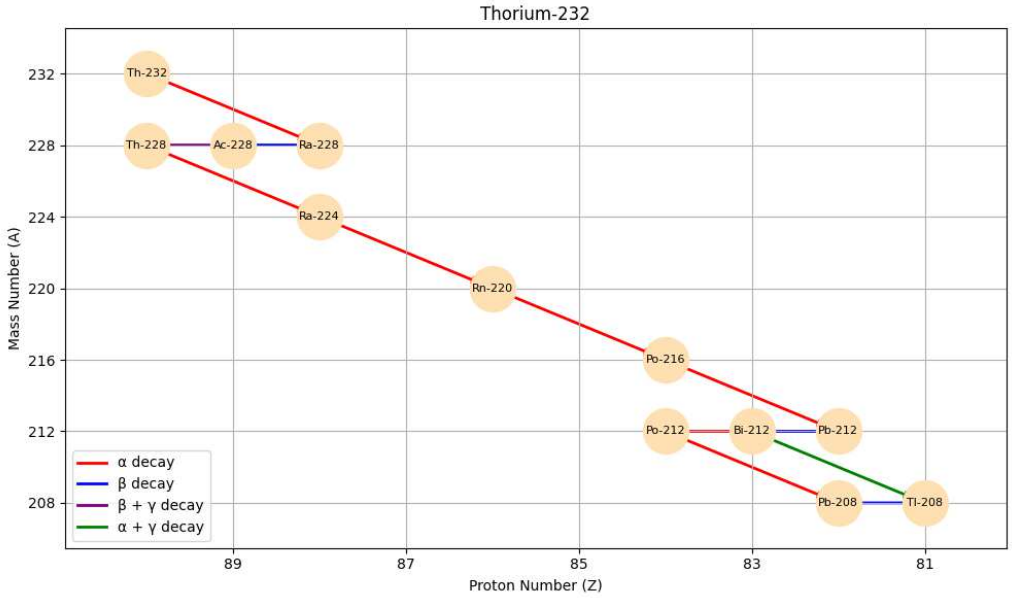
[124]MALEKI, M. et al. Phosphorus sensing for fresh soils using visible and near infrared spectroscopy. *Biosystems Engineering*, Elsevier, v. 95, n. 3, p. 425–436, 2006.

[125]KIZEWSKI, F. et al. Spectroscopic approaches for phosphorus speciation in soils and other environmental systems. *Journal of Environmental Quality*, Wiley Online Library, v. 40, n. 3, p. 751–766, 2011.

A DECAY CHAINS

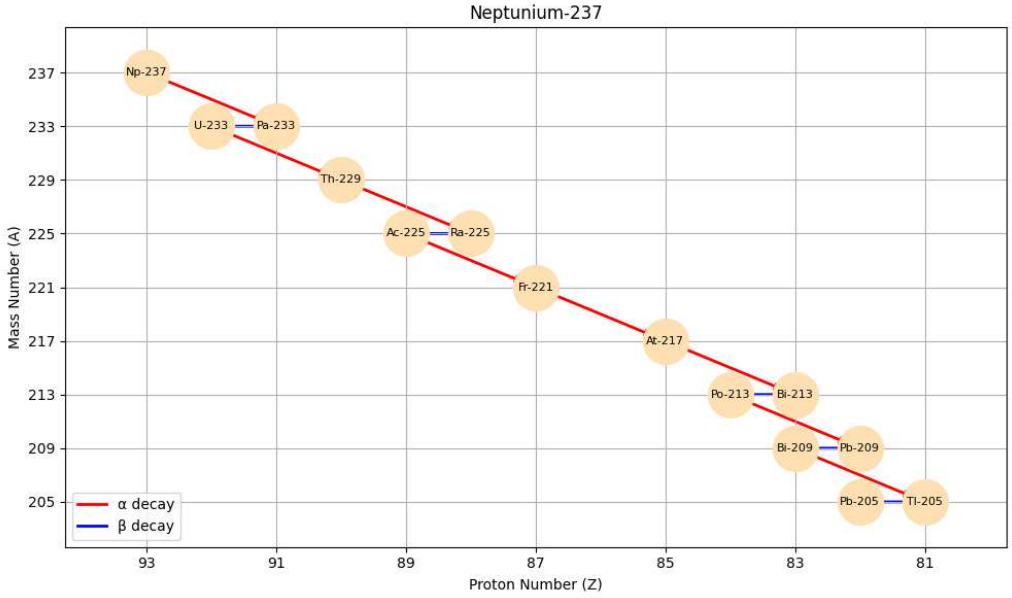
A.1 Thorium-232

Figure 53 – Representation of the ^{232}Th decay chain.



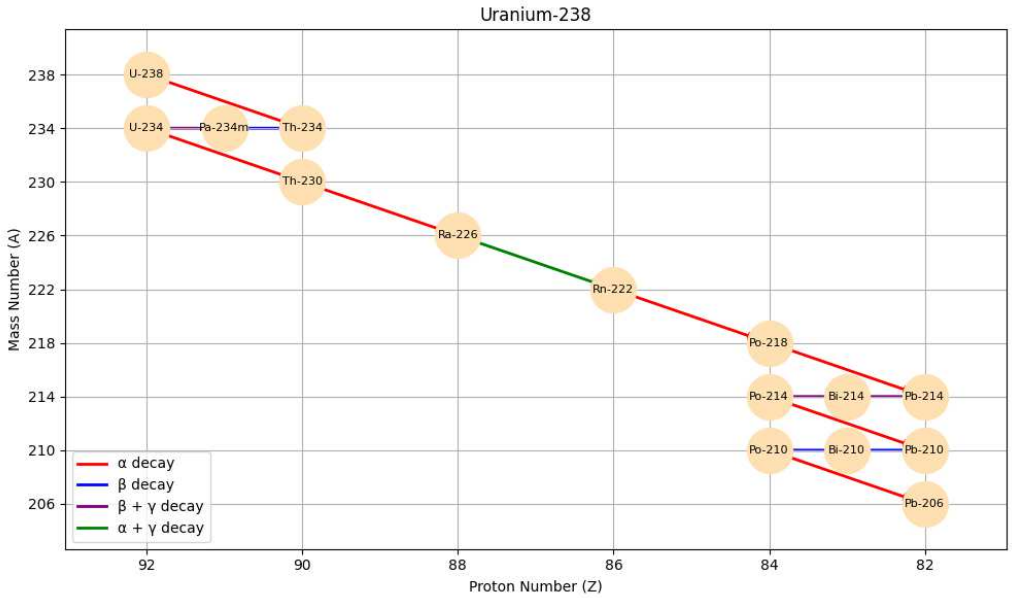
A.2 Neptunium-237

Figure 54 – Representation of the ^{237}Np decay chain.



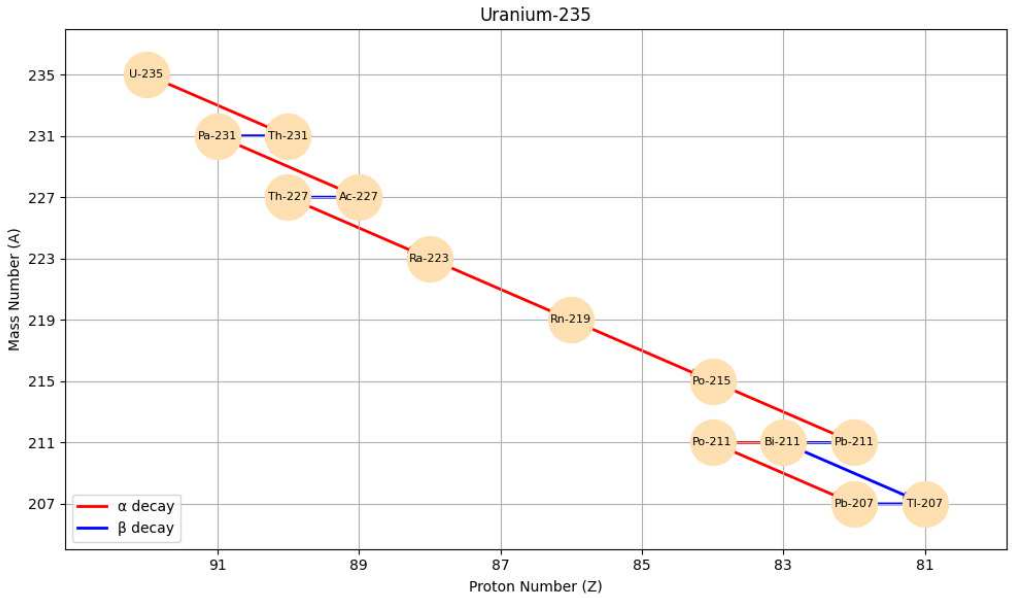
A.3 Uranium-238

Figure 55 – Representation of the ^{238}U decay chain.



A.4 Uranium-235

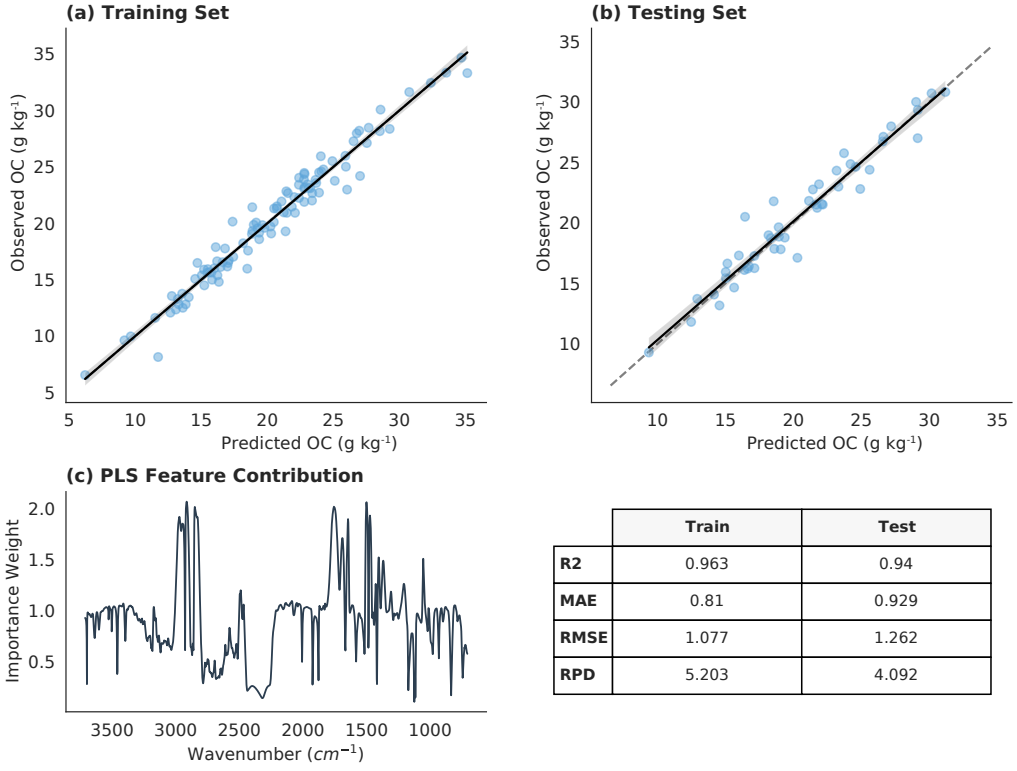
Figure 56 – Representation of the ^{235}U decay chain.



B RESULTS FOR SOIL FERTILITY PARAMETERS WITH STANDALONE MODELS

B.1 MIRS Models

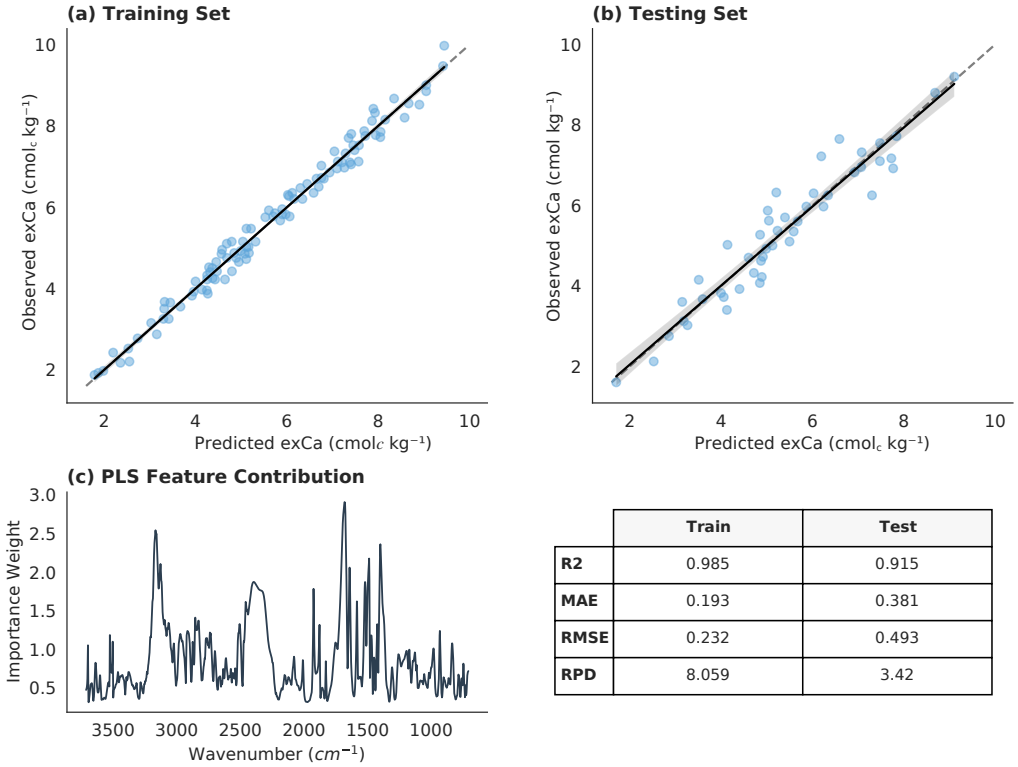
Figure 57 – PLSR results for soil OC prediction based on MIRS spectral data.



	Train	Test
R2	0.963	0.94
MAE	0.81	0.929
RMSE	1.077	1.262
RPD	5.203	4.092

Model Configuration: n_components=4

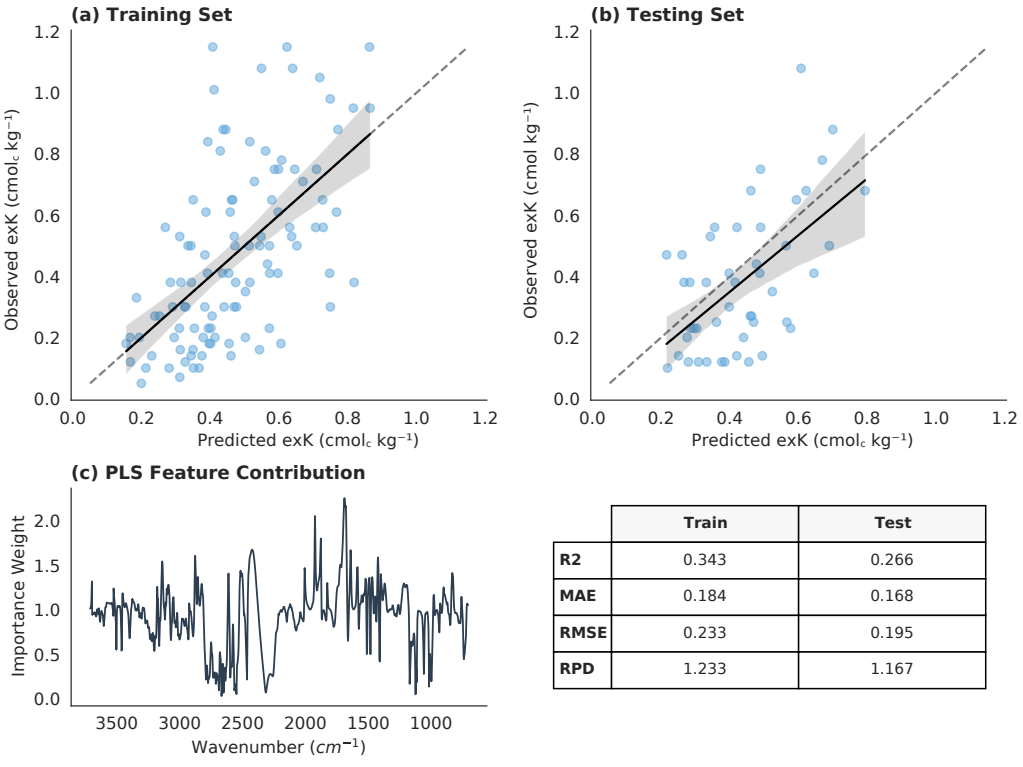
Figure 58 – PLSR results for soil exCa prediction based on MIRS spectral data.



	Train	Test
R2	0.985	0.915
MAE	0.193	0.381
RMSE	0.232	0.493
RPD	8.059	3.42

Model Configuration: n_components=12

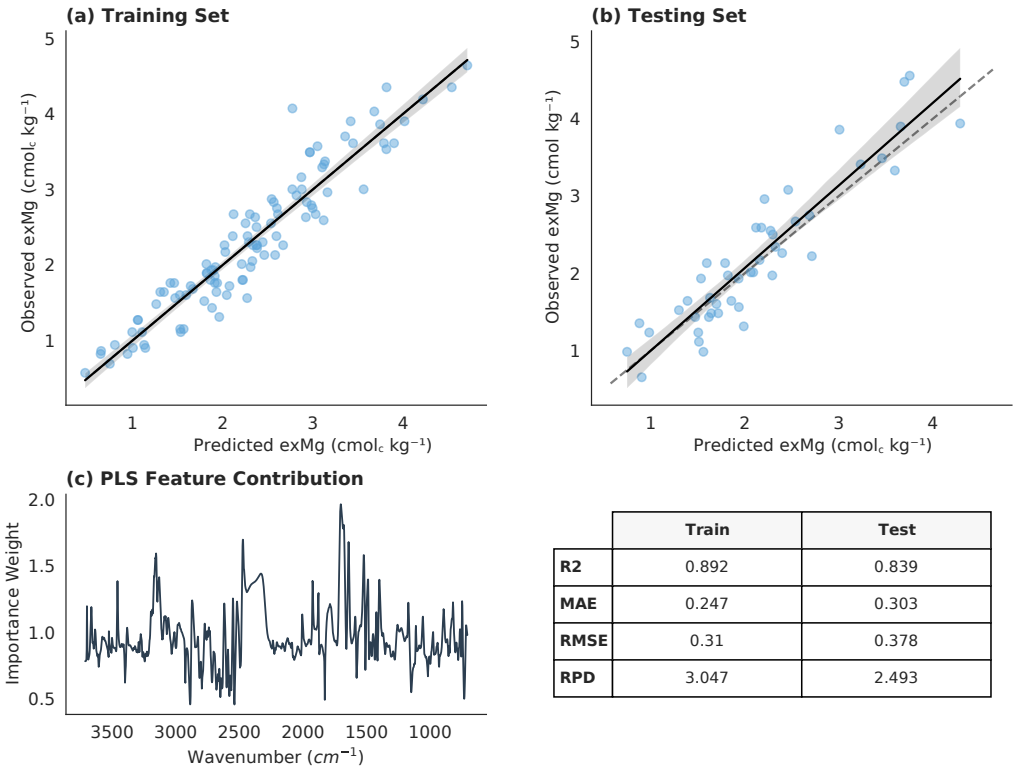
Figure 59 – PLSR results for soil exK prediction based on MIRS spectral data.



	Train	Test
R2	0.343	0.266
MAE	0.184	0.168
RMSE	0.233	0.195
RPD	1.233	1.167

Model Configuration: n_components=2

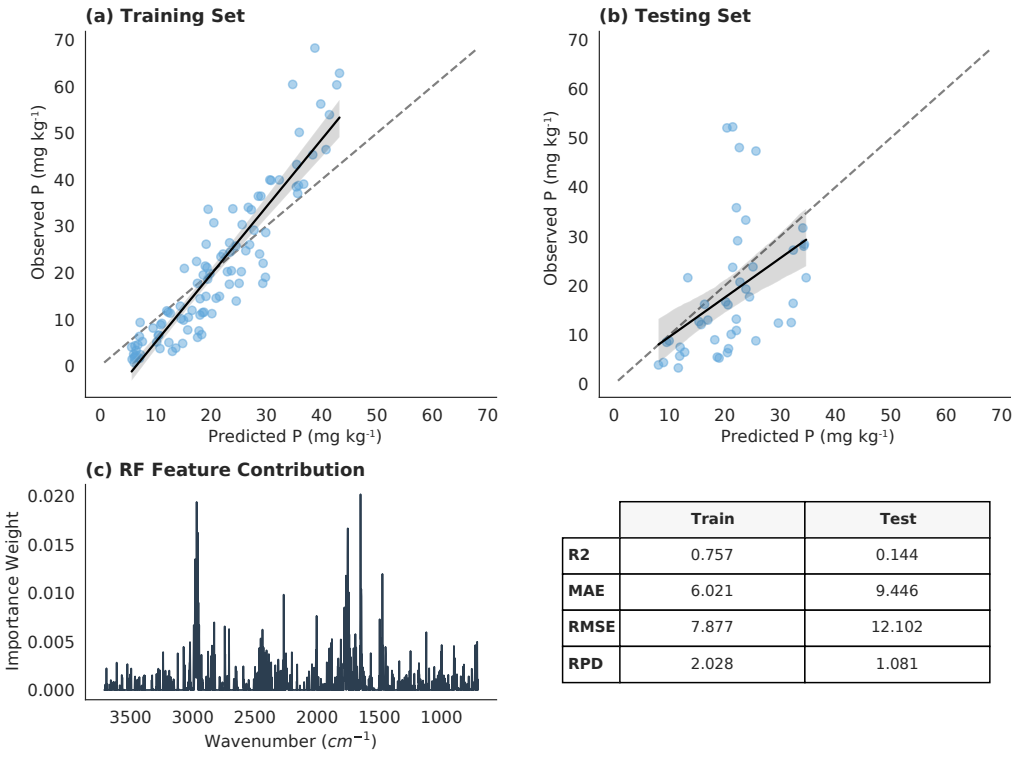
Figure 60 – PLSR results for soil exMg prediction based on MIRS spectral data.



	Train	Test
R2	0.892	0.839
MAE	0.247	0.303
RMSE	0.31	0.378
RPD	3.047	2.493

Model Configuration: n_components=7

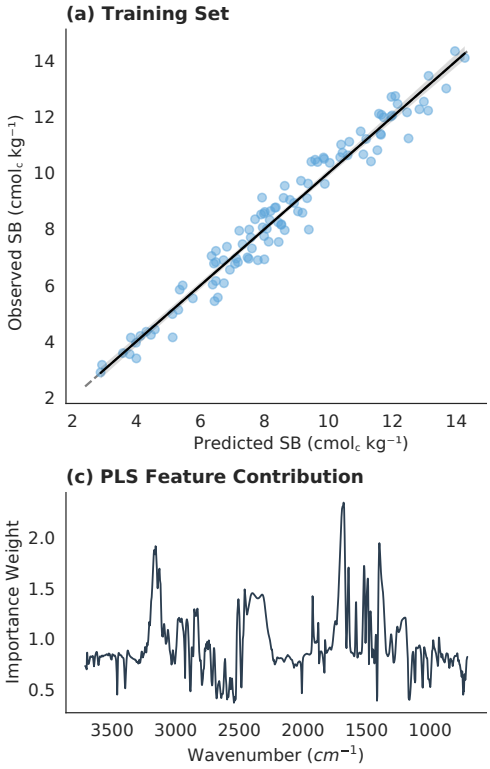
Figure 61 – RFR results for soil exP prediction based on MIRS spectral data.



	Train	Test
R2	0.757	0.144
MAE	6.021	9.446
RMSE	7.877	12.102
RPD	2.028	1.081

Model Configuration: max_depth=25, max_features=sqrt, min_samples_split=15, n_estimators=97

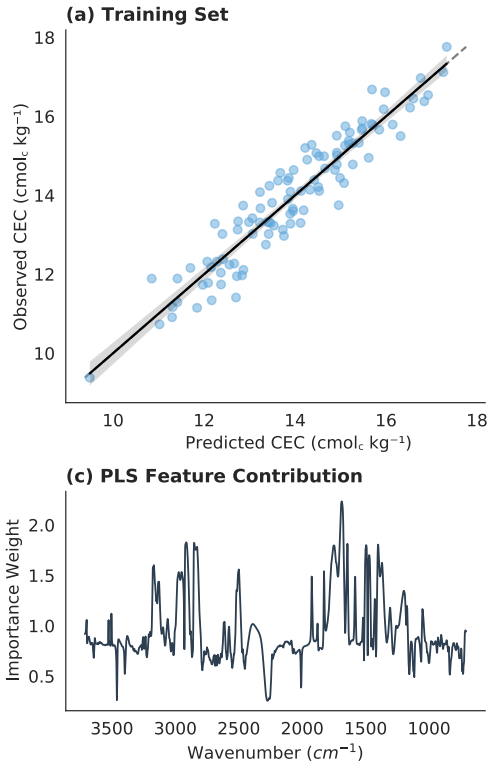
Figure 62 – PLSR results for soil SB prediction based on MIRS spectral data.



	Train	Test
R2	0.96	0.946
MAE	0.454	0.481
RMSE	0.545	0.598
RPD	5.008	4.313

Model Configuration: n_components=7

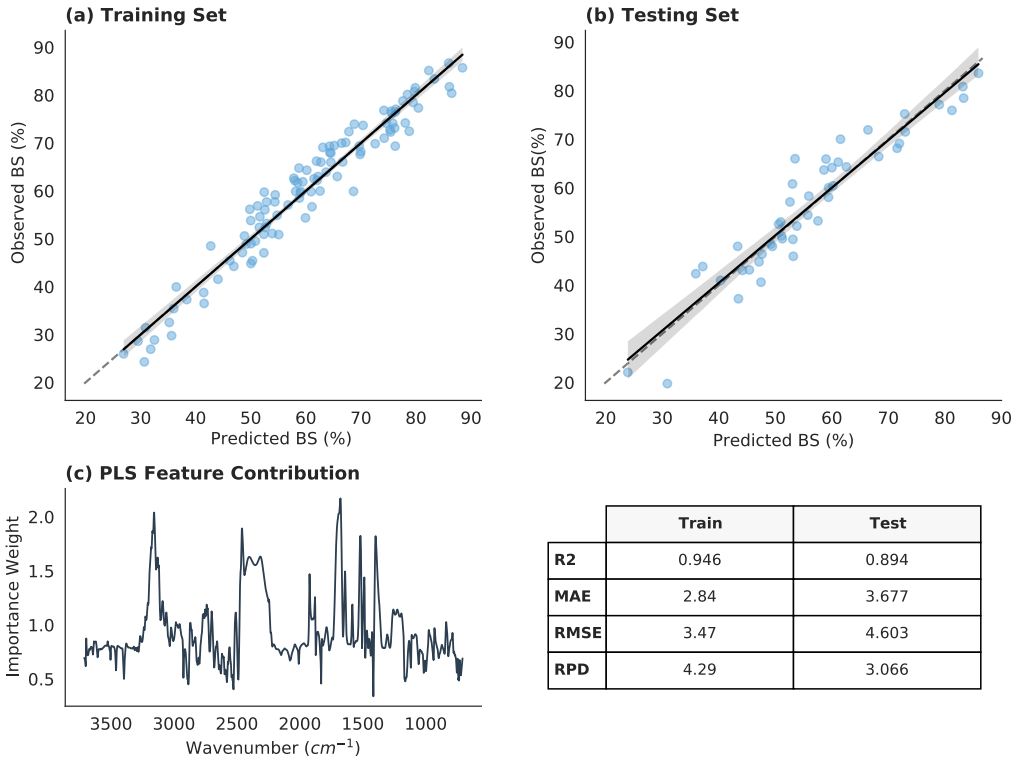
Figure 63 – PLSR results for soil CEC prediction based on MIRS spectral data.



	Train	Test
R2	0.898	0.745
MAE	0.429	0.669
RMSE	0.522	0.794
RPD	3.136	1.982

Model Configuration: n_components=6

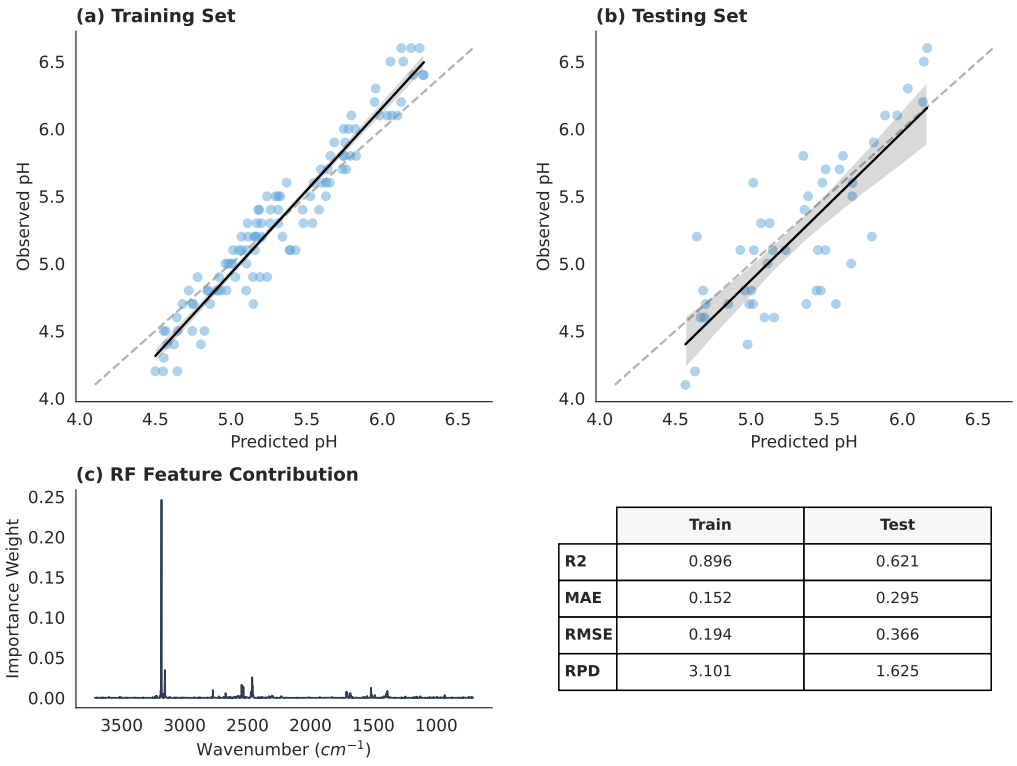
Figure 64 – PLSR results for soil BS prediction based on MIRS spectral data.



	Train	Test
R2	0.946	0.894
MAE	2.84	3.677
RMSE	3.47	4.603
RPD	4.29	3.066

Model Configuration: n_components=7

Figure 65 – RFR results for soil pH prediction based on MIRS spectral data.

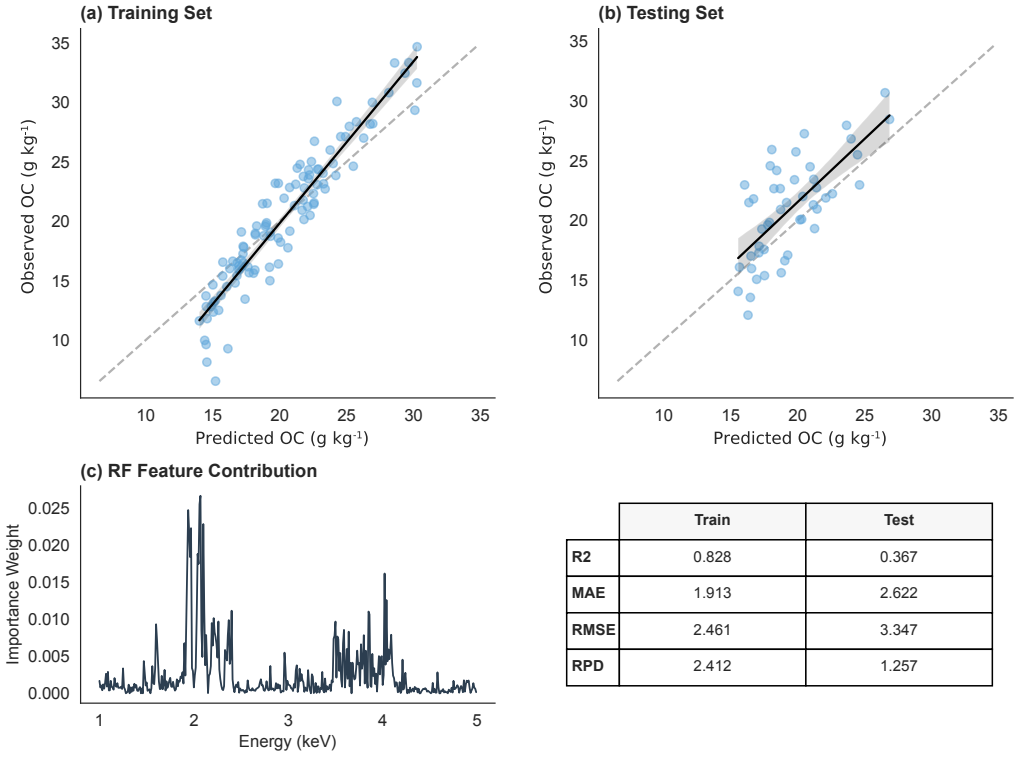


	Train	Test
R2	0.896	0.621
MAE	0.152	0.295
RMSE	0.194	0.366
RPD	3.101	1.625

Model Configuration: max_depth=25, max_features=1.0, min_samples_split=13, n_estimators=244

B.2 XRF Models

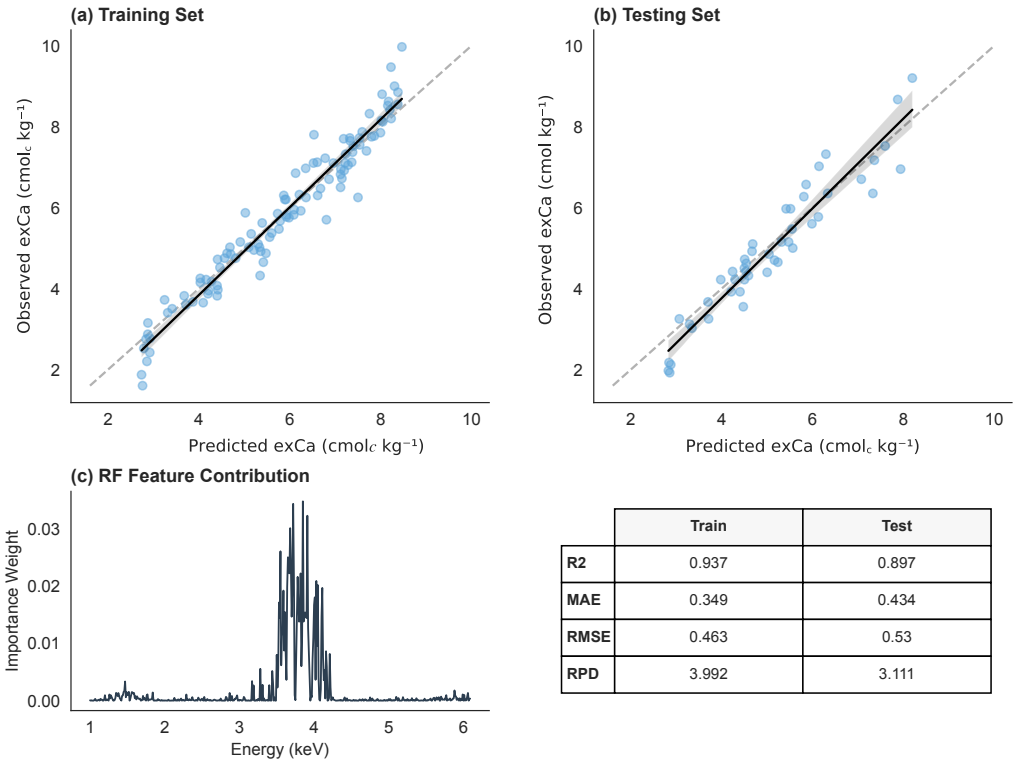
Figure 66 – RFR results for soil OC prediction based on XRF spectral data.



	Train	Test
R2	0.828	0.367
MAE	1.913	2.622
RMSE	2.461	3.347
RPD	2.412	1.257

Model Configuration: {'n_estimators': 200, 'min_samples_split': 10, 'min_samples_leaf': 4, 'max_features': 'sqrt', 'max_depth': None}

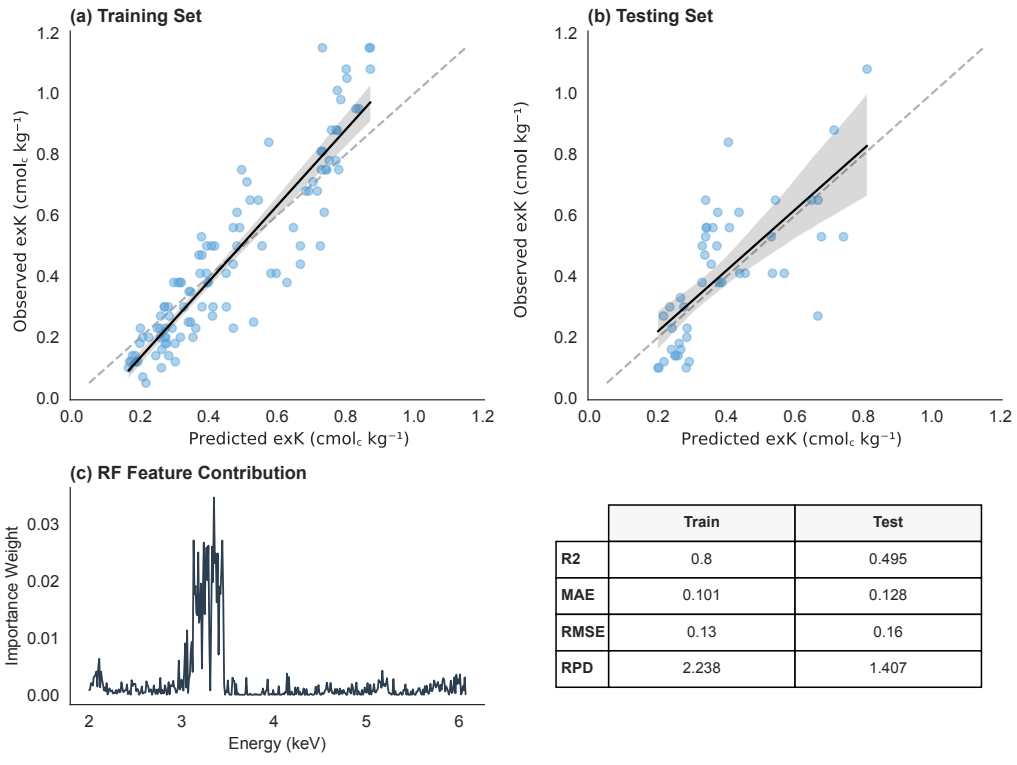
Figure 67 – RFR results for soil exCa prediction based on XRF spectral data.



	Train	Test
R2	0.937	0.897
MAE	0.349	0.434
RMSE	0.463	0.53
RPD	3.992	3.111

Model Configuration: max_depth=10, max_features=sqrt, min_samples_split=13, n_estimators=145

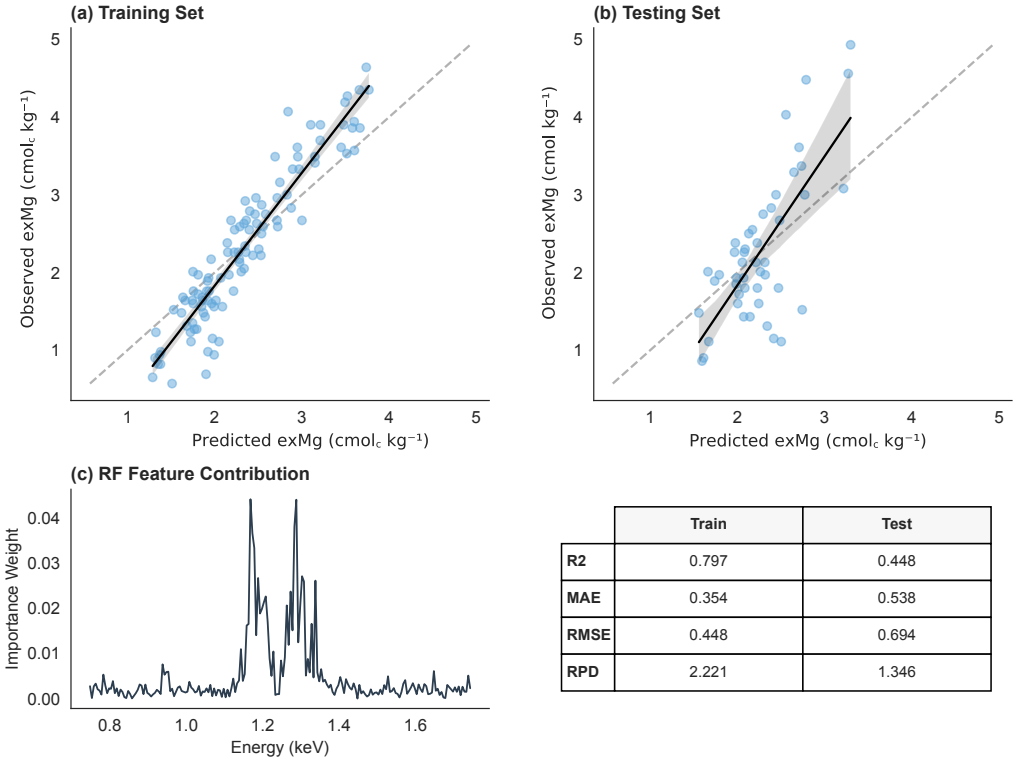
Figure 68 – RFR results for soil exK prediction based on XRF spectral data.



	Train	Test
R2	0.8	0.495
MAE	0.101	0.128
RMSE	0.13	0.16
RPD	2.238	1.407

Model Configuration: max_depth=25, max_features=sqrt, min_samples_split=15, n_estimators=157

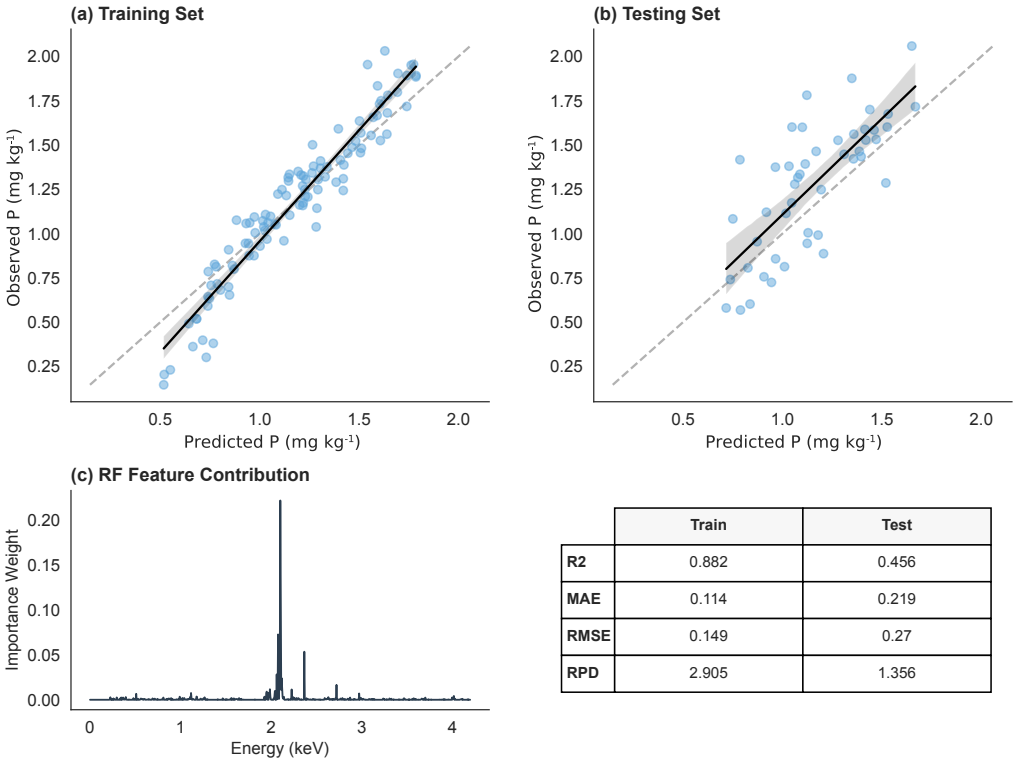
Figure 69 – RFR results for soil exMg prediction based on XRF spectral data.



	Train	Test
R2	0.797	0.448
MAE	0.354	0.538
RMSE	0.448	0.694
RPD	2.221	1.346

Model Configuration: max_depth=None, max_features=sqrt, min_samples_split=13, n_estimators=179

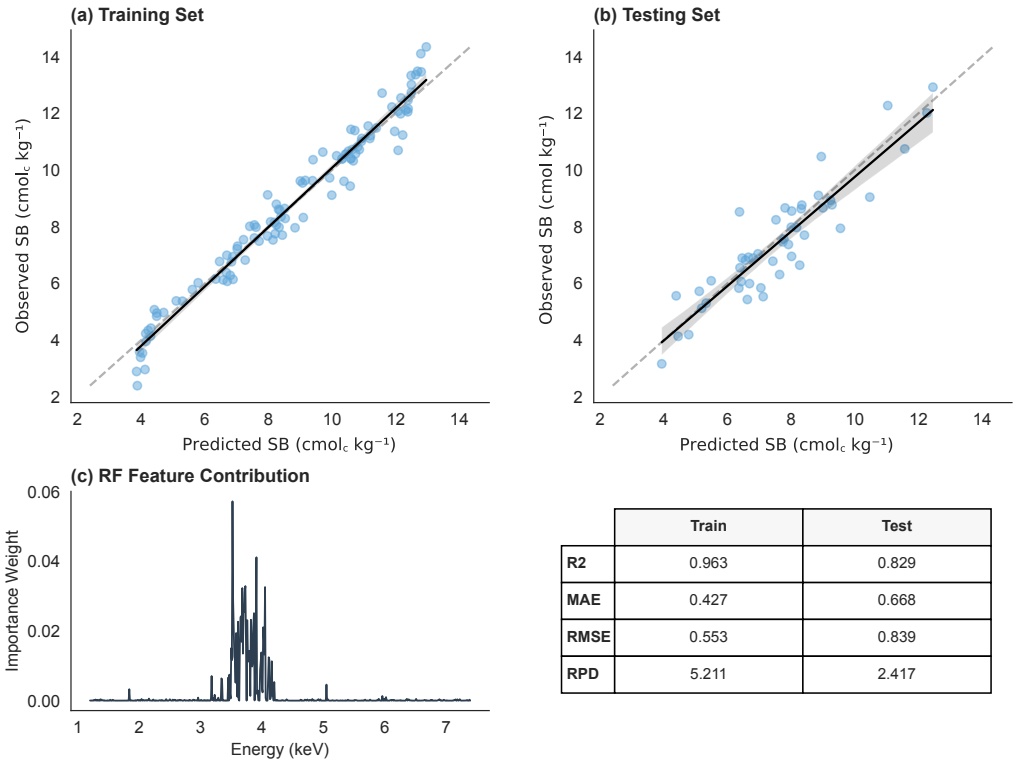
Figure 70 – RFR results for soil exP prediction based on XRF spectral data.



	Train	Test
R2	0.882	0.456
MAE	0.114	0.219
RMSE	0.149	0.27
RPD	2.905	1.356

Model Configuration: max_depth=25, max_features=1.0, min_samples_split=13, n_estimators=74

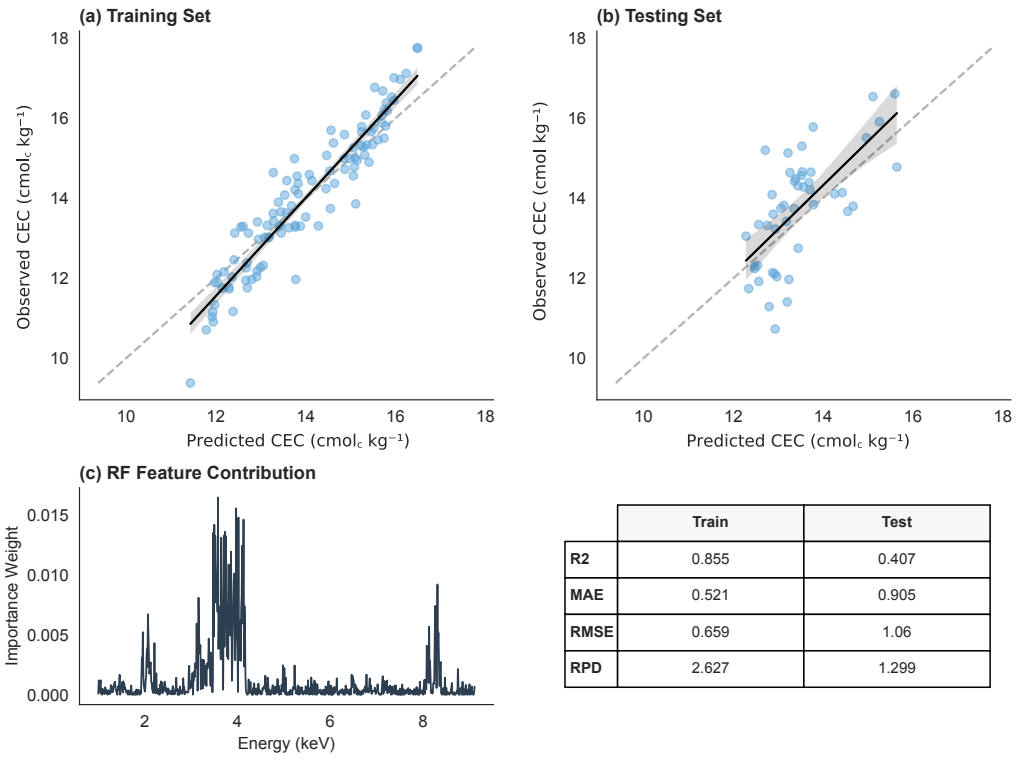
Figure 71 – RFR results for soil SB prediction based on XRF spectral data.



	Train	Test
R2	0.963	0.829
MAE	0.427	0.668
RMSE	0.553	0.839
RPD	5.211	2.417

Model Configuration: max_depth=25, max_features=sqrt, min_samples_split=13, n_estimators=77

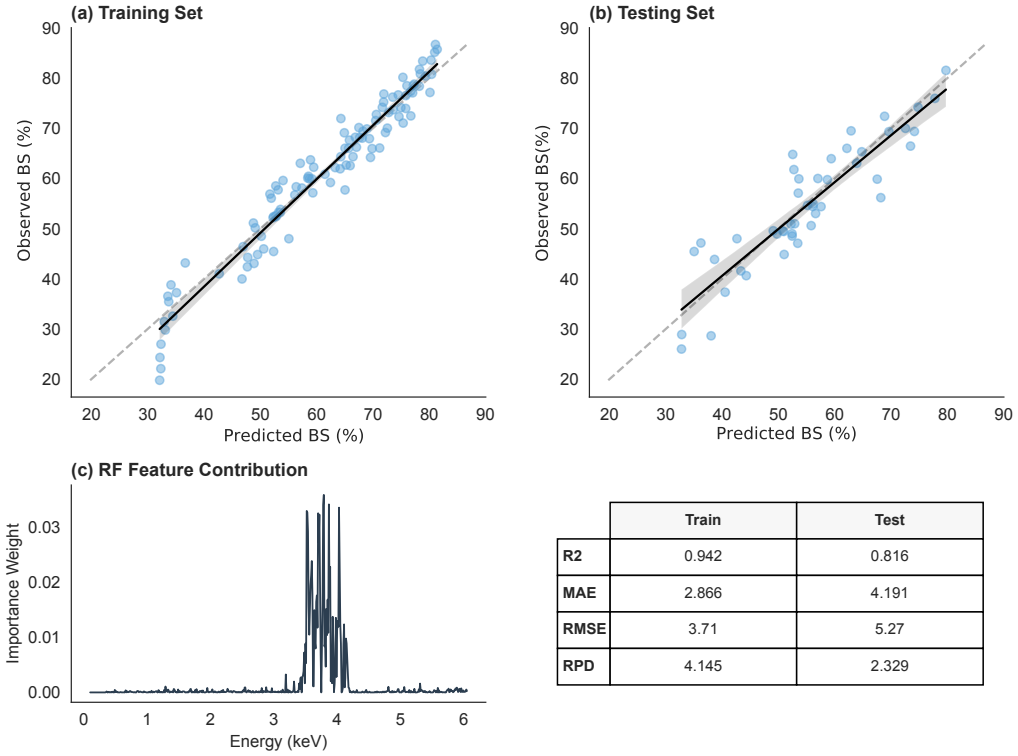
Figure 72 – RFR results for soil CEC prediction based on XRF spectral data.



	Train	Test
R2	0.855	0.407
MAE	0.521	0.905
RMSE	0.659	1.06
RPD	2.627	1.299

Model Configuration: max_depth=10, max_features=log2, min_samples_split=14, n_estimators=287

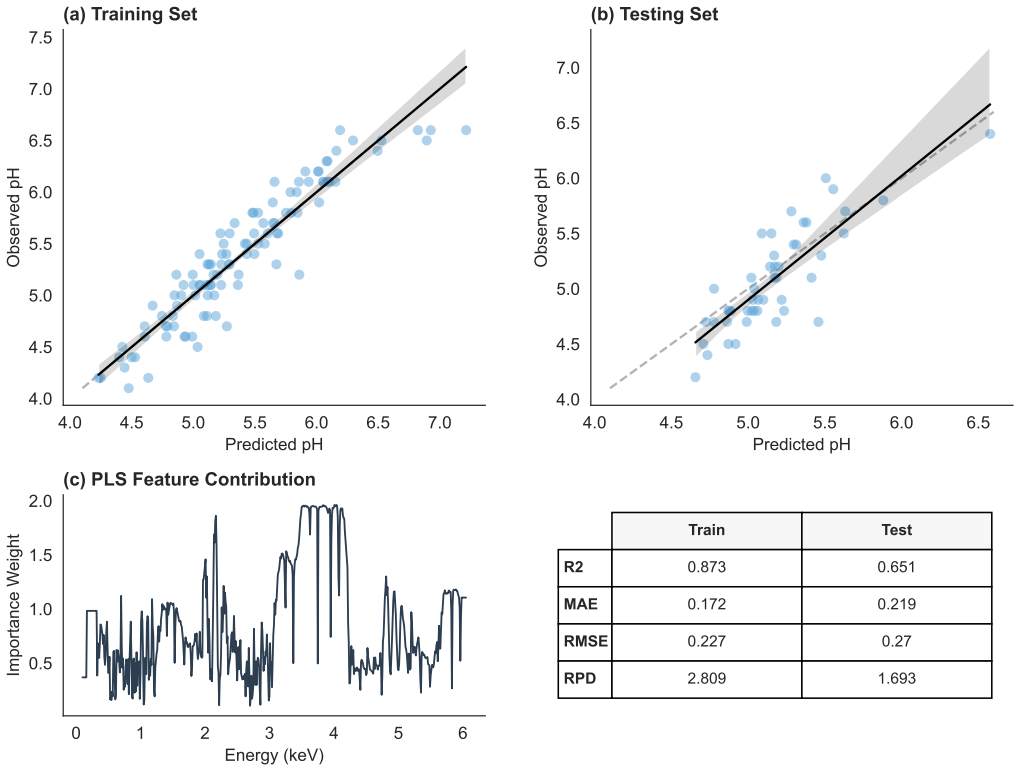
Figure 73 – RFR results for soil BS prediction based on XRF spectral data.



	Train	Test
R2	0.942	0.816
MAE	2.866	4.191
RMSE	3.71	5.27
RPD	4.145	2.329

Model Configuration: max_depth=10, max_features=sqrt, min_samples_split=13, n_estimators=162

Figure 74 – PLSR results for soil pH prediction based on XRF spectral data.



	Train	Test
R2	0.873	0.651
MAE	0.172	0.219
RMSE	0.227	0.27
RPD	2.809	1.693

Model Configuration: n_components=4

PUBLISHED WORK

Peer-reviewed papers and book chapters published during the PhD period are listed below.

Thesis outputs

1. **Lopes, J. M. F.**, Ribeiro, J. V., Andrello, A. C., Melquiades, F. L. (2024). **Gamma-Rays and X-Rays Spectrometries Applied to Evaluate Soil Redistribution**. *Journal of Environmental Radioactivity*, Volume 278, 107479.

2. **Lopes, J. M. F.**, Melquiades, F. L., Bastos, R. O., de Oliveira, J. F., Barbosa, G. M., Andrello, A. C. (2024). **Spectral Deconvolution Approach for Beryllium-7 Detection: Evaluating Soil Redistribution with NaI(Tl) Detectors**. *Radiation Physics and Chemistry*, 225, 112153.

3. Barbosa, G. M. de C., Colozzi Filho, A., de Oliveira, J. F., Andrello, A. C., Menoncin, A. S., Martins, B. H., Bertagnoli, B. G. P., Londero, A. L., Schneider, F. J. A., **Lopes, J. M. F.** (2025). **Qualidade do Solo e a Dinâmica Hidrosedimentológica na Mesorregião Norte**. In *Manejo e Conservação de Solo e Água* (Cap. 4, pp. 201–254). Rede AgroPesquisa IDR-PR.

4. **Lopes, J. M. F.**, Melquiades, F. L., Bastos, R. O., Neto, A. Z., Andrello, A. C. (2026). **Calibration Transfer from HPGe to NaI(Tl) Detectors for Radionuclides Quantification in Soil Samples**. *Journal of Environmental Radioactivity*, Volume 293, 107885.

5. **Lopes, J. M. F.**, Ribeiro, J. V., Melquiades, F. L., de Oliveira, J. F., Barbosa, G. M., Andrello, A. C. (2026). **Data Fusion of Gamma and X-ray Spectra via ComDim for Soil Erosion Characterization**. *Soil Science Society of America Journal*, 90, e70268.

Additional research and collaborations

1. Ribeiro, J. V., **Lopes, J. M. F.**, Andrello, A. C., de Oliveira, J. F., Barbosa, G. M., Bastos, R. O., Melquiades, F. L. (2025). **XRF and Gamma-Ray Data Fusion for Predicting Key Soil Fertility Attributes**. *Radiation Physics and Chemistry*, Volume 234, 112750.

2. Birelo, L. M., França, A. M. de, **Lopes, J. M. F.**, Andrello. A. C., Jussiani, E. I. (2025). **Analysis of the First Rain After a Prolonged Drought in Southern Brazil Using TXRF**. *Radiation Physics and Chemistry*, Volume 237, 113117.

3. Borsoi, B. W., Bastos, R. O., Neto, A. Z., **Lopes, J. M. F.**, Andrello, A. C., Muller, M. M. L., Pott, C. A., Melquiades, F. L. (2026). **^{7}Be Measurements in Agricultural Soil Using Low Resolution Detector**. *Radiation Measurements*, Volume 191, 107591.

4. **Lopes, J. M. F.**, Gahle, D. S., Migliori, A., Kanaki, K. (2026). **Bayesian Calibration of XRF Data for Geological Applications**. *Chemometrics and Intelligent Laboratory Systems*, Volume 272, 105662.

5. Neto, A. Z., Bastos, R. O., Borsoi, B. W., Ikeoka, R. A., Melquiades, F. L., **Lopes, J. M. F.**, Neri, V. S. (2026). **Be-7 Measurements in an Urban Area Using a Low-Resolution Gamma Spectrometer**. *Applied Radiation and Isotopes*, Volume 231, 112529.

6. Desanti, C. S. A., Appoloni, C. R., Ikeoka, R. A., Suñer, M. M. A., Fagundes, M., Silva, F. A., **Lopes, J. M. F.**, Silva, P. R. C. da, Battilani, G. A., Samulewski, R. B. (2026). **Archaeometric study of archaeological pottery from the Ventarrón-Collud Complex – Lambayeque, Peru: A multi-technique approach using EDXRF, FTIR, XRD, Gamma Spectrometry, and Principal Component Analysis**. *Radiation Physics and Chemistry*, Volume 246, 113926.

SCIENTIFIC EVENTS

Scientific events attended and works presented during the PhD:

1. VI Escola de Inverno de Quimiometria. **Análise Exploratória de Solo por Espectrometria Gama para Identificação de Berílio-7**. Brasília – DF, Brazil. Poster. July 2023.

2. VII International Conference on Environmental Radioactivity. **Exploratory Analysis of Gamma Spectrometry and EDXRF Applied to Soil Redistribution Evaluation in Agricultural Areas**. Oral Presentation. Seville, Spain. September 2023.

3. XXIII Escola de Verão Jorge André Swieca de Física Nuclear Experimental. **Mid-Level Data Fusion with Gama and X-Rays Spectra Applied to Soil Redistribution Evaluation: A Novel Approach**. Salvador — BA, Brazil. Poster, awarded. February 2024.

4. IV INCT-FNA Symposium. **Spectral Deconvolution Approach For Beryllium-7 Detection: Evaluating Soil Redistribution With NaI(Tl) Detectors**. Niterói – RJ, Brazil. Oral Presentation. April 2024.

5. IV Brazilian Physical Society Spring Meeting. **Optimizing Radiation Detection: Python-Based Methodologies for CZT Detectors**. Belo Horizonte – MG, Brazil. Poster. September 2024.

6. V INCT-FNA Symposium. **Calibration Transfer from HPGe to NaI(Tl) Detectors for Radionuclides Quantification in Soil**. Niterói – RJ, Brazil. Poster. December 2025.

7. VII Workshop of the Graduate Program in Physics. **Bayesian Calibration of XRF Data for Geological Applications**. Curitiba – PR, Brazil. Oral Presentation, awarded. December 2025.

8. II Workshop of the Graduate Program in Physics. **Análise Espectral Multimodal para Avaliação de Fenômenos Erosivos em Solos Agrícolas**. Londrina – PR, Brazil. Oral Presentation. December 2025.

9. EGU General Assembly 2026. **MIRS and XRF Data Fusion for Improving Soil Fertility Attributes Prediction**. Vienna, Austria. Poster. May 2026.