UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

ANGELO JAMIL MAIA

SOIL DEGRADATION MONITORING STRATEGIES USING INFRARED SPECTROSCOPY

RECIFE
2025

Angelo Jamil Maia
Engenheiro Agrônomo

Soil degradation monitoring strategies using infrared spectroscopy

Tese apresentada ao Programa de Pós-Graduação em Ciência do Solo da Universidade Federal Rural de Pernambuco, como parte dos requisitos para obtenção do título de Doutor em Ciência do Solo

Orientador: Prof. Dr. Yuri Jacques Agra Bezerra da Silva

Recife

2025

Dados Internacionais de Catalogação na Publicação
Sistema Integrado de Bibliotecas da UFRPE
Bibliotecário(a): Suely Manzi – CRB-4 809

M217s Maia, Angelo Jamil.
Soil degradation monitoring strategies using infrared spectroscopy / Angelo Jamil Maia. – Recife, 2025.
147 f.; il.

Orientador(a): Yuri Jacques Agra Bezerra da Silva.

Tese (Doutorado) – Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Ciência do Solo, Recife, BR-PE, 2025.

Inclui referências e apêndice(s).

1. Quimiometria . 2. Sensoriamento remoto . 3. Aprendizado computacional. 4. Degradação do solo 5. Espectroscopia infravermelha. I. Silva, Yuri Jacques Agra Bezerra da, orient. II. Título

CDD 631.4

ANGELO JAMIL MAIA

Soil degradation monitoring strategies using infrared spectroscopy

Tese apresentada ao Programa de Pós-Graduação em Ciência do Solo da Universidade Federal Rural de Pernambuco, como parte dos requisitos para obtenção do título de Doutor em Ciência do Solo.

Aprovada em 29 de Agosto de 2025

_____

Prof. Dr. Yuri Jacques Agra Bezerra da Silva
Orientador
Universidade Federal Rural de Pernambuco

BANCA EXAMINADORA

_____

Prof. Dr. Valdomiro Severino de Souza Júnior
Universidade Federal Rural de Pernambuco

_____

Prof. Dr. Tales Tiecher
Universidade Federal do Rio Grande do Sul

_____

Profa. Dra. Maria Eugenia Ortiz Escobar
Universidade Federal do Ceará

_____

Prof. Dr. Adrian Collins (Membro)
Rothamsted Research

Dedico a meus avôs e avós, que já não caminham conosco, e aos tantos avôs e avós que os precederam, pois a estes devo a vida, e, como resultado de suas histórias, existo.

AGRADECIMENTOS

Ao meu filho Ivan, por todo significado e propósito que me trouxe com sua chegada ao mundo. Te ver crescer é motivação suficiente pra viver mil vidas.

À minha companheira de vida, Michelle, por sua lealdade incondicional e por ser meu porto seguro todos esses anos. Navegar no mar da vida é muito mais fácil sabendo que te tenho ao meu lado.

Aos meus pais, Isis e Germano, pela criação e incentivo ao estudo que sempre me foi dado. Minha gratidão aos dois é silenciosa, mas sempre presente.

À UFRPE, minha casa científica, lugar onde forjei minha visão de mundo e criei amizades que espero perdurarem décadas. Foram mais de 12 anos na Rural, e o orgulho de fazer parte desta instituição é proeminente na minha formação pessoal e profissional.

A meu orientador, Yuri Jacques, pela parceria e pelos valiosos ensinamentos. Foram vários anos de colaboração, muitas histórias pra contar e bastante conhecimento compartilhado. Mais importante do que o destino é a companhia na viagem, e nesse caso em específico eu dei bastante sorte.

Ao Dr. Rennan Nascimento e aos professores Ygor Jacques e Clístenes Nascimento pelas parcerias científicas sem as quais não teria sido possível a realização do presente trabalho e de muitos outros.

Ao programa de pós-graduação em Ciência do Solo da UFRPE, pelo acolhimento e comprometimento com a formação de bons pesquisadores. Estendo essa gratidão a todos que compõem o programa: professores, técnicos, profissionais de limpeza e da secretaria.

À CAPES, pela concessão da bolsa de doutorado, que me permitiu o desenvolvimento pessoal e profissional nesses últimos quatro anos.

Ao CNPq, pelo apoio financeiro na realização das pesquisas que participei, direta e indiretamente.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Meu muito obrigado!

"A gem cannot be polished without friction, nor a man perfected without trials."

— Seneca

Soil degradation monitoring strategies using infrared spectroscopy

ABSTRACT

Soil degradation represents a critical global challenge, threatening food security, biodiversity, and environmental stability through processes including organic matter depletion, erosion, contamination by potentially toxic elements (PTEs), salinization, and desertification. Traditional soil monitoring approaches, while accurate, are costly, labor-intensive, and logistically demanding, raising a need for more efficient assessment methods. Infrared spectroscopy (IRS) has emerged as a promising alternative, offering rapid, non-destructive, and cost-effective soil assessment. This work evaluated innovative methodological approaches for monitoring soil degradation with IRS, with emphasis on predicting soil attributes such as PTEs. The research is comprised of three axes: 1. a comprehensive review synthesizing current scientific knowledge on IRS applications for soil degradation monitoring, with the addition of case studies comprised of diverse datasets; 2. an assessment of near-infrared (NIR), mid-infrared (MIR), and spectral fusion (NIR+MIR) approaches for predicting metal contents in Technosols derived from scheelite mining residues across different accumulation periods (2, 10, and 40 years), and; 3. an evaluation of Convolutional Neural Networks (CNN) for predicting Ba, Cd, Cu, and Pb concentrations using a global mid-infrared soil spectral library. Both Partial Least Squares regression (PLSR) and CNN modeling approaches were systematically compared across diverse datasets. The results demonstrated significant potential for IRS-based soil degradation monitoring across multiple settings. The comprehensive review revealed substantial advances in IRS modeling capabilities through machine learning integration, successful applications across diverse environments supported by spectral libraries, and promising developments in the soil assessment of carbon, pollutants, erodibility, and salinity-related properties. For the Technosols, MIR and fusion spectra outperformed NIR, with the best performances for Cu (R^2 = 0.69, RPD = 1.81) and Sr (R^2 = 0.67, RPD = 1.76). The global spectral library modeling showed a superior performance for CNN over PLSR, achieving excellent predictions for Cd (R^2 = 0.81, RPD = 2.32) and reliable results for Ba, Cu, and Pb (R^2 > 0.70, RPD > 1.8). Elements with strong geochemical associations with potential spectrally active components exhibited better predictive performance across studies. These findings reinforce IRS as an efficient tool for environmental monitoring, supporting sustainable land management and the development of rapid soil contamination assessment systems.

Keywords: Chemometrics. Proximal sensing. Machine learning. Land degradation.

Estratégias para monitoramento da degradação do solo utilizando espectroscopia no infravermelho

RESUMO

A degradação do solo representa um desafio global crítico, ameaçando a segurança alimentar, a biodiversidade e a estabilidade ambiental por meio de processos como a depleção de matéria orgânica, erosão, contaminação por elementos potencialmente tóxicos (PTEs), salinização e desertificação. As abordagens tradicionais de monitoramento do solo, embora precisas, são onerosas, exigem muita mão de obra e são logisticamente complexas, o que ressalta a necessidade de métodos de avaliação mais eficientes. A espectroscopia no infravermelho (IRS) surgiu como uma alternativa promissora, oferecendo uma avaliação do solo rápida, não destrutiva e de baixo custo. Este trabalho avaliou abordagens metodológicas inovadoras para o monitoramento da degradação do solo com IRS, com ênfase na predição de atributos do solo como PTEs e propriedades relacionadas à salinidade. A pesquisa foi composta por três eixos: 1. uma revisão abrangente que sintetiza o conhecimento científico atual sobre aplicações da IRS no monitoramento da degradação do solo, com a adição de estudos de caso compostos por conjuntos de dados diversos; 2. uma avaliação das abordagens no infravermelho próximo (NIR), infravermelho médio (MIR) e de fusão espectral (NIR+MIR) para a predição de teores de metais em Tecnossolos derivados de resíduos de mineração de scheelita, ao longo de diferentes períodos de acúmulo (2, 10 e 40 anos); 3. uma avaliação de Redes Neurais Convolucionais (CNN) para a predição das concentrações de Ba, Cd, Cu e Pb utilizando uma biblioteca espectral global de solos no infravermelho médio. As abordagens de modelagem por regressão de mínimos quadrados parciais (PLSR) e CNN foram comparadas sistematicamente em diferentes conjuntos de dados. Os resultados demonstraram um potencial significativo para o monitoramento da degradação do solo com IRS em múltiplos contextos. A revisão revelou avanços substanciais nas capacidades de modelagem da IRS por meio da integração com aprendizado de máquina, aplicações bem-sucedidas em ambientes diversos apoiadas por bibliotecas espectrais e desenvolvimentos promissores na avaliação do solo em relação a carbono, poluentes, erodibilidade e propriedades relacionadas à salinidade. Nos Tecnossolos, os espectros MIR e de fusão superaram o NIR, com os melhores desempenhos obtidos para Cu (R^2 = 0,69; RPD = 1,81) e Sr (R^2 = 0,67; RPD = 1,76). A modelagem com a biblioteca espectral global mostrou desempenho superior das CNNs em relação ao PLSR, alcançando predições excelentes para Cd (R^2 = 0,81; RPD = 2,32) e resultados confiáveis para Ba, Cu e Pb (R^2 > 0,70; RPD > 1,8). Elementos com fortes associações geoquímicas com componentes potencialmente ativos espectralmente apresentaram melhor desempenho preditivo nos estudos. Esses resultados reforçam a IRS como uma ferramenta eficiente para o monitoramento ambiental, apoiando a gestão sustentável da terra e o desenvolvimento de sistemas rápidos de avaliação da contaminação do solo.

Palavras-chave: Quimiometria. Sensoriamento proximal. Aprendizado de máquina. Degradação da terra.

LIST OF FIGURES

LIST OF TABLES

SUMMARY

1 GENERAL INTRODUCTION

Soil is a vital, non-renewable resource essential for the survival of terrestrial ecosystems. It supports agriculture, sustains biodiversity, regulates hydrological cycles, and contributes significantly to climate regulation. Yet, despite its importance, soils worldwide face escalating threats from degradation processes. These include the loss of organic matter, erosion, contamination by hazardous substances, salinization, and desertification. Often interlinked, these processes undermine soil functionality, impose difficulties on food production, reduce ecological stability, and carry substantial socioeconomic consequences.

Addressing soil degradation is fundamental to sustainability goals. However, effective mitigation depends on the ability to monitor soil conditions with accuracy, efficiency, and spatial coverage. Traditional monitoring approaches, which are based on extensive sampling and traditional laboratory analyses, are accurate but costly, labor-intensive, and logistically demanding. These limitations are especially problematic in resource-constrained settings or when timely data is needed for decision-making. As pressures from land-use change, population growth, and climate change increase, more efficient monitoring solutions become essential.

Infrared spectroscopy (IRS) has emerged as a promising alternative for soil assessment. It enables rapid, non-destructive, and cost-effective analysis of soil properties by measuring the interaction of infrared radiation with matter. This spectral information is able to reflect a range of characteristics, including organic carbon content, nutrient content, texture, mineralogy, salinity, and contaminant content. When combined with appropriate data modeling, IRS allows for efficient assessments suitable for environmental and agricultural monitoring.

Recent advances in instrumentation, computing, and chemometrics have propelled IRS from an experimental technique to a versatile tool capable of field deployment and remote sensing integration. The construction of soil spectral libraries and the adoption of machine learning models have enhanced both the accuracy and the applicability of IRS across diverse environments. Portable devices now extend IRS reach, while satellite-based spectral data open possibilities for large-scale diagnostics.

The multidimensional nature of soil degradation makes IRS particularly relevant. Depletion of organic carbon, for instance, weakens soil structure and reduces fertility. Erosion strips away topsoil and associated nutrients. Pollution, both organic and inorganic,

compromises biological activity, ecosystem safety, and may directly affect human health. Salinity alters soil chemistry, reducing agricultural viability. Desertification, driven by these and other factors, marks the ultimate collapse of once productive lands. Each of these degradation processes are in need of proper monitoring strategies, which can be directly or indirectly assisted by spectral data obtained with IRS.

Maximizing IRS effectiveness requires robust predictive models that correlate spectral data with physical or chemical soil attributes. This calibration process relies on paired spectral and laboratory reference datasets. Once validated, these models can enable fast, repeatable analysis of new samples. There are many different modeling methods, such as partial least squares regression, support vector machines, random forests and neural networks, among others. Novel modeling approaches allow for the enhancement of predictive performance, and drive further innovation in the applicability of this technique.

One of the core advantages of IRS is that it enables integrated soil assessments. Unlike conventional methods focused on single properties, IRS supports simultaneous prediction of multiple properties from a single measurement. This is particularly useful for assessing degradation as a composite process, where changes in one variable (e.g., carbon content) often reflect or influence others (e.g., erodibility). When linked to spatial modeling, IRS may help identify degradation hotspots and support targeted interventions.

While promising, IRS is not without challenges. Variability in soil composition demands region-specific calibrations. Low pollutant concentrations fall below detection thresholds as they lack visible spectral features. Spectral noise and signal overlap can complicate interpretation. However, ongoing research continues to improve spectral preprocessing, feature selection, and model standardization, addressing these limitations and reinforcing the reliability of this technique.

This thesis investigates the application of infrared spectroscopy for assessing soil degradation, emphasizing the estimation of potentially toxic elements contents, soil erodibility, and salinity-related properties. The study also explores the benefits of methodological innovations, including the use of convolutional neural networks and spectral fusion, to enhance IRS performance and broaden its applicability. The overall aim is to demonstrate how IRS can provide efficient and cost-effective soil assessments that support sustainable land management. Positioned within the broader discourse on soil monitoring, this thesis aims to contribute to a refined understanding of IRS as a tool capable of assisting in the capture of the complex, multi-dimensional nature of soil degradation processes.

1.1 Hypotheses

Emerging evidences indicate that infrared spectroscopy, when coupled with well-calibrated statistical and machine learning models, is capable of accurately predicting key soil degradation attributes, including carbon content, salinity indicators, and concentrations of potentially toxic elements.

The integration of Convolutional Neural Networks (CNN) with spectral data fusion techniques, combining near- and mid-infrared spectra, enhances the predictive performance for soil erodibility, salinity-related properties, and potentially toxic elements, surpassing conventional chemometric approaches such as Partial Least Squares Regression (PLSR), across diverse environmental contexts.

1.2 General objective

To evaluate innovative technologies and methodological approaches for the monitoring of soil degradation using infrared spectroscopy, with emphasis on the prediction of soil quality attributes such as potentially toxic elements contents, soil erodibility, and salinity-related properties.

1.3 Specific objectives

To review and synthesize current scientific knowledge on the potential of infrared spectroscopy as an alternative assessment tool for soil degradation processes.

To assess the ability of infrared spectroscopy to estimate soil properties associated with degradation (soil erodibility, salinity-related properties, and potentially toxic elements) across diverse environmental conditions.

To evaluate the performance of different modeling approaches, including Partial Least Squares regression (PLSR) and Convolutional Neural Networks (CNN), for the prediction of soil properties from infrared spectral data.

To analyze the feasibility of a global soil spectral library for the development of robust prediction models for potentially toxic elements contents.

2 LITERATURE REVIEW

2.1 Soil degradation

The decline in soil quality – driven by natural and anthropogenic factors such as erosion, contamination, salinization, and loss of organic matter – is known as soil degradation (GOMIERO, 2016). As soils deteriorate, their capacity to support vegetation, cycle nutrients, store carbon, and regulate water is impaired, threatening food production and ecosystem stability (KARLE; RICE, 2015). This is a global concern, affecting a wide range of environments and requiring a clear understanding of its causes and impacts. Erosion is both a major form of degradation and a consequence of the combination of diverse degradation processes, removing topsoil and disturbing soil structure. Its effects go beyond yield loss, altering carbon dynamics and increasing water pollution (LAL, 2001).

The loss of soil organic matter (SOM) is another central issue. SOM supports the maintenance of soil structure, fertility, and microbial activity, making it a key indicator of soil health. Its decline increases vulnerability to other forms of degradation like erosion and compaction. Because it influences so many properties, tracking SOM, especially its labile fractions, is essential for monitoring and early intervention (OBALUM et al., 2017). Chemical degradation, though less visible, is equally damaging. Contamination from fertilizers and pesticides degrade soil function and fertility (TETTEH, 2015). Soil pollution can occur through many different inputs and processes, and both widespread and localized sources can cause long-term harm, making prevention and monitoring critical issues.

Human-driven land use changes have accelerated soil degradation. Studies show that converting forests and pastures into cropland or urban areas lowers soil quality, particularly in indicators like cation exchange capacity and labile carbon (KHALEDIAN et al., 2017). In urban areas, soil quality is often the lowest, reflecting the cumulative effects of construction and compaction. Soil degradation also carries substantial economic costs. In England and Wales, the annual damage is estimated at £0.9–1.4 billion, mostly from organic matter loss, compaction, and erosion (GRAVES et al., 2015). Some of these negative economic outcomes can occur off-site, such as through increased flooding and water contamination. Soil degradation also undermines food production, increasing food insecurity risks (BINDRABAN et al., 2012).

Furthermore, desertification represents the extreme spectrum of soil degradation, especially in arid and semi-arid regions. This process results from a combination of climate pressures – such as drought and high temperatures – and unsustainable human practices. Remote sensing analyses show that this process can result from significant increases in soil salinity and moisture stress over time, particularly in regions with poor land management (SHOBA; RAMAKRISHNAN, 2016). Desertification can be considered the ultimate form of soil degradation, transforming productive lands into fragile systems, further escalating food insecurity and ecosystem decline. Restoring soil quality is vital for long-term sustainability. Increasing soil organic carbon beyond key thresholds is especially effective in reversing degradation trends (Lal, 2015). However, these strategies must be paired with broader shifts in land-use policy and environmental monitoring (GOMIERO, 2016; KARLEN; RICE, 2015).

2.2 Soil pollution

The presence of toxic contaminants in soil at levels that pose risks to human health, ecosystems, or agricultural productivity is known as soil pollution. It arises from both natural and anthropogenic sources, including industrial activity, agricultural practices, and improper waste disposal (GAUTAM et al., 2023). Among the most concerning soil contaminants are heavy metals, which can be introduced through fertilizers, sewage, industrial emissions, and pesticides. These metals, such as lead (Pb), cadmium (Cd), and mercury (Hg), can persist in soils for long periods, accumulate in food crops, and pose significant risks to human health via the food chain (ZWOLAK et al., 2019). Heavy metals are considered potentially toxic elements (PTEs) due to their toxicity, persistence, long residence times in soils, and tendency to bioaccumulate through trophic chains and direct contact (ALI et al., 2019; WANG et al., 2020). Once introduced into the food chain, they can have acute and chronic toxic effects on humans, as they are not metabolized and instead accumulate in tissues, potentially leading to various diseases, including carcinogenic and mutagenic outcomes (JAFARZADEH et al., 2020).

In addition to inorganic pollutants, soils are also affected by organic contaminants like polycyclic aromatic hydrocarbons (PAHs), which originate from incomplete combustion and industrial discharges. These compounds can alter soil structure, reduce microbial diversity, and impair overall soil function (SAKSHI; SINGH; HARITASH, 2019). Pesticides represent another critical category of soil pollutants. Their diverse chemical nature and transformation

pathways challenge our ability to assess and mitigate their impacts on soil biological activity (WOŁEJKO et al., 2020). In both developed and developing nations, soil pollution is considered a critical environmental concern driven by industrial activities, urban development, and the intensification of agriculture.

A wide array of pollutants, both organic and inorganic, enters soils from sources such as industrial discharges, vehicular emissions, and the overuse of agrochemicals (SKALA et al., 2018). These contaminants pose significant threats not only to soil health and biodiversity but also to human health. For example, Cd contamination in agricultural soils can result in its accumulation in staple crops like maize, raising food chain exposure risks, particularly in regions with differing soil properties that affect metal bioavailability and plant uptake (RETAMEL-SALGADO et al., 2017).

More recently, rare earth elements (REEs) – a group of 17 elements including the lanthanide series, scandium (Sc), and yttrium (Y) – have also been recognized as emerging soil pollutants. Like heavy metals, REEs are considered PTEs due to their potential to impact human and environmental health (XIONG et al., 2019). Their increasing input into soils via fertilizers, mining waste, and industrial discharges calls for consistent environmental monitoring (PEREIRA et al., 2019). REEs tend to associate with sediments and are useful not only for tracking pollution sources but also for understanding biogeochemical and weathering processes (TRIFUOGGI et al., 2018). This highlights the need to expand pollution assessment beyond traditional contaminants to include REEs, especially in areas subject to agricultural intensification or mining.

The spatial variability and distribution of soil contaminants are strongly influenced by factors such as soil type, land use, proximity to pollution sources, and urban infrastructure. In roadside environments, for instance, elevated levels of heavy metals like Pb, zinc (Zn), and copper (Cu) are often detected due to traffic emissions and surface runoff, with concentrations increasing alongside road age and traffic density (WANG; ZHANG, 2018). Beyond surface patterns, studies have demonstrated that pollutant behavior also varies with soil depth, as illustrated by the distinct vertical and horizontal distribution of chromium in historically contaminated industrial soils (ZHOU et al., 2018). Geostatistical and modeling approaches have been instrumental in mapping these patterns and guiding environmental risk assessments (HU et al., 2018). The complexity of soil contamination raises the need for integrated approaches to monitor, understand, and enable the remediation of affected environments.

2.3 Soil salinity

Soil salinity is a significant form of land degradation that threatens agricultural productivity and environmental sustainability on a global scale. It results from the accumulation of soluble salts in the root zone to levels that adversely affect plant growth and soil health (HARDIE; DOYLE, 2012). The phenomenon has both historical and contemporary relevance, having contributed to the decline of ancient civilizations such as Mesopotamia due to irrigation-induced salinization. Today, salinity affects over one billion hectares across more than 100 countries, with between 25% and 30% of irrigated lands suffering from salinity-related issues, rendering them unsuited for agriculture or commercially unproductive (SHAHID; ZAMAN; HENG, 2018a).

This widespread issue poses not only environmental challenges but also economic ones, with annual losses estimated in the tens of billions of dollars. Saline soils contain excess salts that limit plant growth, while sodic soils have elevated levels of sodium, harming soil structure and water absorption. When both issues occur together, soils are labeled saline-sodic and are especially harmful. These categories are widely used by scientists and land managers to evaluate soil health and determine how to manage or restore affected areas (SHAHID; ZAMAN; HENG, 2018b). Although the basic concepts are globally recognized, the ways salinity is measured can differ between countries and regions (HARDIE; DOYLE, 2012).

Climate change and unsustainable irrigation further exacerbate the spread of soil salinization, with projections indicating that up to 50% of arable land could be affected by 2050 (BUTCHER et al., 2016). The effects are already alarming in Europe, where salinization pressures linked to both natural and anthropogenic drivers impair critical soil functions (DALIAKOPOULOS et al., 2016). Modern diagnostic techniques, including remote sensing, electromagnetic sensors, and geostatistical modeling, have enhanced the capacity to detect and map salinity at multiple scales, yet regional calibration and methodological harmonization remain necessary (SHAHID; ZAMAN; HENG, 2018a). Furthermore, salinity threatens global food security by limiting crop yields, particularly for sensitive species such as corn and soybean, and compromises progress toward several Sustainable Development Goals (SINGH, 2022). Addressing this complex challenge requires integrated strategies encompassing monitoring, mitigation, and adaptation under diverse environmental conditions.

2.4 Soil infrared spectroscopy

Infrared Spectroscopy (IRS) has gained increasing attention as a valuable tool for the rapid, non-destructive analysis of soil properties. Compared to traditional laboratory methods, IRS offers significant advantages in terms of cost, time, and environmental impact. One of its key strengths lies in its ability to estimate both qualitative and quantitative information without the need for complex sample preparation and chemical reagents, making it a safer and more sustainable alternative (CAPUANO; VAN RUTH, 2016). The technique operates by irradiating soil samples with electromagnetic radiation in the infrared range, measuring the energy absorbed or reflected, which is linked to molecular vibrations (PASQUINI, 2018).

According to the Beer-Lambert law, these absorbance readings are directly related to the concentration of compounds within the sample. Among its various applications, near-infrared (NIR) spectroscopy (750–2500 nm) has proven particularly useful for monitoring contaminants such as heavy metals in soils, sediments, and water (COZZOLINO, 2016). Its versatility across spectral regions – including visible (Vis), NIR, and mid-infrared (MIR) – has made IRS a widely used method for characterizing soil organic matter (GHOLIZADEH et al., 2013). Studies have also demonstrated its capacity to estimate soil texture with high accuracy when proper data transformation and preprocessing techniques are applied (JACONI; VOS; DON, 2019). Soil mineralogy, especially the identification of clay minerals, likewise benefited from IRS applications. Traditional approaches such as X-ray diffraction (XRD) are accurate but impractical for large-scale studies due to their cost and complexity. In contrast, visible and near-infrared reflectance spectroscopy (Vis-NIR) has emerged as a faster, non-destructive alternative for analyzing mineralogical patterns (FANG et al., 2018). The development of robust calibration models using multivariate statistical methods has enabled IRS to predict a wide range of soil attributes based on spectral data and standard reference measurements (VISCARRA ROSSEL; BEHRENS, 2010).

Recent advances have also underscored IRS value in evaluating overall soil quality, particularly for agricultural decision-making. Instead of relying on multiple discrete laboratory tests, IRS allows for the simultaneous estimation of several physical, chemical, and mineralogical properties with a single spectral reading (DEMATTÊ et al., 2019a). This integrative approach can support a more efficient land management and even real-time monitoring of soil health (ASKARI; O'ROURKE; HOLDEN, 2015). IRS application to the speciation and quantification of adsorbed chemical compounds, such as in the analysis of surface complexes or contaminants, has also been validated in environmental studies

(VILMIN et al., 2015). The integration of machine learning techniques has further enhanced the predictive performance of IRS, particularly for complex datasets involving multiple contaminants or interacting soil variables (PYO et al., 2020). While limitations still exist, IRS is recognized as a reliable, scalable, and efficient method for routine soil monitoring and environmental assessment (SØRENSEN; DALSGAARD, 2005; ZIMMERMAN; LEIFELD; FUHRER, 2007).

One of the key challenges in applying IRS to soil analysis lies in the complex and heterogeneous nature of soil itself. As a physical matrix composed of minerals, organic matter, water, and gases, soil lacks well-defined spectral features for many of its chemical components. Moreover, the reflectance spectrum is significantly affected by light scattering effects, which can obscure the relatively small absorption features related to the actual concentration of specific analytes (HORTA et al., 2015). As a result, the useful information within the spectrum – particularly for chemical quantification – is often limited and blurred within a noisy background signal. To overcome this, dimensionality reduction techniques and variable selection algorithms are employed. These approaches help identify the most informative wavelengths and improve the performance of calibration models used in chemometric analysis.

The construction of robust calibration models is fundamental for predicting soil properties using IRS. A variety of linear and non-linear modeling techniques – such as partial least squares regression (PLSR), support vector machines, regression trees, random forests, and artificial neural networks – are used to correlate spectral reflectance data with laboratory reference measurements (SARATHJITH et al., 2016). The effectiveness of these models often depends on appropriate spectral preprocessing, which enhances signal quality by removing irrelevant noise. Common preprocessing methods include multiplicative scatter correction (MSC), standard normal variate (SNV), and Savitzky-Golay derivatives (GHOLIZADEH et al., 2015). Together, these strategies play a crucial role in maximizing the predictive power of IRS and ensuring its reliability across diverse soil conditions.

2.5 Soil spectral libraries

One of the primary challenges for the advancement of soil infrared spectroscopy is the limited availability of high-quality spectral and analytical data. This scarcity restricts model development, testing, and the exploration of novel approaches. In response, multiple

initiatives have emerged to create comprehensive soil spectral libraries. These repositories serve as essential resources that promote further research, particularly in regions where conventional laboratory analyses are costly or inaccessible. Notably, the Global Soil Spectral Calibration Library and Estimation Service provides crucial data access for developing nations (SHEPHERD et al., 2022). Similarly, the Brazilian Soil Spectral Library was developed for the same purpose and was recently expanded with thousands of MIR samples, offering a robust dataset for national and international studies (MENDES et al., 2022). These efforts highlight the development of open-access spectral libraries as keys to fostering inclusive, data-driven progress in soil infrared spectroscopy.

The effectiveness of soil spectral libraries has been demonstrated in various contexts. In New Zealand, both Vis-NIR and MIR libraries were successfully developed and validated, with MIR consistently showing superior performance for predicting diverse soil properties (MA et al., 2023). Recent studies further exemplify the application of open datasets, such as the use of urban soil libraries for estimating organic carbon through data mining techniques (HONG et al., 2022), and the GLOBAL-LOCAL approach that leverages global Vis-NIR libraries to build precise local models (LUCE; ZIADI; ROSSEL, 2022). These examples illustrate how spectral libraries are not only instrumental for methodological innovation but also essential for promoting standardization and reproducibility in soil infrared spectroscopy.

References

ALI, L.; RASHID, A.; KHATTAK, S. A.; ZEB, M.; JEHAN, S. Geochemical control of potential toxic elements (PTEs), associated risk exposure and source apportionment of agricultural soil in Southern Chitral, Pakistan. **Microchemical Journal**, v. 147, p. 516-523. 2019. https://doi.org/10.1016/j.microc.2019.03.034

ASKARI, M. S.; O'ROURKE, S. M.; HOLDEN, N. M. Evaluation of soil quality for agricultural production using visible–near-infrared spectroscopy. **Geoderma**, v. 243, p. 80–91, 2015. https://doi.org/10.1016/j.geoderma.2014.12.012

BINDRABAN, P. S.; VAN DER VELDE, M.; YE, L.; VAN DEN BERG, M.; MATERECHERA, S.; KIBA, D. I.; TAMENE, L.; RAGNARSDÓTTIR, K. V.; JONGSCHAAP, R.; HOOGMOED, W.; VAN BEEK, C.; VAN LYNDEN, G. Assessing the impact of soil degradation on food production. **Current Opinion in Environmental Sustainability**, v. 4, n. 5, p. 478–488, 2012. https://doi.org/10.1016/j.cosust.2012.09.015

BUTCHER, K.; WICK, A. F.; DESUTTER, T.; CHATTERJEE, A.; HARMON, J. Soil salinity: A threat to global food security. **Agronomy Journal**, v. 108, n. 6, p. 2189–2200, 2016. https://doi.org/10.2134/agronj2016.06.0368

CAPUANO, E.; VAN RUTH, S. M. Infrared Spectroscopy: Applications. **Encyclopedia of Food and Health**, p. 424-431, 2016. https://doi.org/10.1016/B978-0-12-384947-2.00644-9

COZZOLINO, D. Near infrared spectroscopy as a tool to monitor contaminants in soil, sediments and water—State of the art, advantages and pitfalls. **Trends in Environmental Analytical Chemistry**, v. 9, p. 1–7, 2016.

DALIAKOPOULOS, I. N.; TSANIS, I. K.; KOUTROULIS, A.; KOURGIALAS, N. N.; VAROUCHAKIS, A. E.; KARATZAS, G. P.; RITSEMA, C. J. The threat of soil salinity: A European scale review. **Science of the Total Environment**, v. 573, p. 727–739, 2016. https://doi.org/10.1016/j.scitotenv.2016.08.177

DEMATTÊ, J. A. M.; DOTTO, A. C.; BEDIN, L. G.; SAYÃO, V. M.; E SOUZA, A. B. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. **Geoderma**, v. 337, p. 111-121, 2019a. https://doi.org/10.1016/j.geoderma.2018.09.010

DEMATTÊ, J. A. M.; DOTTO, A. C.; PAIVA, A. F. S.; SATO, M. V.; DALMOLIN, R. S.; DE ARAÚJO, M. D. S. B.; DA SILVA, E. B.; NANNI, M. R.; TEN CATEN, A.; NORONHA, N. C.; LACERDA, M. P. C.; DE ARAÚJO FILHO, J. C.; RIZZO, R.; BELLINASO, H.; FRANCELINO, M. R.; SCHAEFER, C. E. G. R.; VICENTE, L. E.; DOS SANTOS, U. J.; SAMPAIO, E. V. DE S. B.; MENEZES, R. S. C.; DE SOUZA, J. J. L. L.; ABRAHÃO, W. A. P.; COELHO, R. M.; GREGO, C. R.; LANI, J. L.; FERNANDES, A. R.; GONÇALVES, D. A. M.; SILVA, S. H. G.; DE MENEZES, M. D.; CURI, N.; COUTO, E. G.; DOS ANJOS, L. H. C.; CEDDIA, M. B.; PINHEIRO, É. F. M.; GRUNWALD, S.; VASQUES, G. M.; MARQUES JÚNIOR, J.; DA SILVA, A. J.; DE VASCONCELOS BARRETO, M. C.; NÓBREGA, G. N.; DA SILVA, M. Z.; DE SOUZA, S. F.; VALLADARES, G. S.; VIANA, J. H. M.; TERRA, F. S.; HORÁK-TERRA, I.; FIORIO, P. R.; DA SILVA, R. C.; FRADE JÚNIOR, E. F.; LIMA, R. H. C.; FILIPPINI ALBA, J. M.; DE SOUZA JUNIOR, V. S.; BREFIN, M. DE L. M. S.; RUIVO, M. DE L. P.; FERREIRA, T. O.; BRAIT, M. A.; CAETANO, N. R.; BRINGHENTI, I.; MENDES, W. DE S.; SAFANELLI, J. L.; GUIMARÃES, C. C. B.; POPPIEL, R. R.; SOUZA, A. B.; QUESADA, C. A.; DO COUTO, H. T. Z. The Brazilian soil spectral library (BSSL): A general view, application and challenges. **Geoderma**, v. 354, p. 113793, 2019. https://doi.org/10.1016/j.geoderma.2019.05.043

FANG, Q.; HONG, H.; ZHAO, L.; KUKOLICH, S.; YIN, K.; WANG, C. Visible and near-infrared reflectance spectroscopy for investigating soil mineralogy: A review. **Journal of Spectroscopy**, v. 2018, n. 1, p. 3168974, 2018. https://doi.org/10.1155/2018/3168974

GAUTAM, K.; SHARMA, P.; DWIVEDI, S.; SINGH, A.; GAUR, V. K.; VARJANI, S.; SRIVASTAVA, J. K.; PANDEY, A.; CHANG, J.-S.; NGO, H. H. A review on control and abatement of soil pollution by heavy metals: Emphasis on artificial intelligence in recovery of contaminated soil. Environmental Research, v. 225, p. 115592, 2023. https://doi.org/10.1016/j.envres.2023.115592

GHOLIZADEH, A.; BORŮVKA, L.; SABERIOON, M.; VAŠÁT, R. Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. **Applied Spectroscopy**, v. 67, n. 12, p. 1349–1362, 2013. https://doi.org/10.1366/13-07288

GHOLIZADEH, A.; BORŮVKA, L.; SABERIOON, M. M.; KOZÁK, J.; VAŠÁT, R.; NĚMEČEK, K. Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features. **Soil and Water Research**, v. 10, n. 4, p. 218-227. 2015. https://doi.org/10.17221/113/2015-SWR

GOMIERO, T. Soil degradation, land scarcity and food security: Reviewing a complex challenge. **Sustainability**, v. 8, n. 3, p. 281, 2016. https://doi.org/10.3390/su8030281

GRAVES, A. R.; MORRIS, J.; DEEKS, L. K.; RICKSON, R. J.; KIBBLEWHITE, M. G.; HARRIS, J. A.; FAREWELL, T. S.; TRUCKLE, I. The total costs of soil degradation in England and Wales. **Ecological Economics**, v. 119, p. 399–413, 2015. https://doi.org/10.1016/j.ecolecon.2015.07.026

HARDIE, M.; DOYLE, R. Measuring soil salinity. In: SHABALA, S.; CUIN, T. (eds.). Plant Salt Tolerance. **Methods in Molecular Biology**, v. 913. Humana Press, Totowa, NJ, 2012. https://doi.org/10.1007/978-1-61779-986-0_28

HONG, Y.; CHEN, Y.; CHEN, S.; SHEN, R.; HU, B.; PENG, J.; WANG, N.; GUO, L.; ZHUO, Z.; YANG, Y.; LIU, Y.; MOUAZEN, A. M.; SHI, Z. Data mining of urban soil spectral library for estimating organic carbon. **Geoderma**, v. 426, p. 116102, 2022. https://doi.org/10.1016/j.geoderma.2022.116102

HORTA, A.; MALONE, B.; STOCKMANN, U.; MINASNY, B.; BISHOP, T.F.A.; MCBRATNEY, A.B.; PALLASSER, R.; POZZA, L. Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. **Geoderma**, v. 241, p. 180-209, 2015. https://doi.org/10.1016/j.geoderma.2014.11.024

HU, B.; ZHAO, R.; CHEN, S.; ZHOU, Y.; JIN, B.; LI, Y.; SHI, Z. Heavy metal delineation based on uncertainty in a coastal industrial city in the Yangtze River Delta, China. **International Journal of Environmental Research and Public Health**, v. 15, p. 710, 2018. https://doi.org/10.3390/ijerph15040710

JACONI, A.; VOS, C.; DON, A. Near infrared spectroscopy as an easy and precise method to estimate soil texture. **Geoderma**, v. 337, p. 906–913, 2019. https://doi.org/10.1016/j.geoderma.2018.10.038

JAFARZADEH, S.; FARD, R. F.; GHORBANI, E.; SAGHAFIPOUR, A.; MORADI-ASL, E.; GHAFURI, Y. Potential risk assessment of heavy metals in the Aharchai River in northwestern Iran. **Physics and Chemistry of the Earth**, Parts A/B/C, v. 115, p. 102812. 2019. https://doi.org/10.1016/j.pce.2019.102812

KARLEN, D. L.; RICE, C. W. Soil degradation: Will humankind ever learn? **Sustainability**, v. 7, n. 9, p. 12490–12501, 2015. https://doi.org/10.3390/su70912490

KHALEDIAN, Y.; KIANI, F.; EBRAHIMI, S.; BREVIK, E. C.; AITKENHEAD-PETERSON, J. Assessment and monitoring of soil degradation during land use change using multivariate analysis. **Land Degradation & Development**, v. 28, n. 1, p. 128–141, 2017. https://doi.org/10.1002/ldr.2541

LAL, R. Soil degradation by erosion. **Land Degradation & Development**, v. 12, n. 6, p. 519–539, 2001. https://doi.org/10.1002/ldr.472

LAL, R. Restoring soil quality to mitigate soil degradation. **Sustainability**, v. 7, n. 5, p. 5875–5895, 2015. https://doi.org/10.3390/su7055875

LUCE, M. S.; ZIADI, N.; ROSSEL, R. A. V. GLOBAL-LOCAL: A new approach for local predictions of soil organic carbon content using large soil spectral libraries. **Geoderma**, v. 425, p. 116048, 2022. https://doi.org/10.1016/j.geoderma.2022.116048

MA, Y.; ROUDIER, P.; KUMAR, K.; PALMADA, T.; GREALISH, G.; CARRICK, S.; LILBURNE, L.; TRIANTAFILIS, J. A soil spectral library of New Zealand. **Geoderma Regional**, v. 35, e00726, 2023. https://doi.org/10.1016/j.geodrs.2023.e00726

MENDES, W. S.; DEMATTÊ, J. A. M.; ROSIN, N. A.; DA SILVA TERRA, F.; POPPIEL, R. R.; URBINA-SALAZAR, D. F.; BOECHAT, C. L.; SILVA, E. B.; CURI, N.; SILVA, S. H. G.; SANTOS, U. J.; VALLADARES, G. S. The Brazilian soil mid-infrared spectral library: The power of the fundamental range. **Geoderma**, v. 415, p. 115776, 2022. https://doi.org/10.1016/j.geoderma.2022.115776

OBALUM, S. E.; CHIBUIKE, G. U.; PETH, S.; OKAE-ANTWI, M. Soil organic matter as sole indicator of soil degradation. **Environmental Monitoring and Assessment**, v. 189, p. 176, 2017. https://doi.org/10.1007/s10661-017-5881-y

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8-36, 2018. https://doi.org/10.1016/j.aca.2018.04.004

PEREIRA, B.A.; SILVA, Y. J. A. B.; NASCIMENTO, C. W. A.; SILVA, Y. J. A. B.; NASCIMENTO, R. C.; BOECHAT, C. L.; BARBOSA, R.S. SINGH, V. P. Watershed scale assessment of rare earth elements in soils derived from sedimentary rocks. **Environmental monitoring and assessment**, v. 191, n. 8, p. 514. 2019. https://doi.org/10.1007/s10661-019-7658-y

PYO, J.; HONG, S. M.; KWON, Y. S.; KIM, M. S.; CHO, K. H. Estimation of heavy metals using deep neural network with visible and infrared spectroscopy of soil. **Science of the Total Environment**, v. 741, p. 140162, 2020. https://doi.org/10.1016/j.scitotenv.2020.140162

RETAMEL-SALGADO, J.; HIRZEL, J.; WALTER, I.; MATUS, I. Bioabsorption and bioaccumulation of cadmium in the straw and grain of maize (Zea mays L.) in growing soils contaminated with cadmium in different environment. **International Journal of Environmental Research and Public Health**, v. 14, p. 1399, 2017. https://doi.org/10.3390/ijerph14111399

SAKSHI; SINGH, S. K.; HARITASH, A. K. Polycyclic aromatic hydrocarbons: Soil pollution and remediation. **International Journal of Environmental Science and Technology**, v. 16, p. 6489–6512, 2019. https://doi.org/10.1007/s13762-019-02414-3

SARATHJITH, M.C.; DAS, B.S.; WANI, S.P.; SAHRAWAT, K.L. Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. **Geoderma**, v. 267, p. 1-9, 2016. https://doi.org/10.1016/j.geoderma.2015.12.031

SHAHID, S. A.; ZAMAN, M.; HENG, L. Soil Salinity: Historical Perspectives and a World Overview of the Problem. In: _____. (Ed.). **Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques**. Cham: Springer, 2018a. p. 43-53. https://doi.org/10.1007/978-3-319-96190-3_2

SHAHID, S. A.; ZAMAN, M.; HENG, L. Introduction to Soil Salinity, Sodicity and Diagnostics Techniques. In: _____. (Ed.). **Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques**. Cham: Springer, 2018b. p. 1-42. https://doi.org/10.1007/978-3-319-96190-3_1

SHEPHERD, K. D.; FERGUSON, R.; HOOVER, D.; VAN EGMOND, F.; SANDERMAN, J.; GE, Y. A global soil spectral calibration library and estimation service. **Soil Security**, v. 7, p. 100061, 2022. https://doi.org/10.1016/j.soisec.2022.100061

SHOBA, P.; RAMAKRISHNAN, S. S. Modeling the contributing factors of desertification and evaluating their relationships to the soil degradation process through geomatic techniques. **Solid Earth**, v. 7, n. 2, p. 341–354, 2016. https://doi.org/10.5194/se-7-341-2016

SINGH, A. Soil salinity: A global threat to sustainable development. **Soil Use and Management**, v. 38, n. 1, p. 39–67, 2022. https://doi.org/10.1111/sum.12772

SKALA, J.; VACHA, R.; CUPR, P. Which compounds contribute most to elevated soil pollution and the corresponding health risks in floodplains in the headwater areas of the central European watershed? **International Journal of Environmental Research and Public Health**, v. 15, p. 1146, 2018. https://doi.org/10.3390/ijerph15061146

SØRENSEN, L. K.; DALSGAARD, S. Determination of clay and other soil properties by near infrared spectroscopy. **Soil Science Society of America Journal**, v. 69, n. 1, p. 159–167, 2005. https://doi.org/10.2136/sssaj2005.0159

TETTEH, R. N. Chemical soil degradation as a result of contamination: A review. **Journal of Soil Science and Environmental Management**, v. 6, n. 11, p. 301–308, 2015. https://doi.org/10.5897/JSSEM15.0499

TRIFUOGGI, M.; DONADIO, C.; FERRARA, L.; STANISLAO, C.; TOSCANESI, M.; ARIENZO, M. Levels of pollution of rare earth elements in the surface sediments from the Gulf of Pozzuoli (Campania, Italy). **Marine pollution bulletin**, v. 136, p. 374-384. 2018. https://doi.org/10.1016/j.marpolbul.2018.09.034

VILMIN, F.; BAZIN, P.; THIBAULT-STARZYK, F.; TRAVERT, A. Speciation of adsorbates on surface of solids by infrared spectroscopy and chemometrics. **Analytica chimica acta**, v. 891, p. 79-89, 2015. https://doi.org/10.1016/j.aca.2015.06.006

VISCARRA ROSSEL, R.A.; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, v. 158, n. 1-2, p. 46-54, 2010. https://doi.org/10.1016/j.geoderma.2009.12.025

WANG, M.; ZHANG, H. Accumulation of heavy metals in roadside soil in urban area and related impact factors**. International Journal of Environmental Research and Public Health**, v. 15, p. 1064, 2018. https://doi.org/10.3390/ijerph15061064

WANG Z.; XIAO, J.; WANG, L.; LIANG, T.; GUO, Q.; GUAN, Y.; RINKLEBE, J. Elucidating the differentiation of soil heavy metals under different land uses with geographically weighted regression and self-organizing map. **Environmental Pollution**, v. 260, p. 114065, 2020. https://doi.org/10.1016/j.envpol.2020.114065

Wołejko, E., Jabłońska-Trypuć, A., Wydro, U., Butarewicz, A., & Łozowicka, B. (2020). Soil biological activity as an indicator of soil pollution with pesticides–a review. **Applied Soil Ecology**, 147, 103356. https://doi.org/10.1016/j.apsoil.2019.09.006

XIONG, X.; LIU, X.; IRIS, K. M.; WANG, L.; ZHOU, J.; SUN, X.; RINKLEBE, J.; SHAHEEN, S.M.; OK, Y.S.; LIN, Z. TSANG, D. C. Potentially toxic elements in solid waste streams: Fate and management approaches. **Environmental Pollution**, v. 253, p. 680-707, 2019. https://doi.org/10.1016/j.envpol.2019.07.012

ZHOU, B.; HUANG, D.; WU, J.; ZHU, Q.; ZHU, H. Horizontal and vertical distributions of chromium in a chromate production district of South Central China. **International Journal of Environmental Research and Public Health**, v. 15, p. 571, 2018. https://doi.org/10.3390/ijerph15040571

ZIMMERMANN, M.; LEIFELD, J.; FUHRER, J. Quantifying soil organic carbon fractions by infrared-spectroscopy. **Soil Biology and Biochemistry**, v. 39, n. 1, p. 224–231, 2007. https://doi.org/10.1016/j.soilbio.2006.07.010

ZWOLAK, A.; SARZYŃSKA, M.; SZPYRKA, E.; STAWARCZYK, K. Sources of soil pollution by heavy metals and their accumulation in vegetables: A review. **Water, Air, and Soil Pollution**, v. 230, p. 164, 2019. https://doi.org/10.1007/s11270-019-4221-y

3 INFRARED SPECTROSCOPY APPLICATIONS FOR SOIL DEGRADATION MONITORING: STATE-OF-THE-ART, CASE STUDIES, AND POTENTIAL TRAJECTORIES

Abstract

Soil degradation, driven by complex and interacting processes such as organic carbon depletion, erosion, pollution, salinity, and possibly leading to desertification, poses a critical threat to global food security, biodiversity, and environmental stability. Traditional monitoring techniques are limited by cost and logistical constraints, creating a need for alternative methods. Infrared spectroscopy (IRS) has emerged as a promising tool, enabling rapid, non-destructive, and cost-effective estimation of soil properties linked to degradation. This review synthesizes the state-of-the-art applications of IRS for monitoring soil degradation, with a special emphasis on soil carbon, pollution, salinity, and erodibility. IRS has demonstrated significant advances in modeling capabilities, particularly through the integration of machine learning approaches, as well as innovations in spectral data preprocessing and fusion. Applications have expanded across diverse environments, supported by regional and global soil spectral libraries. IRS has also shown promise in tracking erosion through associated soil properties, estimating potentially toxic elements and organic pollutants via convolutional neural networks, and predicting salinity-related properties through fused spectral data. Furthermore, IRS contributes to desertification studies by enabling early, localized assessments of soil quality decline. It is presented here case studies from Brazil employing Convolutional Neural Networks and Partial Least Squares regression models to estimate different soil attributes linked to pollution and salinity. These include concentrations of rare earth elements and heavy metals, as well as salinity indicators like electrical conductivity and exchangeable sodium percentage. Results demonstrated moderate to high predictive performance, reinforcing IRS's value for monitoring efforts. The evolution of IRS from laboratory-based analyses to field-portable devices and satellite integration highlights its growing potential for environmental diagnostics. Although challenges persist, especially in pollution assessment and large-scale desertification monitoring, advances in data fusion, hybrid modeling, and open-access data position IRS as a key tool for future soil assessment frameworks. With greater standardization and collaborative efforts, IRS is positioned to support advancements in sustainable land management and environmental monitoring.

Keywords: Soil conservation. Proximal sensing. Chemometrics. Soil monitoring.

APLICAÇÕES DA ESPECTROSCOPIA NO INFRAVERMELHO PARA MONITORAMENTO DA DEGRADAÇÃO DO SOLO: ESTADO-DA-ARTE, ESTUDOS DE CASO E POTENCIAIS TRAJETÓRIAS

Resumo

A degradação do solo, impulsionada por processos complexos e interativos como a depleção de carbono orgânico, erosão, poluição, salinidade e, em alguns casos, desertificação, representa uma ameaça crítica à segurança alimentar global, à biodiversidade e à estabilidade ambiental. Técnicas tradicionais de monitoramento enfrentam limitações relacionadas a custos e questões logísticas, evidenciando a necessidade de métodos alternativos. A espectroscopia no infravermelho (IRS) tem se destacado como uma ferramenta promissora, permitindo estimativas rápidas, não destrutivas e de baixo custo de propriedades do solo associadas à degradação. Esta revisão sintetiza o estado da arte das aplicações da IRS no monitoramento da degradação do solo, com ênfase especial no carbono, poluição, salinidade e erodibilidade. A IRS tem avançado significativamente em capacidade de modelagem, especialmente com a integração de abordagens de aprendizado de máquina, além de inovações no pré-processamento e fusão de dados espectrais. As aplicações têm se expandido em diferentes ambientes, apoiadas por bibliotecas espectrais regionais e globais. A IRS também tem mostrado potencial na detecção indireta da erosão por meio de propriedades do solo associadas, na estimativa de elementos potencialmente tóxicos e poluentes orgânicos usando redes neurais convolucionais, e na predição de atributos relacionados à salinidade com fusão espectral. Além disso, contribui com estudos de desertificação ao possibilitar avaliações locais e precoces da perda de qualidade do solo. Aqui são apresentados estudos de caso no Brasil utilizando modelos de redes neurais convolucionais e regressão por mínimos quadrados parciais para estimar atributos relacionados à poluição e salinidade, incluindo elementos terras-raras, metais pesados, condutividade elétrica e percentual de sódio trocável. Os resultados demonstraram desempenho preditivo moderado a alto, reforçando o valor da IRS para o monitoramento ambiental. A evolução da IRS de análises laboratoriais para dispositivos portáteis e integração com satélites evidencia seu crescente potencial para diagnósticos ambientais. Embora desafios persistam, sobretudo na avaliação de poluentes e monitoramento em larga escala da desertificação, avanços em fusão de dados, modelagem híbrida e dados abertos posicionam a IRS como ferramenta chave para futuras estratégias de avaliação do solo. Com maior padronização e esforços colaborativos, a IRS pode apoiar o avanço do manejo sustentável do solo e do monitoramento ambiental.

Palavras-chave: Conservação do solo. Sensoriamento proximal. Quimiometria. Monitoramento do solo.

3.1 Introduction

Soil is a thin critical layer at the interface between living and non-living systems, fundamental to life on Earth and essential for human survival, yet often neglected by humans (KOPITTKE et al., 2024). Although it is considered one of the world's most important natural resources, global soils are under threat (MONTANARELLA et al., 2016), primarily due to soil erosion, which can result of many degradation processes. Such processes impose a gradual decline in soil quality, which can lead to a decrease in food security, biodiversity, and the overall health of ecosystems. As a result of soil degradation and erosion processes, the current rate of soil loss far exceeds the formation of new soil (HORRIGAN; LAWRENCE; WALKER, 2002), underscoring the urgency of addressing these issues.

In the context of soil degradation, a major concern is the loss of soil organic carbon (SOC), which plays a critical role in soil fertility, soil structure, and the global carbon cycle. Intensive agricultural practices, deforestation, and land use change can deplete SOC levels (STOCKMANN et al., 2015; VILLARINO et al., 2017; HADDAWAY et al., 2017), reducing soil productivity and contributing to increased carbon emissions. The decrease in SOC also alters soil water retention capacity (Rawls et al., 2003), which may lead to further degradation of soil quality. The soil carbon storage is also related to climate change, presenting mitigation potential at global scale (SOMMER; BOSSIO, 2014; SONG et al., 2018), as well as susceptibility for depletion as a response to the expected global effects of the phenomenon (XU; LIU; KIELY, 2011; ALBALADEJO et al., 2013; ZHOU et al., 2019).

Another critical soil issue is the accumulation of potentially toxic elements (PTEs) such as heavy metals, which can result from industrial activities (KISKU; BARMAN; BHARGAVA, 2000; HAN et al., 2021), mining (JIANG et al., 2022; YUAN et al., 2023; ARCHUNDIA; PRADO-PANO; MOLINA-FREANER, 2024), and inputs of fertilizers and pesticides (SHI et al., 2024). These PTEs can reside within the soil for long periods, posing risks to human health and ecosystems through food chain contamination and soil toxicity (JIANG et al., 2020a; LUO et al., 2020). High levels of PTEs can also disrupt microbial communities, impairing key soil functions like nutrient cycling and organic matter decomposition (LIU et al., 2007; FAGNANO et al., 2020). Likewise, organic pollution can result from the input of fertilizers (MO et al., 2008), pesticides (ARIAS-ESTÉVEZ et al., 2008), and industrial activities (CACHADA et al., 2012a; CACHADA et al., 2012b). The presence of organic pollutants further deteriorates soil health. These pollutants can contaminate soil and water resources, leading to the disruption of natural processes such as

nutrient cycling (GIANFREDA et al., 2005) and plant growth (EOM et al., 2007). Over time, this leads to a decline in biodiversity, both above and below ground (MANZO et al., 2008), as the soil becomes increasingly hostile to life.

Soil salinity, often driven by poor irrigation practices (RENGASAMY, 2010) and over-extraction of groundwater (FAN et al., 2012), leads to the buildup of salts in the soil profile, rendering the land unsuitable for many crops (BUTCHER et al., 2016; DALIAKOPOULOS et al., 2016; SINGH, 2022). This process not only decreases agricultural productivity but also promotes desertification in arid and semi-arid regions, where vegetation is already sparse (AMEZKETA, 2006; HAJ-AMOR et al., 2022). Technically, desertification consists in the process of land degradation occurring in arid, semi-arid, and dry sub-humid regions, characterized by a decline in soil quality and a reduction in the overall capacity to support life (MARTÍNEZ-VALDERRAMA et al., 2020). This phenomenon is exacerbated by unplanned land use change (Bestelmeyer et al., 2015) combined with climate change (JIANG et al., 2019). Desertification is a major threat to food security and biodiversity and also forces communities to abandon degraded land (D'ODORICO et al., 2013), further stressing global ecosystems.

Addressing these interconnected degradation issues requires targeted approaches in environmental conservation, soil restoration, and sustainable land management. Furthermore, to act upon these problems effectively by means of environmental conservation policies, high quality information on all of these degradation processes is essential (MAO et al., 2018). However, an effective gathering of such information is often constrained by limitations in large-scale soil monitoring. Traditional monitoring methods rely on extensive fieldwork and, most importantly, complex laboratory analyses which are mostly time-consuming and resource-intensive. These challenges make it difficult to acquire timely, accurate, and comprehensive data on soil degradation processes.

Infrared spectroscopy (IRS) has emerged as a powerful and cost-effective alternative for soil analysis. Encompassing both near-infrared (NIR) and mid-infrared (MIR) spectroscopy, IRS facilitates a rapid and non-destructive evaluation of soil properties. Its greatest advantage lies in the possibility to analyze a wide array of soil attributes from a single measurement. However, this capability depends on the availability of reliable chemometric models, which need to be previously trained on enough data to support an accurate prediction. Over recent decades, this scenario has driven the growth of an entire field of study dedicated to the application of IRS for soil assessment. With ongoing advancements in IRS technology and the development of novel chemometric models, this method is evolving into a

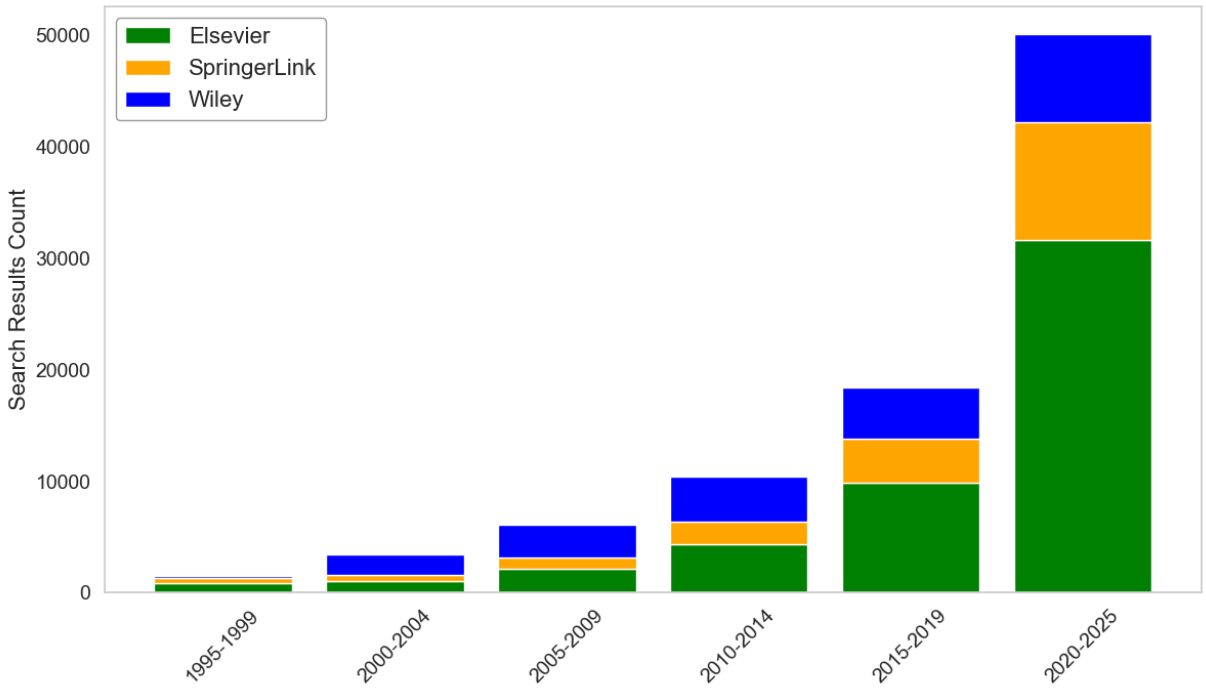
transformative solution that can offer valuable insights into soil degradation processes. In this context, the present review synthesizes the state-of-the-art applications of IRS for monitoring soil degradation, with a special emphasis on soil carbon, pollution, salinity, and erodibility.

3.2 Infrared spectroscopy for soil assessment: State-of-the-art

The history of infrared spectroscopy (IRS) as a tool for soil assessment is rooted in the broader evolution of spectroscopic methods during the mid-20th century (SUTHERLAND, 1959; THOMAS, 1991). Initially, IRS found its primary applications in general chemistry (GORE, 1958), in which it was used to identify molecular structures and functional groups through the interaction of infrared radiation with matter (STRONG, 1951; FREDERICKSON, 1954). By the 1960s and 1970s, IRS began to be explored for soil analysis, mainly focused on the study of clay minerals (AHLRICHS et al., 1965; FARMER, 1968; WADA; GREENLAND, 1970). However, widespread adoption was limited by the technological boundaries of the time, which have made it difficult to collect, process, and interpret the complex spectral data effectively.

A significant shift occurred along with the introduction of advanced statistical tools in the late 20th century (JANIK; SKJEMSTAD, 1995; PALMBORG; NORDGREN, 1996; GEYER; BRÜGGERMAN; HANSCHMANN, 1998), which allowed researchers to extract meaningful insights from spectral datasets obtained with cutting-edge equipment at that time. This progress opened pathways for practical applications in soil analysis and significantly increased the utility of the technique. Such advancements have propelled soil IRS into broader use, particularly through the modeling of mid-infrared and near-infrared spectra (JANIK; MERRY; SKJEMSTAD, 1998; MALLEY et al., 1998; REEVES; MCCARTHY; GESINGER, 1999; MCCARTHY et al., 2002; COZZOLINO; MORON, 2003). In the 21st century, there has been a sharp rise in the use of IRS for soil assessment (ZORNOZA et al., 2008; DU; ZHOU, 2009; STENBERG et al., 2010; MAIA et al., 2022; METZGER et al., 2024), which is also a well-defined pattern considering the general scientific output related to this subject for the last three decades (Figure 1).

Figure 1 - Combined search results count for ‘Soil Infrared Spectroscopy’ over time extracted from Elsevier (ScienceDirect), SpringerLink, and the Wiley Online Library.



As of today, IRS for soil assessment is an established technique, with its most common practical scenarios being the estimation of soil carbon and organic matter (ZIMMERMAN; LEIFELD; FUHRER, 2007; BAUMANN et al., 2016; HONG et al., 2022), soil texture and clay contents (SØRENSEN; DALSGAARD, 2005; GHOLIZADEH et al., 2013; TÜMSAVAŞ et al., 2019), as well as soil fertility and nutrient contents (He et al., 2007; ASKARI; O’ROURKE; HOLDEN, 2015; MUNAWAR; YUNUS; SATRIYO, 2020; YANG et al., 2020a). In addition, many studies have focused on predicting extensive sets of soil properties simultaneously (TINTI et al., 2015; SANTANA et al., 2018; SEYBOLD et al., 2019; NG et al., 2022; MAMMADOV et al., 2024; METZGER et al., 2024), highlighting the core strength of IRS.

Recent innovations in statistical modeling have further enhanced IRS capabilities, such as the application of deep learning (NG et al., 2020; BAI et al., 2023) and advanced variable selection algorithms (WANG et al., 2022; ABRANTES et al., 2023). Likewise, machine learning, coupled with spectral data fusion, have yielded high accuracy in predicting soil properties (XU et al., 2020; PYO et al., 2020; HONG et al., 2023a), even under complex field conditions (LI et al., 2024a). Convolutional neural networks trained with infrared spectra have also presented exceptional results for the estimation of soil properties (NG et al., 2019; YANG et al., 2020b; TANG et al., 2024a). Furthermore, the development of global and

regional soil spectral libraries, such as the Open Soil Spectral Library (VISCARRA ROSSEL et al., 2016; SAFANELLI et al., 2023) and the Brazilian Soil Spectral Library (DEMATTÊ et al., 2019), underscores the importance of standardization and accessibility in the context of soil IRS. These repositories enable researchers to build robust models and extend the applicability of the technique to diverse environmental settings and geographic regions.

Infrared spectroscopy has the potential to play a crucial role in addressing the escalating issues of soil degradation. By enabling efficient, cost-effective, and non-destructive soil assessment, this technology supports a more efficient gathering of information on soil conditions, which is a key to mitigate the impacts of soil degradation. Looking ahead, continued advancements in IRS hold the potential to drive sustainable agricultural practices and environmental conservation efforts, offering a powerful tool for soil quality monitoring.

3.3 Infrared spectroscopy for soil degradation monitoring

Soil degradation is a complex phenomenon resulting from various interacting processes, therefore requiring holistic and timely information for effective monitoring. Large-scale monitoring of soil degradation faces two fundamental constraints: sampling and laboratory analysis. Infrared spectroscopy (IRS) can offer a potential solution to the latter by providing faster and cost-effective estimations of soil properties (WAISER et al., 2007; HUTENGS et al., 2019), thereby improving the overall logistics of soil monitoring. However, uncertainties remain regarding the ability of IRS to predict soil properties accurately. Additionally, in order to investigate the possibility of IRS as a tool for monitoring soil degradation as a phenomenon, further work is needed to evaluate its applications on the distinct processes which impact soil quality and health (BARTHÈS; CHOTTE, 2020). Therefore, this section aims to review and synthesize the current applications of IRS for monitoring soil carbon, erosion, pollution, salinity, and desertification, including context, data, and results from our ongoing case studies.

3.3.1 Carbon

Carbon assessment in the context of soil degradation can focus on different aspects of soil carbon, such as soil organic carbon (SOC), soil inorganic carbon (SIC), soil carbon stocks, and carbon sequestration. SOC is a particularly critical indicator of soil health (NUNES et al., 2021) and plays a pivotal role in the overall productivity of ecosystems (GEORGIOU et al., 2022). Such importance has sparked the interest of improving the monitoring of this soil property by means of IRS (ZIMMERMAN; LEIFELD; FUHRER,

2007; REEVES III, 2010), an endeavor still under active progress (LI et al., 2024a; TANG et al., 2024b; DAI et al., 2025). SIC is also an important attribute in the context of environmental quality, since it contributes to almost half of terrestrial carbon stocks (SHARIFIFAR et al., 2023), with special regards to arid and semiarid areas. However, SIC is less studied than SOC and its dynamics and correlations with climate change remain still relatively unexplored (TAO et al., 2022). The prediction of SIC by means of IRS is often undertaken in combination with SOC (RIEFOLO et al., 2020; GOMEZ et al., 2020; WALDEN et al., 2024). However, recent studies have primarily focused on the IRS prediction of SIC (BAI et al., 2023; HONG et al., 2023b).

Infrared spectroscopy for soil carbon assessment can be applied using both near-infrared (NIR, 780–2500 nm) and mid-infrared (MIR, 2500–25,000 nm) regions. These spectral regions are particularly sensitive to the stretching and bending vibrations within specific functional groups that are rich in carbon and oxygen bonds (TINTI et al., 2015; VISCARRA ROSSEL; HICKS, 2015). These functional groups, including carboxyl (–COOH), carbonyl (C=O), hydroxyl (–OH), aliphatic (C–H), and aromatic (C=C) structures, are essential in characterizing both SOC and SIC from infrared spectra. The vibrational modes of these molecules generate spectral features that allow for the quantitative and qualitative characterization of carbon forms in diverse soil matrices (HONG et al., 2022). These spectral properties enable the application of IRS to a wide range of environmental conditions, driving advances for more accurate and widespread soil carbon analysis.

To provide a clear picture of recent trends, the most relevant articles from 2023 to 2025 regarding infrared spectroscopy for soil carbon analysis were gathered, allowing for a systematic overview of current research. The articles were divided in six main categories by their core approaches and ordered by number of articles: 1. Novel algorithms and modeling advances; 2. Site-specific and environmental applications; 3. Soil spectral libraries and open datasets; 4. Laboratory innovations and *in situ* applications; 5. Agricultural applications; 6. Remote sensing and aerial imagery. A total of 48 research articles were considered, and it is worthy to highlight that only 5 out of 48 use MIR spectroscopy, with the overwhelming majority of these recent studies use NIR or Vis-NIR spectroscopy.

The first category is related to innovations in the technologies used to process data generated by soil spectroscopy. Specifically, it includes recent advances in computational methods and data science, which enable enhancements to the accuracy and efficiency of soil carbon prediction. These studies focus on the development and application of sophisticated models and algorithms that improve spectral data interpretation, prediction performance, and

overall analytical robustness. The 13 research articles in this category are detailed in Table 1. The modeling innovation approach appears to be dominant not only for soil carbon spectroscopy, but for the broader advancements of the IRS technique itself.

Table 1 - Infrared spectroscopy for soil carbon research articles with novel algorithms and modeling advancements as a core approach.

Year	Title	Spectra	Authors
2023	Estimation of soil inorganic carbon with visible near-infrared spectroscopy coupling of variable selection and deep learning in arid region of China	Vis-NIR	Bai et al.
2025	Modeling soil organic carbon content using mid-infrared absorbance spectra and a nonnegative MCR-ALS analysis	MIR	Borisover et al.
2023	Partial least square regression based machine learning models for soil organic carbon prediction using visible-near infrared spectroscopy	Vis-NIR	Das et al.
2024	Subsetting reduces the error of MIR spectroscopy models for soil organic carbon prediction in the U.S. Great Plains	MIR	Dorantes et al.
2024a	Optimizing soil carbon content prediction performance by multi-band feature fusion based on visible near-infrared spectroscopy	Vis-NIR	Li et al.
2023a	Estimation of soil organic carbon content by Vis-NIR spectroscopy combining feature selection algorithm and local regression method	Vis-NIR	Liu et al.
2024	Soil Organic Carbon Prediction Based on Vis-NIR Spectral Classification Data Using GWPCA-FCM Algorithm	Vis-NIR	Miao et al.
2024	Predicting Particle Size and Soil Organic Carbon of Soil Profiles Using VIS-NIR-SWIR Hyperspectral Imaging and Machine Learning Models	Vis-NIR-SWIR	Oliveira et al.
2024b	A stacking ensemble model for predicting soil organic carbon content based on visible and near-infrared spectroscopy	Vis-NIR	Tang et al.
2025	Ensemble and transfer learning of soil inorganic carbon with visible near-infrared spectra	Vis-NIR	Wang et al.
2023b	Evaluation of data pre-processing and regression models for precise estimation of soil organic carbon using Vis-NIR spectroscopy	Vis-NIR	Wang et al.
2023a	Quantifying soil properties relevant to soil organic carbon biogeochemical cycles by infrared spectroscopy: The importance of compositional data analysis	MIR	Zhao et al.
2023b	Reducing Moisture Effects on Soil Organic Carbon Content Estimation in Vis-NIR Spectra With a Deep Learning Algorithm	Vis-NIR	Zhao et al.

Bai et al. (2023) coupled variable selection with deep learning to estimate SIC in arid regions, achieving exceptional accuracy with R^2 = 0.93 and RMSE = 1.26 g kg^{-1} in calibration, and R^2 = 0.92 and RMSE = 1.37 g kg^{-1} in validation, while reducing spectral variables by over 97%. Borisover et al. (2025) pioneered the use of nonnegative multivariate curve resolution with alternating least squares (MCR-ALS) for modeling SOC content from MIR spectra, offering a novel mechanistic perspective that better captures the underlying chemical processes. Das et al. (2023) optimized partial least square regression models with advanced machine learning techniques to improve SOC prediction, achieving significant improvements in accuracy through careful model selection and optimization. Dorantes et al. (2024) showed that strategic subsetting of soil samples can significantly reduce prediction errors in MIR spectroscopy models, particularly in the U.S. Great Plains region.

Li et al. (2024a) proposed a multi-band feature fusion technique to optimize prediction performance by leveraging information from different spectral ranges, showing improved accuracy through better feature utilization. Liu et al. (2023a) combined feature selection algorithms with local regression methods to improve the precision of soil carbon estimation, achieving better results than global modeling approaches. Miao et al. (2024) introduced a novel method by combining geographically weighted principal component analysis with fuzzy c-means clustering to enhance SOC prediction, achieving remarkable results with R^2 = 0.83 and RPIQ = 2.95, representing 10.33% and 18% improvements respectively compared to unclassified modeling. Oliveira et al. (2024) integrated Vis-NIR-SWIR hyperspectral imaging with machine learning for simultaneous prediction of particle size and SOC content, demonstrating the potential of multi-property prediction from spectral data.

Tang et al. (2024b) advanced the field by creating a stacking ensemble model that integrates multiple machine learning algorithms for superior prediction performance, demonstrating enhanced accuracy through model diversity. Wang et al. (2025) introduced an ensemble and transfer learning approach that enhances SIC prediction across diverse soil types, achieving R^2 = 0.81 and demonstrating superior model transfer capabilities with improvements of 0.05-0.21 in R^2 and reductions of 0.33-1.44 g kg^{-1} in RMSE compared to individual models. Wang et al. (2023b) provided a comprehensive evaluation of pre-processing techniques and regression models, establishing best practices for spectral analysis through systematic comparison of different approaches. Zhao et al. (2023a) applied compositional data analysis to infrared spectroscopy, enabling more accurate quantification of soil properties relevant to organic carbon biogeochemical cycles through better handling of the compositional nature of soil data. Zhao et al. (2023b) developed a deep learning-based

network (MIRNet) that effectively removes moisture interference from spectral measurements, achieving R^2 values of 0.92.

This first set of articles highlights the significant strides made in advancing computational methods and modeling approaches for soil carbon analysis with infrared spectroscopy. These innovations not only enhance the precision and robustness of soil carbon predictions but also underscore a broader methodological evolution that is shaping the future of soil spectroscopy research. Notably, the integration of machine learning, ensemble models, and advanced data preprocessing techniques has yielded remarkable improvements in prediction accuracy, paving the way for more reliable and scalable applications.

Following these methodological advancements, the next category focuses on site-specific and environmental applications, where infrared spectroscopy is employed to assess soil carbon dynamics across diverse environments. This line of research is crucial, since one of the most important limitations for the generalized use of this technique is the lack of standardization considering different local environmental factors. The 11 articles listed for this category are detailed in Table 2.

Table 2 - Infrared spectroscopy for soil carbon research articles with site-specific and environmental applications.

Year	Title	Spectra	Authors
2024	Can Soil Organic Carbon in Long-Term Experiments Be Detected Using Vis-NIR Spectroscopy?	Vis-NIR	Barbetti et al.
2025	Developing a near-infrared spectroscopy calibration algorithm for soil organic carbon content in South Africa	NIR	Cloete et al.
2023	Near-Infrared Spectroscopy Assessment of Soil Organic Carbon Stock in a Colombian Oxisol	NIR	Fernández-Martínez et al.
2024	Establishing Visible-Near Infrared Spectroscopy for the Prediction of Soil Organic Carbon in Arid Croplands of Arequipa, Peru	Vis-NIR	Foster et al.
2024	Estimation of soil organic carbon content using visible and near-infrared spectroscopy in the Red River Delta, Vietnam	Vis-NIR	Hau et al.
2024	Prediction of wetland soil carbon storage based on near infrared hyperspectral imaging and deep learning	NIR	Jia et al.
2024	Advancing Soil Organic Carbon and Total Nitrogen Modelling in Peatlands: The Impact of Environmental Variable Resolution and Vis-NIR Spectroscopy Integration	Vis-NIR	Mendes et al.

2024	Using visible-near infrared spectroscopy to estimate whole-profile soil organic carbon and its fractions	Vis-NIR	Qi et al.
2023	Comparison of near and mid-infrared reflectance spectroscopy for the estimation of soil organic carbon fractions in Madagascar agricultural soils	NIR/MIR	Ramifehiarivo et al.
2023	Predicting soil carbon in granitic soils using Fourier-transform mid-infrared (FT-MIR) spectroscopy	MIR	Seboto et al.
2024	Prediction of soil organic carbon and total nitrogen affected by mine using Vis-NIR spectroscopy coupled with machine learning algorithms in calcareous soils	Vis-NIR	Zhang et al.

Barbetti et al. (2024) evaluated the reliability of Vis-NIR spectroscopy in long-term experiments, finding that site-specific spectral libraries combined with Cubist and SVM models effectively detected SOC changes, with optimal performance achieved when including standard analyses of at least 10% of monitoring samples in the training set. In South Africa, Cloete et al. (2025) developed NIR spectroscopy calibration algorithms for SOC content, revealing that while regional-scale models showed limited accuracy, local calibration models for specific catchments demonstrated high accuracy (RMSE < 0.1 and RPIQ > 1.5). The study highlighted the importance of sample state and sampling design in model performance, and noted that South African samples were underrepresented in global datasets, leading to poor predictions using open spectral libraries. Fernández-Martínez et al. (2023) used near-infrared diffuse reflectance spectroscopy, combined with depth-interval sampling and geostatistical validation, to develop a site-specific model for estimating soil organic carbon stocks in Colombian oxisols, showcasing the method's potential for cost-effective, environmentally focused monitoring.

Foster et al. (2024) developed a Vis-NIR spectroscopy-based predictive model for monitoring soil organic carbon in arid croplands of southern Peru, innovatively integrating soils from contrasting traditional and newly irrigated systems into a regional spectral library to improve applicability across rapidly changing desert agriculture. Jia et al. (2024) combined NIR hyperspectral imaging with deep learning for wetland soil carbon storage prediction, while Zhang et al. (2024) demonstrated the effectiveness of Vis-NIR spectroscopy coupled with machine learning algorithms for predicting SOC in calcareous soils affected by mining activities. Hau et al. (2024) successfully estimated SOC content in the Red River Delta of Vietnam using Vis-NIR spectroscopy, while Ramifehiarivo et al. (2023) compared near and

mid-infrared reflectance spectroscopy for estimating SOC fractions in Madagascar agricultural soils.

Mendes et al. (2023) demonstrated that combining Vis-NIR spectroscopy with high-resolution LiDAR-derived and medium-resolution SRTM-derived digital elevation models (DEMs) improved the prediction of SOC and total nitrogen in peatlands, highlighting the relevance of integrating local environmental factors to enhance model accuracy across heterogeneous landscapes. Qi et al. (2024), developed approaches for estimating whole-profile SOC and its fractions using visible-near infrared spectroscopy. Meanwhile, Seboto et al. (2023) successfully predicted soil carbon in granitic soils using Fourier-transform MIR spectroscopy. These studies highlight the potential of infrared spectroscopy for comprehensive soil carbon assessment across diverse environmental contexts, as well as different soil types and depths. However, they also emphasize the importance of considering local soil characteristics and environmental conditions when developing calibration models, which can be considered the current main constraint for the generalization of this technique.

Shifting to the next category, soil spectral libraries and open datasets are fundamental resources for advancing and supporting the broader adoption of IRS for soil assessment. These libraries provide standardized, high-quality reference spectra linked to measured soil properties, enabling the development of robust calibration models and facilitating data sharing across regions and research initiatives. The availability of these open datasets enhances reproducibility, accelerates methodological innovation, and lowers barriers for researchers and practitioners working in data-scarce regions. The following set of 7 articles (Table 3) highlights key efforts in the application of soil spectral libraries and open-access data in support of sustainable land management and soil monitoring efforts.

Table 3 - Infrared spectroscopy for soil carbon research articles with soil spectral libraries and open datasets applications.

Year	Title	Spectra	Authors
2024	Soil organic carbon and total nitrogen multivariate modelling from diverse FT-NIR spectral dataset	NIR	Adejumo et al.
2024	Quantification of soil organic carbon in particle size fractions using a near-infrared spectral library in West Africa	NIR	Cambou et al.
2025	Prediction of soil organic carbon fractions in tropical cropland using a regional visible and near-infrared spectral library and machine learning	Vis-NIR	Dai et al.

2023	Prediction of Soil Organic Carbon Contents in Tibet Using a Visible Near-Infrared Spectral Library	Vis-NIR	Jia et al.
2024a	Estimation of soil organic carbon in LUCAS soil database using Vis-NIR spectroscopy based on hybrid kernel Gaussian process regression	Vis-NIR	Liu et al.
2023	Improved Soil Organic Carbon Prediction in a Forest Area by Near-Infrared Spectroscopy Spiking of a Soil Spectral Library	NIR	Long et al.
2024	Organic carbon estimation in a regional soil Vis-NIR database supported by unsupervised learning and chemometrics techniques	Vis-NIR	Miloš et al.

Adejumo et al. (2024) leveraged a diverse Fourier Transform NIR spectral dataset to develop robust multivariate models for SOC and total nitrogen prediction, highlighting the value of comprehensive spectral databases. Cambou et al. (2024) used an existing NIR spectral library comprised of West African soils with 181 samples from five countries, enabling accurate quantification of SOC in particle size fractions from 94 target samples, demonstrating the importance of region-specific spectral librarires. Dai et al. (2025) utilized a regional Vis-NIR spectral library, successfully predicting SOC fractions while testing different machine learning approaches, also showcasing the potential of regionally calibrated spectral databases.

Jia et al. (2023) leveraged a dedicated Vis-NIR spectral library for Tibetan soils, further demonstrating how region-specific libraries can enhance prediction accuracy in unique environments. Liu et al. (2024a) innovatively combined the LUCAS soil database with a hybrid kernel Gaussian process regression to improve SOC estimation, showcasing the potential of integrating large-scale soil databases with novel modeling techniques. Long et al. (2023) demonstrated the effectiveness of spiking existing spectral libraries with local samples, a strategy that significantly improved SOC prediction accuracy in forest areas and highlighted the importance of local calibration in spectral databases. Novel computational methods have shown potential to enhance the utility of soil spectral libraries. Miloš et al. (2024) leveraged a comprehensive regional soil Vis-NIR database from Croatia, building models with an approach that combines unsupervised learning with chemometrics, demonstrating how advanced analytical methods can maximize the value of spectral libraries.

These studies collectively illustrate the growing importance of well-curated soil spectral libraries and open datasets in SOC prediction, with particular emphasis on the need

for region-specific calibration and the integration of advanced computational techniques to extract maximum value from these databases.

The next category explores recent advances in laboratory methodologies and *in situ* field applications of infrared spectroscopy. These approaches are pushing the boundaries of how infrared spectroscopy is applied. Innovations such as new sample preparations and portable spectroscopy technologies have enhanced the efficiency of soil properties prediction both in controlled settings and directly in the field. In the case of field measurements, real-time decision-making is enabled, reducing reliance on costly laboratory infrastructure and expanding the accessibility of soil diagnostics to a wider range of users. Table 4 presents the 6 listed articles that illustrate these approaches.

Table 4 - Infrared spectroscopy for soil carbon research articles with laboratory innovations and *in situ* applications.

Year	Title	Spectra	Authors
2025	Soil organic carbon measurements influence FT-NIR model training in calcareous soils of Saskatchewan	NIR	Adejumo et al.
2024	In-situ prediction of soil organic carbon contents in wheat-rice rotation fields via visible near-infrared spectroscopy	Vis-NIR	Dai et al.
2023	Large-scale measurement of soil organic carbon using compact near-infrared spectrophotometers effect of soil sample preparation and the use of local modelling	NIR	Fonseca et al.
2024b	Micro-Near-Infrared (Micro-NIR) sensor for predicting organic carbon and clay contents in agricultural soil	NIR	Liu et al.
2023	On-the-Go Vis-NIR Spectroscopy for Field-Scale Spatial-Temporal Monitoring of Soil Organic Carbon	Vis-NIR	Reyes et al.
2025	Combining temperature ramp dry combustion and mid-infrared spectroscopy for enhanced soil organic carbon characterisation	MIR	Walden et al.

Adejumo et al. (2025) investigated the influence of different SOC measurement methods on Fourier Transform NIR model prediction performance in calcareous soils, revealing that the choice of reference method significantly impacts model accuracy and calibration transfer. Their work highlights the importance of method standardization for reliable spectroscopic predictions specific for high-pH soils. Dai et al. (2024) developed an innovative *in situ* approach for SOC prediction in wheat-rice rotation fields using Vis-NIR

spectroscopy, achieving promising results without sample preparation, which represents a significant advancement in field-scale monitoring capabilities.

The development of portable and compact spectroscopic solutions has expanded the possibilities for large-scale SOC assessment. Fonseca et al. (2023) demonstrated the effectiveness of compact NIR spectrophotometers for large-scale SOC measurement, with their study showing that local modeling approaches can overcome the limitations of portable instruments. Liu et al. (2024b) introduced a novel micro-NIR sensor application for simultaneous prediction of SOC and clay contents, offering a cost-effective solution for agricultural soil characterization. Their work represents a significant step forward in miniaturized sensor technology for soil analysis.

Field-scale monitoring and enhanced characterization methods have also seen notable improvements. Reyes et al. (2023) implemented an on-the-go Vis-NIR spectroscopy system for spatial-temporal monitoring of SOC, providing valuable insights into the dynamics of soil carbon at the field scale. Their approach demonstrates the potential for real-time, high-resolution soil mapping. In the context of laboratory advancements, Walden et al. (2025) presented an innovative combination of temperature ramp dry combustion with mid-infrared spectroscopy, offering enhanced characterization of SOC through the integration of thermal and spectral analysis. Collectively, these studies underscore the rapid progress in both laboratory and field-based spectroscopic techniques, paving the way for more accurate, efficient, and scalable soil carbon assessment.

The next category explores the integration of remote sensing with infrared spectroscopy, which is opening new possibilities for large-scale soil monitoring. By leveraging data from satellites, drones, and other aerial platforms, the reach of infrared spectroscopy can be extended beyond point-based measurements to be combined with spatially continuous assessments of soil properties. This combination enhances the ability to detect variability across fields and regions, supports precision agriculture, and contributes to informed land management and conservation practices. Table 5 features 6 articles that demonstrate how remote sensing technologies, when paired with spectral data, are being applied to map and monitor soils at broader spatial scales.

Table 5 - Infrared spectroscopy for soil carbon research articles with remote sensing and aerial imagery applications.

Year	Title	Spectra	Authors
2023	Predicting soil organic carbon in cultivated land across	Vis-NIR	Bao et al.

	geographical and spatial scales Integrating Sentinel-2A and laboratory Vis-NIR spectra		
2023	Estimate of carbon stock in the soil via diffuse reflectance spectroscopy (vis/nir), air and orbital remote sensing	Vis-NIR	Faria et al.
2023b	Prediction of soil organic carbon in soil profiles based on visible-near-infrared hyperspectral imaging spectroscopy	Vis-NIR	Liu et al.
2024	Visible, near-infrared, and shortwave-infrared spectra as an input variable for digital mapping of soil organic carbon	Vis-NIR-SWIR	Khosravi et al.
2024	Rapid Analysis of Soil Organic Carbon in Agricultural Lands Potential of Integrated Image Processing and Infrared Spectroscopy	Vis-NIR	Senevirathne et al.
2023a	Using Machine-Learning Algorithms to Predict Soil Organic Carbon Content from Combined Remote Sensing Imagery and Laboratory Vis-NIR Spectral Datasets	Vis-NIR	Zayani et al.

Bao et al. (2023) demonstrated the effectiveness of integrating Sentinel-2A satellite imagery with laboratory Vis-NIR spectra for predicting SOC across different geographical scales, highlighting the potential of combining remote sensing data with ground-based measurements. Similarly, Faria et al. (2023) showed that combining diffuse reflectance spectroscopy with aerial and orbital remote sensing can effectively estimate soil carbon stocks, providing a more efficient alternative to traditional laboratory methods. The integration of machine learning with remote sensing data has opened new possibilities for SOC prediction. Zayani et al. (2023a) developed a novel approach using machine learning algorithms to predict SOC content by combining remote sensing imagery with laboratory Vis-NIR spectral datasets, achieving high accuracy in their predictions. Liu et al. (2023b) further advanced this field by implementing Vis-NIR hyperspectral imaging spectroscopy for SOC prediction in soil profiles, demonstrating the potential of this technology for detailed soil characterization.

Recent innovations have also explored the combination of multiple spectral ranges for enhanced SOC mapping. Senevirathne et al. (2024) demonstrated the potential of integrated image processing and infrared spectroscopy for rapid analysis of SOC in agricultural lands, while Khosravi et al. (2024) proposed a novel methodology using visible, near-infrared, and shortwave-infrared (Vis-NIR-SWIR) spectra for digital mapping of SOC, showing that interpolated spectra combined with environmental predictor covariates can significantly improve mapping accuracy and reduce uncertainty in SOC predictions. Together, these

studies underscore the growing synergy between remote sensing, spectral technologies, and data-driven approaches, paving the way for more scalable, accurate, and cost-effective monitoring of soil organic carbon across diverse landscapes.

Moving on to the last category, it is important to note that soil carbon plays a critical role in maintaining soil fertility, structure, and overall productivity, all of which are essential for sustainable agriculture. The ability to rapidly and accurately assess soil carbon through infrared spectroscopy offers immense potential to inform precision agriculture practices, optimize soil management strategies, and ultimately improve crop yields while supporting environmental sustainability. Therefore, the following set of 5 articles (Table 6) demonstrates how infrared spectroscopy is being applied within agricultural contexts.

Table 6 - Infrared spectroscopy for soil carbon research articles with agricultural applications.

Year	Title	Spectra	Authors
2025	Prediction of soil organic carbon stock along layers and profiles using Vis-NIR laboratory spectroscopy	Vis-NIR	Dharumarajan et al.
2024	Monitoring Land Management Practices Using Vis-NIR Spectroscopy Provides Insights into Predicting Soil Organic Carbon and Limestone Levels in Agricultural Plots	Vis-NIR	Herranz-Luque et al.
2024	Development of near-infrared spectroscopy (NIRS) for estimating organic matter total carbon and total nitrogen in agricultural soil	NIR	Santasup et al.
2023	Detecting the temporal trend of cultivated soil organic carbon content using visible near infrared spectroscopy	Vis-NIR	Zayani et al.
2024	VIS-NIR spectroscopy and environmental factors coupled with PLSR models to predict soil organic carbon and nitrogen	Vis-NIR	Zhu et al.

All of these articles explore applications of soil infrared spectroscopy for soil carbon prediction in a given agricultural context. Dharumarajan et al. (2025) demonstrated the effectiveness of Vis-NIR spectroscopy in predicting soil organic carbon stocks across different soil layers, aiming to offer farmers a non-destructive method to assess soil health. Herranz-Luque et al. (2024) showed how Vis-NIR can monitor the impact of different agricultural practices on soil organic carbon and limestone levels, providing crucial feedback for sustainable farming. Santasup et al. (2024) developed a rapid NIRS method for estimating soil organic matter, carbon, and nitrogen, enabling farmers to quickly assess soil fertility.

Zayani et al. (2023b) tracked soil carbon changes over five years in a French agricultural catchment, demonstrating the long-term monitoring capabilities of Vis-NIR for sustainable agriculture. Finally, Zhu et al. (2024) enhanced soil property predictions by combining Vis-NIR with environmental factors and PLSR models, providing a comprehensive approach to soil analysis that considers both spectral data and environmental conditions. These studies collectively demonstrate how spectroscopy is transforming soil analysis in agriculture, offering an alternative for monitoring soil health and guiding agricultural practices.

Infrared spectroscopy has shown strong potential as a rapid, cost-effective, and scalable method for assessing soil carbon – particularly SOC, which is a key indicator of soil health and degradation. Recent advances in modeling techniques, site-specific and environmental insights, field applications, spectral library development, and integration with remote sensing hold the potential to improve the accuracy and applicability of IRS across diverse scenarios. While challenges such as standardization and regional calibration persist, the growing body of research demonstrates that IRS is moving beyond experimental stages and into practical use for soil carbon monitoring.

Altogether, the reviewed studies in this section highlight a methodological evolution that is transforming how SOC is quantified, from traditional laboratory-based analyses to non-destructive methods, as well as portable sensors for *in situ* measurements, and large-scale spatial assessments. This progress underscores IRS as a valuable tool in efforts to better understand and manage soil carbon dynamics. With its expanding capabilities, IRS is well positioned to support sustainable land management and long-term soil monitoring.

3.3.2 Pollution

Inputs of hazardous chemicals on the soil matrix can raise several problems in regard of environmental quality. These inputs can be comprised of both organic contaminants and potentially toxic elements (PTEs). Organic contaminants in the soil, such as polycyclic aromatic hydrocarbons (PHAs), polychlorinated biphenyls (PCBs), organochlorine pesticides (OCPs), among others, are mostly a direct result of industrialization in both urban and rural areas (Sun et al., 2018). The presence of these compounds in the soil has been an object of intensive research in the pursuit of better management of soil quality (DUARTE; MATOS; SENESI, 2018), since they contribute significantly to the degradation of soil quality (Manzo et al., 2008). Organic pollutants in soil interferes with essential natural processes like nutrient cycling (GIANFREDA et al., 2005) and plant development (EOM et al., 2007), leading to an

overall decline in biodiversity. Monitoring the presence of organic contaminants presents challenges similar to those found for PTEs, and recent approaches aim for the use of IRS for the prediction of such compounds, mostly for petroleum contaminated soils (NG et al., 2017; DOUGLAS et al., 2018; KARIMIAN et al., 2024) using both vis-NIR and MIR spectroscopy.

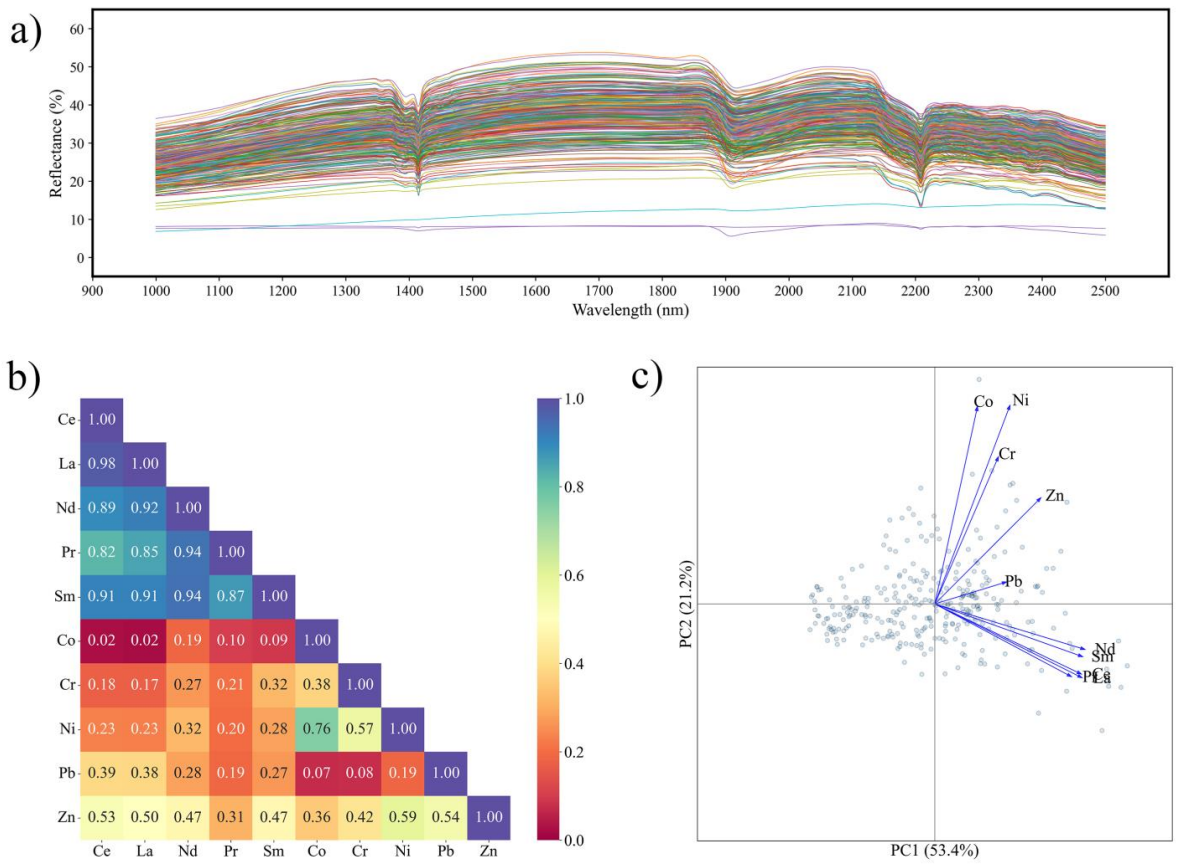
Potentially toxic elements (PTEs) are metals and metalloids that can harm organisms and ecosystems, especially when present at elevated concentrations in the soil. High levels of these elements, such as heavy metals and rare earth elements, can lead to detrimental effects on agriculture and human health (KHAN et al., 2021). Moreover, the persistence of these contaminants in the soil complicates remediation efforts, as many PTEs do not easily degrade or leach out, leading to long-term environmental and health risks (HAN et al., 2021). In addition to their environmental persistence and toxicity, PTEs are significant because of their complex interactions with soil components, which influence their mobility, bioavailability, and potential for plant uptake. These interactions are governed by factors such as pH, redox conditions, organic matter content, and mineralogy, making their behavior highly site-specific. Given the problematic nature of these elements, monitoring their contents in the soil is a task of utmost importance, especially for accurate risk assessment and the development of effective management and remediation strategies.

Recent advances highlight the role of innovative machine learning and artificial intelligence approaches to identify sources, characterize, and quantify soil PTEs (GAUTAM et al., 2023). In contrast to organic contaminants, PTEs have been mainstream targets for IRS prediction approaches (SHI et al., 2014; GHOLIZADEH et al., 2020; NAWAR et al., 2020; KÄSTNER et al., 2022; MAIA et al., 2022). However, the prediction of soil PTEs by IRS is marked by a lack of standardization and its effectiveness is very relative to local characteristics of the study areas, such as environmental heterogeneity, soil-specific attributes, and sampling size.

In this context, we have gathered a dataset consisting of 292 samples (Figure 2) containing elemental concentrations of Ce, La, Nd, Pr, Sm, Co, Cr, Ni, Pb, and Zn, as well as their near-infrared (NIR) spectra (1000 – 2500 nm). This dataset is comprised of samples obtained from different locations in the States of Pernambuco, Rio Grande do Norte, and Piauí, in the northeast of Brazil. The concentrations of the elements were obtained by open acid digestion using nitric, hydrochloric, perchloric, and hydrofluoric acids, followed by analysis with inductively coupled plasma optical emission spectroscopy (ICP-OES). The near-infrared spectra were measured with an FT-IR/NIR spectrometer (Frontier/PerkinElmer) coupled with a Near-Infrared Reflectance Accessory (NIRA), using a 2-nm resolution and a

0.5-nm window, with multiple accumulations per sample to ensure representativeness. The approach here was to compare the performance of a Convolutional Neural Network (CNN) with the classical Partial Least Squares regression (PLSR) in the prediction of the elemental concentrations. Both CNN and PLSR are well-established algorithms in scientific research (GARTHWAITE, 1994; FALEH; KACHOURI, 2023).

Figure 2 - Overview of the pollution sample set: a) Near-infrared spectra; b) Spearman correlation matrix between the elemental concentrations; c) Principal component analysis biplot.



The data were divided into an 80% training subset (n = 233) and a 20% validation subset (n = 59). All data handling – including loading, cleaning, transforming, splitting, and processing – were conducted with Python (version 3.11.0) using the libraries ‘pandas’ (version 1.5.3) and ‘scikit-learn’ (version 1.2.2). Spectral preprocessing was applied with the Savitzky-Golay second-derivative filter to smooth the NIR spectra by attenuating high-frequency noise, correcting baseline drift, and highlighting spectral features (SAVITZKY; GOLAY, 1964). The preprocessed spectra were used only for the PLS models. For each soil attribute, a PLS regression model was tuned via a grid-search over 2-10 number

of components and 10-fold cross-validation; the optimal model was then calibrated with the training set and tested on the validation set, ultimately generating the prediction performance metrics (Table 7).

A Convolutional Neural Network (CNN) was employed to leverage deep learning for the prediction. CNNs are particularly well-suited for handling multidimensional data structures through the convolution operation, where a filter or kernel moves across the input array and performs pointwise multiplications summed into a single output value (Padarian et al., 2019). While an individual convolutional layer can detect simple patterns, stacking multiple layers enables the network to learn increasingly abstract and intricate features (LECUN; BENGIO, 1995). CNNs are commonly used for data with image-like structures (PANDEY; JAIN, 2022; THAKUR et al., 2023). However, infrared spectra can be interpreted as a one-dimensional image (CHAKRAVARTULA et al., 2022; SHANG et al., 2023). Previous studies have shown that CNNs can predict multiple soil properties from infrared spectral data with high accuracy (NG et al., 2019).

In this work, we adopted a CNN architecture based on depthwise separable convolutions (Appendix A), as suggested by Yang et al. (2020b). This design was optimized for efficiently capturing both localized and abstract patterns in raw data. The network was fed with the training sample set using the ‘adam’ optimization algorithm (KINGMA; BA, 2015; REDDI; KALE; KUMAR, 2018), with a default learning rate of 0.001 and no weight decay applied. Training ran for 200 epochs with a batch size of 5. Then, the trained networks were tested with the validation sample set, generating the prediction performance metrics (Table 7). The implementation was done in Python using the ‘keras’ module from the TensorFlow framework (version 2.12.0rc0).

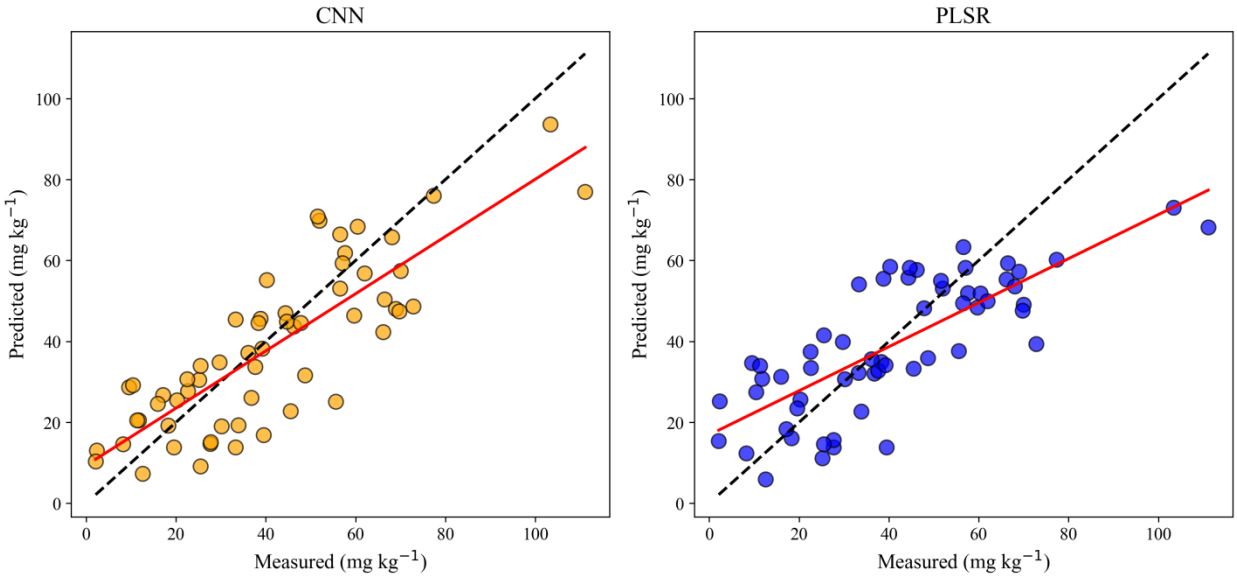
Table 7 - Prediction performance metrics with CNN and PLSR for each potentially toxic element.

	Ce	La	Nd	Pr	Sm	Co	Cr	Ni	Pb	Zn
						CNN				
RMSE	34.65	13.41	8.64	4.94	2.14	3.28	19.51	6.42	8.29	14.34
R^2	0.57	0.67	0.52	0.33	0.60	0.36	-0.01	0.42	0.50	0.54
RPD	1.52	1.77	1.48	1.23	1.59	1.32	1.00	1.43	1.45	1.51
						PLSR				
RMSE	34.30	15.01	9.28	4.79	2.44	3.57	18.67	5.95	8.69	16.04
R^2	0.58	0.58	0.45	0.37	0.48	0.23	0.07	0.50	0.45	0.42
RPD	1.56	1.56	1.35	1.28	1.42	1.16	1.05	1.42	1.36	1.32

CNN Convolutional neural network; PLSR Partial least squares regression; RMSE Root mean squared error; R² Coefficient of determination; RPD Ratio of prediction to deviation

The performance results reveal that both modeling approaches achieved comparable predictive performance across the evaluated elements, though with varying degrees of success depending on the specific variable. The best prediction amongst all elements was for La using CNN (Figure 3), achieving R² and RPD values of 0.67 and 1.77, respectively. The CNN slightly outperformed PLSR for most elements, as evidenced by lower RMSE and higher R² and RPD values, particularly for La, Nd, Sm, Pb, and Zn. However, for some elements such as Ce and Ni, both models demonstrated similar levels of predictive accuracy, with R² values around 0.42-0.58 and RPDs over 1.4, indicating moderate reliability. The predicted *versus* measured scatter plots for all elements can be found in Appendix B.

Figure 3 - Predicted *versus* measured scatter plots for Lanthanum using both Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).



Notably, predictions for Pr, Co, and Cr were poor with both approaches (Table 7), as reflected by R² values below 0.50 and RPDs below 1.4, suggesting that the models struggled to capture the relevant spectral features for these elements, especially for Cr, which achieved near-zero R² values and RPD of 1. These findings align with previous research indicating that model performance is element-specific and often constrained by the inherent spectral sensitivity to particular soil properties (VISCARRA ROSSEL; WEBSTER, 2012). Overall, while CNN showed a slight advantage in handling the raw spectral data without

preprocessing, the PLSR models combined with Savitzky-Golay spectral preprocessing remained competitive in this scenario.

The Spearman correlation matrix for the elements (Figure 2b) underscores strong collinearity among the rare earth elements (Ce, La, Nd, Pr, Sm), with correlation coefficients ranging from 0.82 to 0.98, demonstrating their high covariance in the studied soils. This multicollinearity may have contributed to the superior predictive success observed for these elements, as spectral features related to one may indirectly assist in the prediction of others. In addition, Pr demonstrated the lowest correlations with Ce, La, and Sm among all REEs, and also the poorest prediction performance, further strengthening the perspective that performance improvement is relative to a higher covariance.

On the other hand, transition metals such as Co, Cr, and Ni showed low correlations with REEs and moderate correlations among themselves, hinting at distinct geochemical features that may not be well expressed in the spectral data, further explaining the limited model performance for these elements, especially for Co and Cr. Furthermore, Zn and Pb occupied an intermediate position. Zn exhibited moderate correlations with all REEs and transition metals, possibly reflecting shared geochemical features or indirect associations with spectrally active soil components, while Pb showed weak correlations with most elements except for Zn, indicating a more independent behavior potentially linked to distinct inputs or processes.

The PCA biplot (Figure 2c) reveals that the first two principal components explain a substantial portion of the variance in the dataset (PC1 = 53.4%; PC2 = 21.2%). The strong loading of REEs along PC1 aligns with their high correlations and suggests they share similar variance structures, likely related to soil geochemistry. The transition metals Co, Ni, Cr, and Zn load more heavily along PC2, demonstrating a separation from REEs in the multivariate space. This suggests that different geochemical processes or soil fractions control the distribution of these elements, which may not be fully captured by the spectral signatures used by the models. Meanwhile, Pb exhibited a more diffuse loading with a lower impact on both PC1 and PC2, reinforcing its complex and partially independent variance structure. Overall, these visualizations corroborate and refine the interpretation that prediction success depends heavily on the degree to which soil elemental concentrations are spectrally expressed and how they correlate with other features detectable by NIR spectroscopy.

Although the predictions for most potentially toxic elements in this study showed limited accuracy, the results suggest a hopeful future. With expected improvements in both spectral technology and data analysis algorithms, predicting these elements more accurately

will likely become more feasible. Thus, as sensors get better and models become more advanced, it will be possible to estimate contents of challenging pollutants from infrared spectral data more reliably. This progress could significantly assist environmental monitoring by allowing faster and non-destructive detection of soil contamination, supporting better pollution control across different ecosystems.

3.3.3 Erosion

The current global rate of soil erosion far exceeds the natural rate of soil formation (WUEPPER et al., 2019), posing a significant threat to food security and overall ecosystem productivity (QUINTON et al., 2010). Soil erosion is a natural geomorphological process, but its importance lies in its occurrence as a consequence of land degradation, which leads to the loss of fertile soil and the deposition of sediments in water bodies, reducing water quality. Since soil erosion is a long-established and studied process, there are only highly specific advances to add to its existing body of knowledge. However, there is a pressing need for research into innovative techniques and strategies for monitoring and preventing this process (POESEN, 2017). Recent studies analyzing the current state of soil erosion modeling (BATISTA et al., 2019; BORRELLI et al., 2021) highlight the perspective that significant contributions can still be made to this field, especially through innovation in technology and statistics.

In the context of IRS applied to soil erosion monitoring, most studies have been focusing on the prediction of soil properties that correlate with erosion (CONFORTI et al., 2013; ELLERBROCK; GERKE; DEUMLICH, 2016), mainly soil organic matter. Further within the scope of monitoring, approaches that combine IRS with remote sensing and satellite imagery analysis are also common (TENG et al., 2016), aiming to achieve large-scale solutions that are less dependent on field and laboratory work. Applications of IRS can also be found for sediment tracing and channel erosion monitoring (NI et al., 2019; TIECHER et al., 2016; XIONG et al., 2024). In this context, IRS is being used as a diagnosis tool for negative outcomes. Sediment deposition in water bodies is a direct consequence of soil erosion, and the improvement of its monitoring is also a pressing issue.

Nonetheless, new approaches are currently exploring the direct application of IRS to estimate soil erodibility (WANG et al., 2016; JIANG et al., 2020b; SALEHI-VARNOUSFADERANI et al., 2022), which is an intrinsic soil property that represents its susceptibility to erosion. A recent study by Teixeira et al. (2023) demonstrated the viability of using near-infrared spectroscopy (NIR) to predict soil erodibility indices in a watershed

undergoing desertification in Brazil. By analyzing spectral data from 214 soil samples and applying partial least squares regression (PLSR), the authors achieved acceptable levels of prediction for the K (K-factor), Ki, and Kr indices, with R^2 values ranging from 0.49 to 0.61 and ratio of prediction to deviation (RPD) values above 1.4. The study highlighted the potential of NIR as a low-cost, rapid alternative to conventional methods for assessing soil erodibility, enabling estimated inputs for models such as the USLE and RUSLE. Adding soil erodibility to the array of properties that can be predicted in a single IRS measurement could greatly enhance the efficiency of soil degradation monitoring for both quality and frequency of data collection.

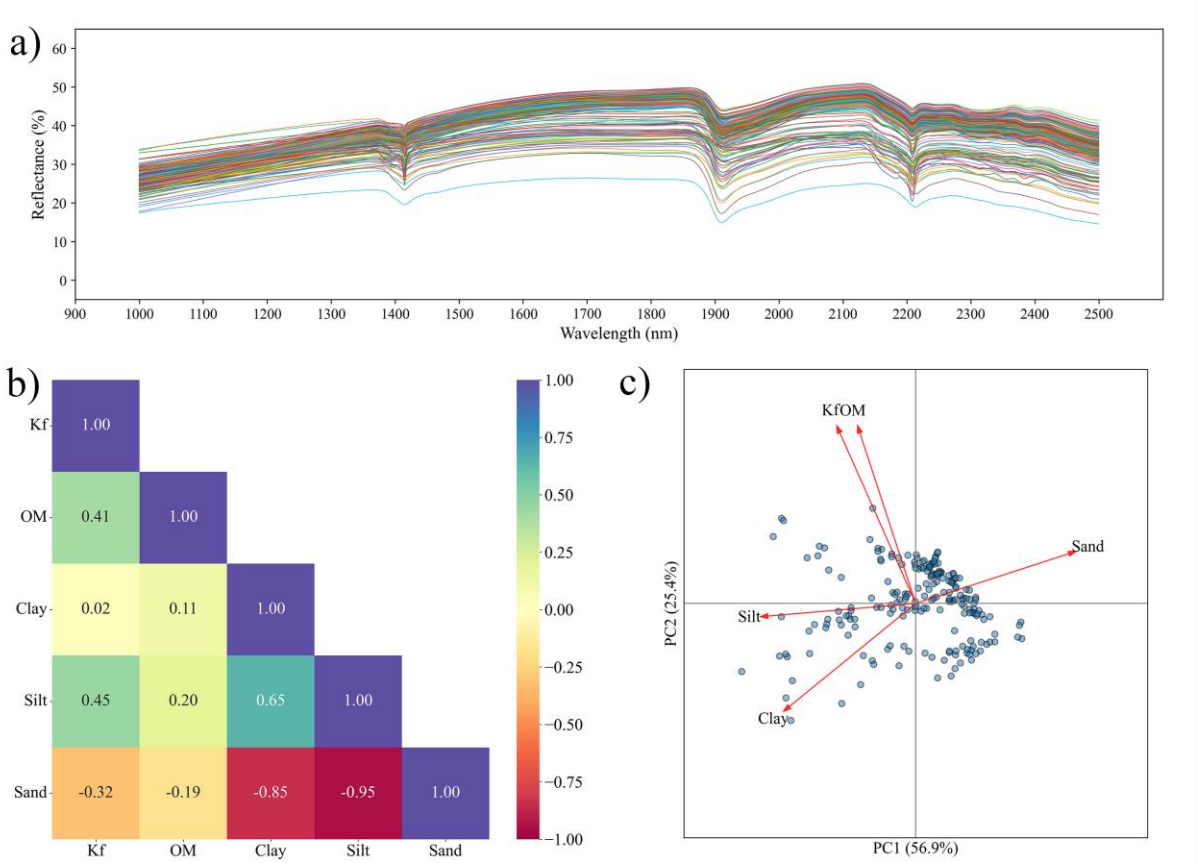
Given the current scarcity regarding efforts for the prediction of erosion-specific soil properties by means of infrared spectroscopy, a dataset was gathered (Figure 4) consisting of 203 samples from the states of Paraíba and Piauí. The data is comprised of absolute particle size distribution (clay, silt, and sand %), organic matter content (OM, g kg^{-1}), and the soil erodibility factor (Kf, 10^{-3} t h MJ^{-1} mm^{-1}). The erodibility factor (Kf) estimates the rate of soil loss per unit of rainfall erosion, and was obtained following the USLE model definition (OSTOVARI et al., 2016), which incorporates soil texture, permeability, aggregate distribution, and organic matter content.

In addition to the soil properties, near-infrared (NIR) reflectance spectra (1000 – 2500 nm) were measured for all samples (Figure 4a). This spectral data was obtained with an FT-IR/NIR spectrometer (Frontier/PerkinElmer) coupled with a Near-Infrared Reflectance Accessory (NIRA), using a 2-nm resolution and a 0.5-nm window, with multiple accumulations per sample to ensure representativeness. Spearman correlation and principal component analyses were run with the soil properties to disclose relationship between the variables. The spectral data was used to build prediction models for all available soil properties (Kf, OM, Clay, Silt, and Sand), focusing on whether it is possible to estimate the soil erodibility factor (Kf) from spectral data. In a similar approach as the previous topic, it employed both Convolutional Neural Networks (CNN) and Partial Least Squares regression (PLSR) models.

Following a similar approach as in the previous section, the dataset was split into a training subset (80%, n = 162) and a validation subset (20%, n = 41). Spectral preprocessing was performed using a Savitzky-Golay second-derivative filter to smooth the NIR spectra. These preprocessed spectra were exclusively used for the PLSR models. For each soil property, a PLSR model was optimized through a grid search over 2 to 10 components using 10-fold cross-validation; the best-performing model was then fitted to the training set and

evaluated with the validation set, producing the PLSR prediction performance metrics (Table 8). A Convolutional Neural Network (CNN) was also applied using the same methodology described in the previous section. The trained CNN models were evaluated on the validation set, resulting in the CNN prediction performance metrics.

Figure 4 - Overview of the erodibility sample set: a) Near-infrared spectra; b) Spearman correlation matrix between the soil properties; c) Principal component analysis biplot.



The correlation matrix and PCA biplot revealed some relationships between the erodibility factor and the other soil properties. Kf shows a moderate positive correlation with silt and organic matter, and a weak negative correlation with sand content, suggesting that finer textures and higher organic content increase soil erodibility within this dataset. Although this result may appear counterintuitive, given that organic matter and higher clay contents typically enhances soil aggregation and resistance to erosion, it likely reflects the specific nature of the studied soils, shaped by local parent materials, weathering conditions, and land use. These patterns are reinforced in the PCA biplot, where Kf and OM vectors align closely at the positive section of PC2, while the particle size distribution unsurprisingly demonstrates clay and silt opposing the sand vector mostly at the PC1.

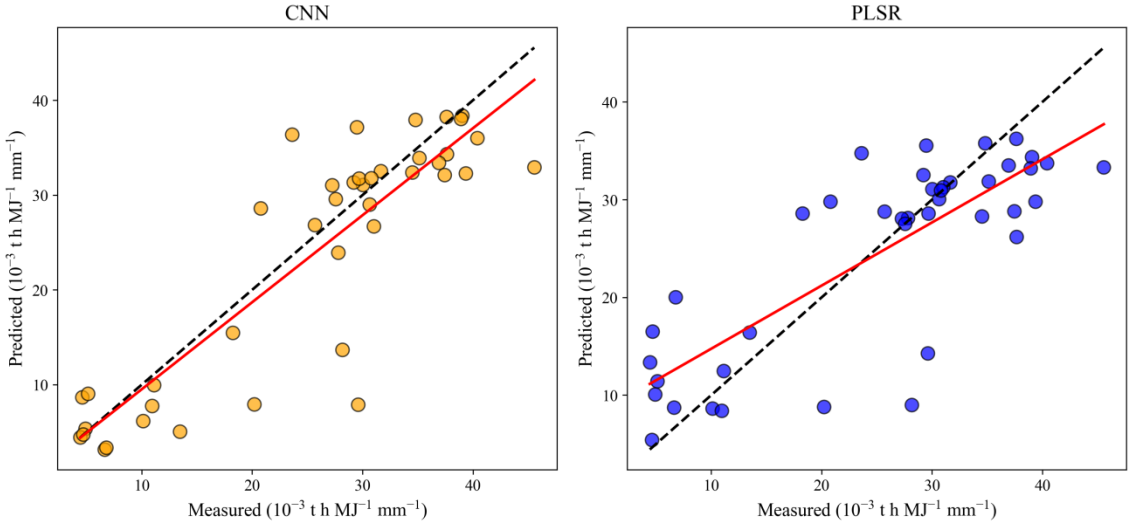
Table 8 - Prediction performance metrics with CNN and PLSR for all soil properties.

		Kf	OM	Clay	Silt	Sand
CNN	RMSE	6.33	0.33	5.34	9.07	13.95
	R^2	0.73	0.55	0.35	0.48	0.42
	RPD	1.99	1.55	1.24	1.39	1.31
PLSR	RMSE	7.32	0.38	4.94	9.88	14.70
	R^2	0.63	0.39	0.44	0.38	0.36
	RPD	1.66	1.28	1.34	1.27	1.25

CNN Convolutional neural network, PLSR Partial least squares regression, RMSE Root mean squared error, R^2 Coefficient of determination, RPD Ratio of prediction to deviation, Kf Erodibility factor, OM Organic matter

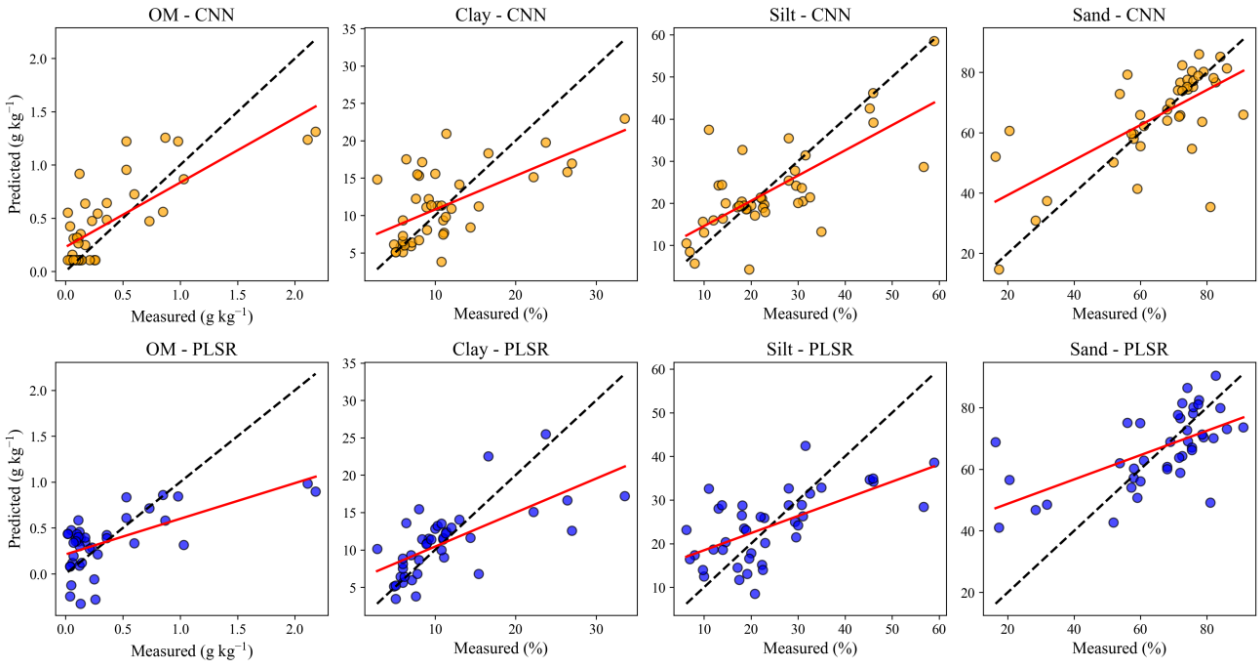
In this study, both PLSR and CNN models achieved promising results in predicting the soil erodibility factor (Kf) (Figure 5), with the CNN model outperforming PLSR across all evaluated properties except clay (Figure 6). Notably, the CNN model yielded an R^2 of 0.73 and an RPD of 1.99 for Kf (Table 8), surpassing the commonly accepted thresholds of $R^2 > 0.70$ and RPD > 1.4 for reliable predictions. These values suggest that the CNN model is capable of reproducing the Kf values with reasonable accuracy. The improved prediction of Kf using CNN aligns with the observed correlations between Kf, OM, and silt, reflecting the model's ability to capture subtle spectral signals associated with soil composition and structure, since organic matter and fine particles play a prominent role in determining erodibility.

Figure 5 - Predicted *versus* measured plots for K (soil erodibility factor).



The predictive performance for organic matter (OM) further confirms the effectiveness of CNN models in leveraging infrared spectral data to assess key soil quality indicators. While PLSR achieved a weak prediction for OM (R^2 = 0.39, RPD = 1.28), CNN improved these metrics to R^2 = 0.55 and RPD = 1.55, crossing the desirable RPD threshold. This enhancement suggests that CNNs are particularly adept at modeling the nonlinear relationships and complex interactions often found in spectral datasets. Given that OM can be an indicator of soil resilience to erosion, the reasonable performance of CNN models in predicting OM also reinforces their value in indirect erosion risk assessment. It is worth noting that OM and Kf were also positively associated in the correlation matrix and aligned in the PCA biplot, reflecting a common spectral signal that the CNN model has probably captured to predict both variables with reasonable accuracy.

Figure 6 - Predicted *versus* measured plots for organic matter (OM), clay, silt, and sand.



Interestingly, while the CNN model improved prediction for most soil properties, clay content remained an exception, with PLSR slightly outperforming CNN in this case (Table 8). This deviation may reflect the distinct spectral features of clay minerals, which are often well-captured by linear techniques like PLSR after proper spectral preprocessing such as the Savitzky-Golay filter. Overall, CNN consistently outperformed PLSR in predicting the most critical variables related to soil erosion (Kf and OM), highlighting its strategic advantage for integrated soil degradation assessments. These findings not only support the feasibility of

using IRS for erosion-related monitoring but also suggest that CNN models are particularly suited to enhancing predictive performance.

3.3.4 Salinity

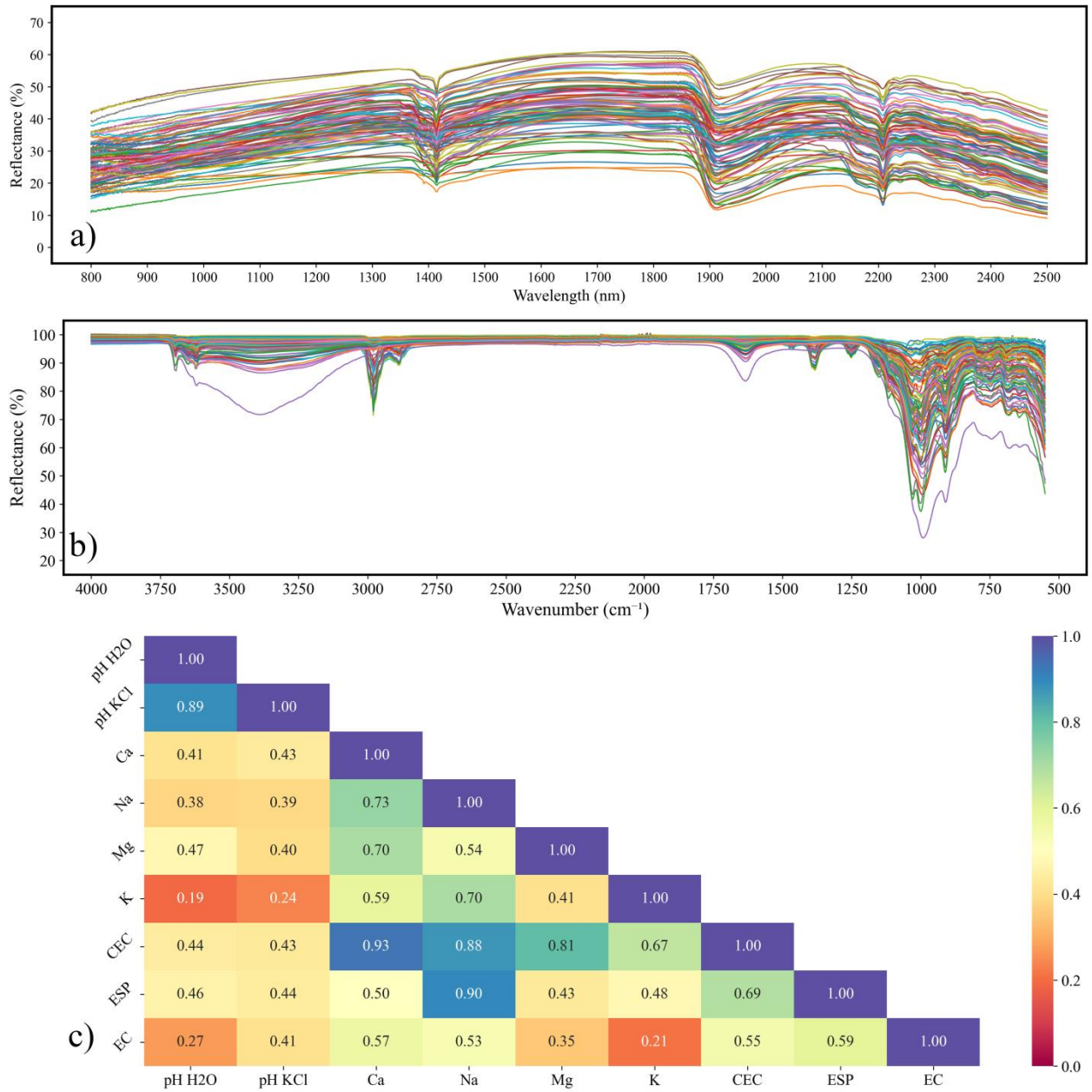
Salts are basic components of both soil and water. However, the accumulation of salts in the soil can lead to problematic outcomes such as decreasing agricultural productivity (BUTCHER et al., 2016), being considered a major soil degradation threat (DALIAKOPOULOS et al., 2016; SINGH, 2022). Climate change also plays a role in increasing soil salinity (CORWIN, 2021), particularly in arid and semiarid regions where ecosystems are especially fragile and vulnerable. Addressing this pressing challenge has inspired global efforts to map and monitor salt-affected soils (IVUSHKIN et al., 2019). Additionally, critical knowledge gaps and research priorities include the development of systematic approaches for analyzing salt-affected areas (HOPMANS et al., 2021). To advance these efforts, infrared spectroscopy emerges as a promising tool for cost-effective, non-destructive monitoring of soil salinity, enabling agile management strategies.

Infrared spectroscopy for soil salinity prediction is mostly centered on the measurement of electrical conductivity (EC) (CHEN et al., 2016), a critical parameter that represents the salinity status of soils. Statistical and machine learning methods, such as Partial Least Squares Regression (PLSR), Random Forest (RF), and other advanced algorithms, have been extensively utilized to model EC with IRS achieving favorable performances (WANG et al., 2018). Recent research highlights the potential of leveraging global datasets, such as the Open Soil Spectral Library, aiming to develop universally applicable models for soil salinity prediction (ZHOU et al., 2022), which could significantly enhance scalability and cross-regional applicability. Moreover, novel methodologies integrating near-infrared (NIR) and mid-infrared (MIR) spectroscopy are gaining traction (LOTFOLLAHI et al., 2023), offering improved spectral resolution and predictive accuracy. These approaches hold particular promise for distinguishing soil salinity variations across different land uses (WANG et al., 2023a), enabling site-specific management practices. Such advancements may further drive sustainable soil management in the face of expanding salinization pressures.

Building upon the global context and technological advances in soil salinity assessment via infrared spectroscopy, the following content presents data and findings derived from soil samples collected in areas with salinity-related challenges in the Brazilian Northeast. The dataset (Figure 7) consists of a total of 98 samples with analysis results for CEC, Ca, Na, Mg, K, ESP, EC (limited to 64 samples), and pH (both H_2O and KCl), as well

as spectral data at the NIR (800 – 2500 nm) and MIR (4000 – 550 cm^{-1}) regions. The core goal here was to assess two novel approaches for the estimation of soil salinity attributes: comparing NIR + MIR spectral fusion with separate NIR and MIR spectra, and; comparing the use of a Convolutional Neural Network (CNN) with the classical Partial Least Squares regression (PLSR) for these three different spectral datasets.

Figure 7 - Overview of the salinity sample set: a) Near-infrared spectra; b) Mid-infrared spectra; c) Spearman correlation matrix for the soil attributes.



NIR spectra were acquired using a Nicolet 26700 FTIR spectrometer (Waltham, MA, USA) operating in diffuse reflectance mode, equipped with an integrating sphere and an

InGaAs detector. Each spectrum consisted of 100 scans at a spectral resolution of 2 cm^{-1}. MIR spectra were obtained using a Nicolet 510-FTIR spectrometer (Thermo Electron Scientific, Madison, WI, USA), also in diffuse reflectance mode, with 100 scans per measurement at a spectral resolution of 2 cm^{-1}. During MIR data collection, a stream of dry, CO_2-free air was maintained through the spectrometer to prevent interference from moisture and carbon dioxide in the spectral readings. For the combination of NIR and MIR, the spectra of each sample were directly fused at the 2500 nm wavelength.

In a similar process as the previous topic, the dataset was divided into an 80% training (n = 78) subset and a 20% validation subset (n = 20). Spectral preprocessing was applied with a Savitzky-Golay second-derivative filter to smooth both NIR and MIR spectra. The preprocessed spectra were used only for the PLS models. For each soil attribute, a PLS regression model was tuned via grid-search over 2-10 number of components and 10-fold cross-validation; the optimal model was then calibrated with the training set and tested on the validation set, ultimately generating the prediction performance metrics for the PLSR (Table 9). A Convolutional Neural Network (CNN) was employed following the same procedures as the last topic. The trained networks were tested with the validation sample set, generating the prediction performance metrics for the CNN (Table 9).

Table 9 - Prediction performance metrics with CNN and PLSR using NIR, MIR, and the NIR+MIR fused spectra.

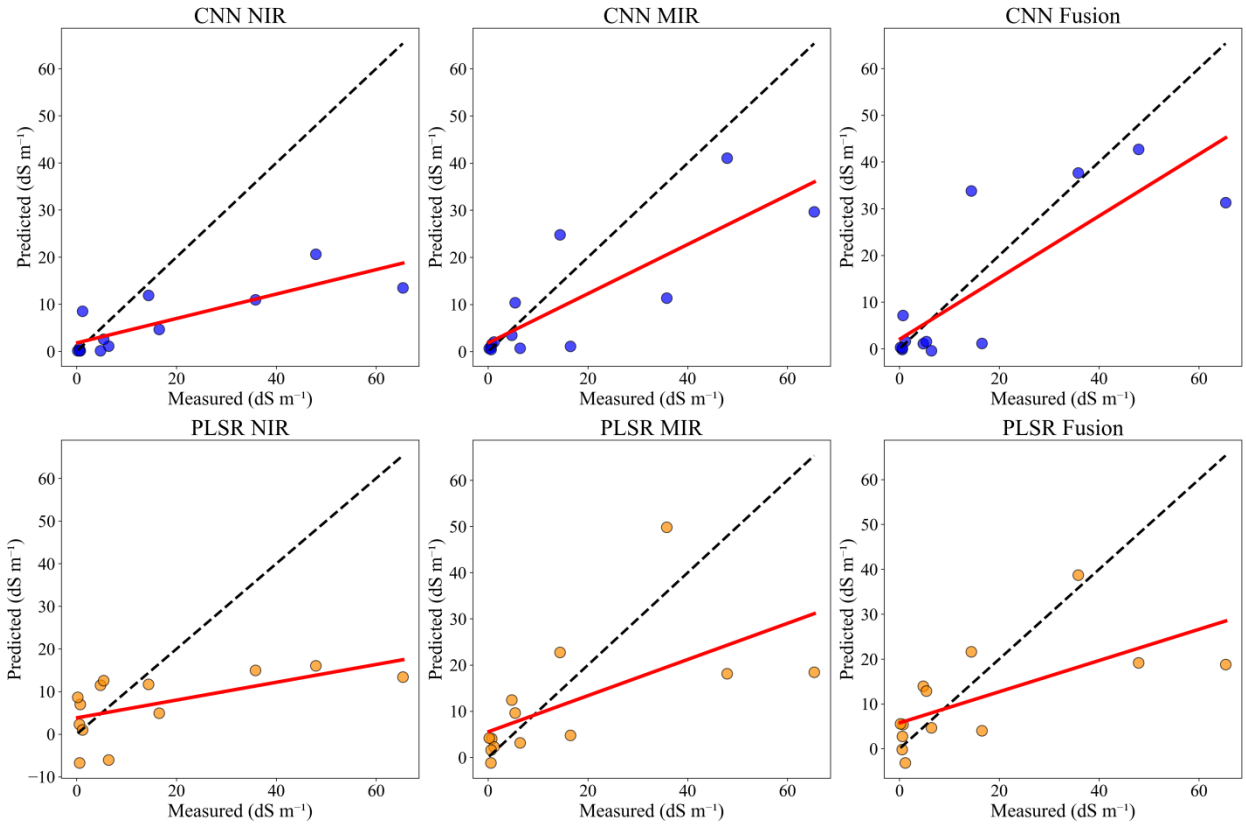
		NIR			MIR			NIR+MIR		
		RMSE	R^2	RPD	RMSE	R^2	RPD	RMSE	R^2	RPD
CNN	CEC	21.98	0.26	1.18	9.5	0.86	2.9	9.08	0.87	2.81
	Ca	7.39	0.26	1.19	4.11	0.77	2.16	5.28	0.62	1.68
	EC	18.21	0.19	1.31	13.37	0.57	1.67	12.15	0.64	1.73
	ESP	14.66	0.04	1.03	8.72	0.66	1.75	10.83	0.48	1.38
	K	1.05	0.2	1.12	0.7	0.65	1.77	0.92	0.39	1.28
	Mg	6.63	0.53	1.53	5.81	0.64	1.81	6.79	0.51	1.6
	Na	9.01	0.14	1.08	4.12	0.82	2.4	5.95	0.63	1.66
	pH H_2O	1	0.38	1.33	1.1	0.25	1.24	0.95	0.44	1.42
	pH KCl	1.15	0.09	1.07	0.96	0.36	1.41	0.98	0.33	1.27
PLSR	CEC	18.7	0.46	1.37	13.43	0.72	1.91	14.3	0.69	1.8
	Ca	7.04	0.33	1.22	5.32	0.62	1.62	5.52	0.59	1.56
	EC	19.03	0.12	1.19	16.66	0.33	1.25	16.24	0.36	1.29
	ESP	12.45	0.31	1.21	12.47	0.31	1.2	11.86	0.37	1.26
	K	1.04	0.22	1.13	0.89	0.43	1.34	0.92	0.39	1.29
	Mg	6.93	0.49	1.46	6.83	0.5	1.47	6.45	0.56	1.58

Na	8.3	0.27	1.18	6.98	0.48	1.39	6.56	0.54	1.48
pH H_2O	0.88	0.52	1.45	1.05	0.32	1.23	0.89	0.51	1.44
pH KCl	1.02	0.28	1.18	1.09	0.18	1.14	1	0.31	1.21

NIR Near-infrared; MIR Mid-infrared; RMSE Root mean squared error; R^2 Coefficient of determination; RPD Ratio of prediction to deviation; CNN Convolutional neural network; PLSR Partial least squares regression; CEC Cation exchange capacity; EC Electrical conductivity; ESP Exchangeable sodium percentage

Overall, CNN presented better results than PLS (Table 9), except for the NIR dataset. MIR results were much superior to NIR, indicating that the NIR region may have brought weak spectral information on the studied soil attributes, while the MIR region presented a more prominent relationship between the spectral features and the studied soil attributes. The spectral fusion presented slightly better results than MIR for some variables, a more visible pattern when using PLS (Table 9). In the context of salt-affected soils, the most important indicators are the electrical conductivity (EC) and exchangeable sodium percentage (ESP). Therefore, a comparison of the predicted vs. measured values at validation for each of these attributes using the three spectral datasets and the two prediction models can be visualized in Figure 8 for EC and Figure 9 for ESP. The predicted vs. measured plots for all studied soil attributes can be found in Appendix B.

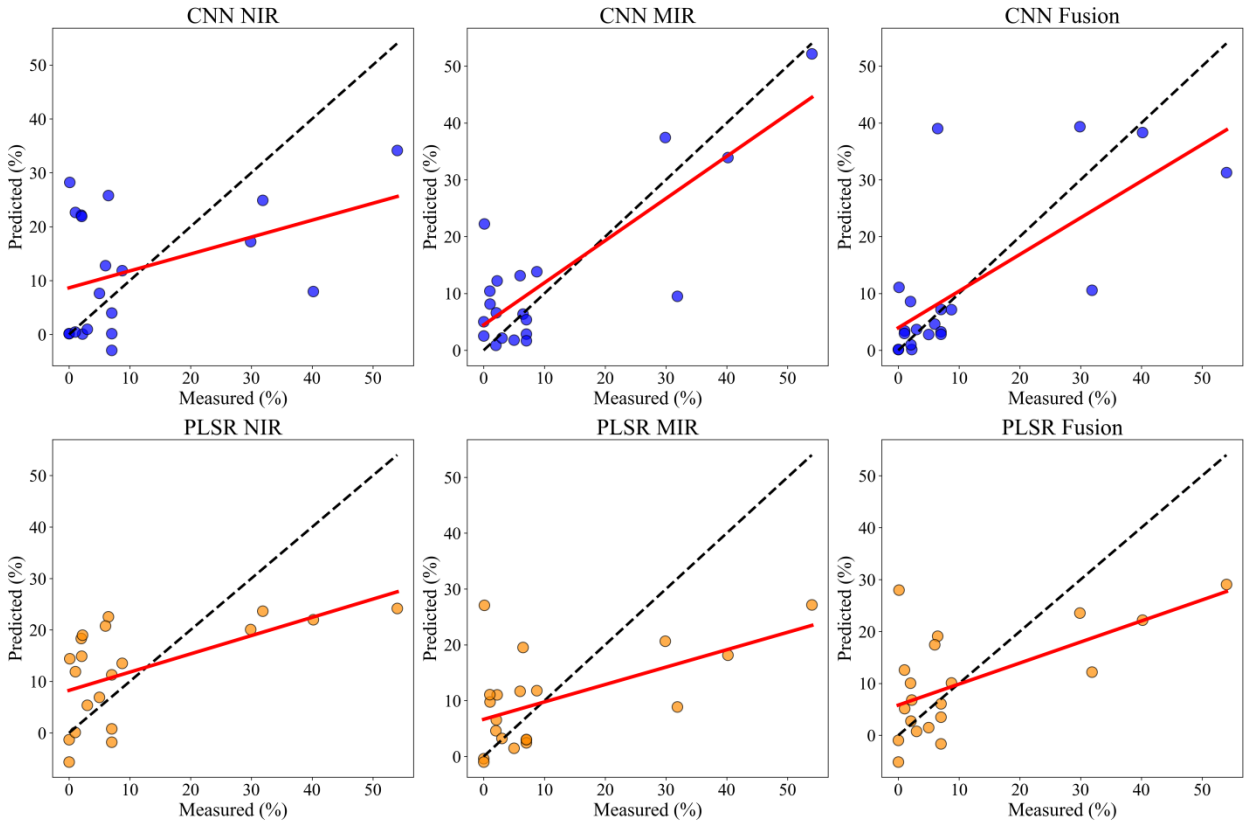
Figure 8 - Electrical conductivity (EC) predicted *versus* measured scatter plots for both Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR) using the three spectral datasets: near-infrared (NIR), mid-infrared (MIR), and the fused spectra of NIR and MIR (Fusion).



In the context of spectral fusion, Li et al. (2024b) have tried Vis-NIR combined with MIR spectra for the prediction of soil nutrients, reporting that the fusion approach achieved better results than all single sensors for their dataset. Furthermore, Ng et al. (2019) have also compared CNN with PLSR with Vis-NIR, MIR and their combined spectra, achieving high prediction accuracy for common soil properties (sand, clay, total C, organic C, CEC, and pH) from a global soil spectral dataset, especially when using CNN with the combined spectra. Meanwhile, Loftollahi et al. (2023) achieved better results for salinity-related soil attributes using MIR in comparison with NIR spectra, as was achieved here, although they did not test the prediction performance by fusing both spectral regions. For the prediction of soil classes, MIR has also been recently found to outperform both NIR and NIR+MIR fusion (Li et al., 2025), demonstrating that the quality of MIR soil spectral data is relevant for both regression and classification. On the other hand, soil organic carbon prediction has been improved by using different fusion methods for Vis-NIR and MIR spectra (Hong et al., 2022; Hong et al.,

2023). Such results underscore the fact that there is no single best method for the prediction of soil properties via infrared spectroscopy, especially in the context of spectral fusion.

Figure 9 - Exchangeable sodium percentage (ESP) predicted *versus* measured values for both convolutional neural network (CNN) and partial least squares regression (PLSR) using the three spectral datasets: near-infrared (NIR), mid-infrared (MIR), and the fused spectra of NIR and MIR (Fusion).



The modeling results reflect, in part, the underlying correlations among the soil properties. The moderate to high positive associations between EC, Ca, Na, Mg, K, CEC, and ESP observed in the correlation matrix (Figure 7c) suggest that these variables share common chemical and physical characteristics. Consequently, their spectra likely overlap, facilitating accurate prediction by multivariate techniques, such as PLSR, and deep learning models like CNN. Observing the performance results, the overlap may be stronger in the MIR region. This multicollinearity among target variables may also account for the comparable predictive performance observed for these attributes, while variables with weaker correlations (pH H_2O and pH KCl) exhibited poorer model accuracies.

The best prediction performance for EC was obtained with CNN using the fused spectra. However, the performance for EC using MIR alone was only slightly weaker than

with the fused spectra. In addition, some samples were lacking EC values, so the training and validation was undertaken with fewer samples (51 for training and 13 for validation) than the rest of the variables (78 for training and 20 for validation). Meanwhile, the best result for ESP was obtained with CNN with MIR alone, presenting a significant better result than with the fused spectra. Such results may indicate that using the fused spectra instead of MIR alone in the context of this study is not an advantageous trade-off. Thus, it can be stated that MIR alone can provide results with better reliability when using CNN for the prediction of the salinity-related soil attributes studied here. It is important to note that this statement is only valid for the limited regional context from where these samples were collected, and further studies exploring different environmental and pedological conditions are recommended.

These findings contribute to a broader understanding on the applicability of infrared spectroscopy as a tool for monitoring soil salinity. In a global context, by enabling the identification and quantification of key salinity-related soil properties, infrared spectroscopy can enhance the spatial and temporal resolution of salinity assessments. Soil salinity is a major driver of the desertification process (CASTRO et al., 2020; HARPER et al., 2021), which is itself one of the most important global environmental challenges today. Therefore, improving soil salinity monitoring through infrared spectroscopy constitutes an important advance for developing strategies to drive sustainable land use practices in arid and semi-arid regions, contributing to more effective desertification control efforts.

3.3.5 Desertification

Desertification is defined as the progressive degradation of arid, semiarid, and dry sub-humid lands into barren ecosystems, configuring a major threat to global agricultural productivity and socioeconomic stability (D'ODORICO et al., 2013; MARTÍNEZ-VALDERRAMA et al., 2020). Caused by synergistic interactions between climatic variability and anthropogenic pressures (BESTELMEYER et al., 2015; JIANG et al., 2019), this process accelerates soil erosion, diminishes biodiversity, and disrupts biogeochemical cycles which are critical for sustaining the affected ecosystems. To combat the spread of such detrimental phenomenon, monitoring has become a critical focus, especially leveraging remote sensing technologies. Imaging spectroscopy, for instance, enables the assessment of earth surface matter by analyzing visible and near-infrared properties, offering insights into soil-vegetation dynamics (PINET; KAUFMANN; HILL, 2006; LIU et al., 2014), which are essential for understanding desertification.

However, challenges such as the lack of reliable baselines, insufficient attention to temporal and spatial scales, and regional disparities in research efforts hinder a more accurate assessment (LIU et al., 2014; RIVERA-MARIN et al., 2022). On the other hand, in regions like China and Central Asia, remote sensing has proven essential for tracking desertification, revealing its drivers as a combination of climatic factors (e.g., prolonged drought, decreased precipitation) and human activities (e.g., agricultural abandonment, overgrazing, oil exploration) (ZHANG; HUISINGH, 2018; JIANG et al., 2019).

Despite the global surge in remote sensing studies in the last decade, the absence of standardized thresholds for key variables remains a significant limitation for desertification studies (RIVERA-MARIN et al., 2022). Addressing these challenges requires integrated innovative approaches to effectively monitor and mitigate desertification. Proximal sensing infrared spectroscopy is not suited for such task at macro scale due to its sampling-dependent spatial coverage. However, it can play a relevant role in the early detection of local desertification trends and in assessing specific, localized damages caused by the phenomenon. By providing a fast and cost-effective analysis of key soil properties, this technology can complement or be combined with remote sensing efforts, offering a more nuanced understanding of local scenarios and enabling targeted interventions to control the impacts of desertification.

3.4 Insights on the future of infrared spectroscopy for soil degradation monitoring

Infrared spectroscopy (IRS) has clearly emerged as a transformative tool in soil science, pushing the limits of what pedometrics can achieve. In the context of soil degradation, regarding its various processes, such as carbon depletion, erosion, pollution, salinity, and desertification, IRS does not directly assess the degradation phenomena themselves but has proven effective in rapidly, cost-effectively, and non-destructively estimating key soil attributes associated with each process. With advancements in chemometric modeling, especially the integration of machine learning and deep learning algorithms, the predictive performance of IRS continues to improve. This has led to increasingly accurate assessments not only in controlled laboratory settings but also in complex field conditions and diverse environments. The technology has also benefited from the creation and expansion of soil spectral libraries, enhancing the potential for regional and even global-scale model transferability.

However, challenges remain that require strategic attention. The standardization of data acquisition protocols, the representativeness of spectral libraries across different soil

types and environmental settings, and the availability of open-source prediction models are all critical to ensuring the reliability and comparability of IRS results. Furthermore, the effectiveness of IRS in predicting properties associated with pollution (especially less spectrally active PTEs such as trace metals), and its spatial limitations in monitoring large-scale desertification, are issues that demand further research. Innovative solutions such as spectral data fusion, remote sensing integration, and hybrid modeling architectures (e.g., CNNs trained on fused spectral and geospatial data) represent promising paths to address these constraints. Collaborative research efforts that incorporate local expertise, region-specific data, and open-access platforms will be essential in overcoming these barriers.

Looking ahead, the future of IRS in soil degradation monitoring appears increasingly synergistic. As spectral technology becomes more accessible and portable, and as machine learning tools become cheaper and more available, IRS is positioned to become a core approach in digital soil assessment frameworks. Its integration with satellite imagery and *in situ* field sensors may enable real-time, multi-scale monitoring, offering unprecedented resolution in both space and time. This convergence of technological innovation and environmental needs sets IRS not only as a novel scientific feature but as a practical tool in the global fight against soil degradation, supporting sustainable agriculture, informed land management, and effective environmental monitoring.

References

ABRANTES, G.; ALMEIDA, V.; MAIA, A. J.; NASCIMENTO, R.; NASCIMENTO, C.; SILVA, Y.; SILVA, Y.; VERAS, G. Comparison between Variable-Selection Algorithms in PLS Regression with Near-Infrared Spectroscopy to Predict Selected Metals in Soil. **Molecules**, v. 28, n. 19, p. 6959, 2023. https://doi.org/10.3390/molecules28196959

ADEJUMO, G. D.; BULMER, D.; SORENSON, P.; PEAK, D. Soil organic carbon and total nitrogen multivariate modelling from diverse FT-NIR spectral dataset. **Geoderma Regional**, v. 38, e00834, 2024. https://doi.org/10.1016/j.geodrs.2024.e00834

ADEJUMO, G.; BULMER, D.; SORENSON, P.; PEAK, D. Soil organic carbon measurements influence FT-NIR model training in calcareous soils of Saskatchewan. **Soil Science Society of America Journal**, v. 89, n. 2, e70034, 2025. https://doi.org/10.1002/saj2.70034

AHLRICHS, J. L.; RUSSELL, J. R.; HARTER, R. D.; WEISMILLER, R. A. Infrared spectroscopy of clay mineral systems. **Proceedings of the Indiana Academy of Science**, v. 75, p. 247–255, 1965.

ALBALADEJO, J.; ORTIZ, R.; GARCIA-FRANCO, N.; RUIZ NAVARRO, A.; ALMAGRO, M.; GARCIA PINTADO, J.; MARTÍNEZ-MENA, M. Land use and climate change impacts on soil organic carbon stocks in semi-arid Spain. **Journal of Soils and Sediments**, v. 13, p. 265–277, 2013. https://doi.org/10.1007/s11368-012-0617-7

AMEZKETA, E. An integrated methodology for assessing soil salinization, a pre-condition for land desertification. **Journal of Arid Environments**, v. 67, n. 4, p. 594–606, 2006. https://doi.org/10.1016/j.jaridenv.2006.03.010

ARCHUNDIA, D.; PRADO-PANO, B.; MOLINA-FREANER, F. Potentially toxic elements in soil–plant–water-animal continuum in a mining area from Northwestern Mexico: animal exposure pathways and health risks for children. **Environmental Geochemistry and Health**, v. 46, p. 99, 2024. https://doi.org/10.1007/s10653-024-01902-x

ARIAS-ESTÉVEZ, M.; LÓPEZ-PERIAGO, E.; MARTÍNEZ-CARBALLO, E.; SIMAL-GÁNDARA, J.; MEJUTO, J. C.; GARCÍA-RÍO, L. The mobility and degradation of pesticides in soils and the pollution of groundwater resources. **Agriculture, Ecosystems & Environment**, v. 123, n. 4, p. 247–260, 2008. https://doi.org/10.1016/j.agee.2007.07.011

ASKARI, M. S.; O'ROURKE, S. M.; HOLDEN, N. M. Evaluation of soil quality for agricultural production using visible–near-infrared spectroscopy. **Geoderma**, v. 243, p. 80–91, 2015. https://doi.org/10.1016/j.geoderma.2014.12.012

BAI, Z.; CHEN, S.; HONG, Y.; HU, B.; LUO, D.; PENG, J.; SHI, Z. Estimation of soil inorganic carbon with visible near-infrared spectroscopy coupling of variable selection and deep learning in arid region of China. **Geoderma**, v. 437, p. 116589, 2023. https://doi.org/10.1016/j.geoderma.2023.116589

BAO, Y.; YAO, F.; MENG, X.; ZHANG, J.; LIU, H.; MOUAZEN, A. M. Predicting soil organic carbon in cultivated land across geographical and spatial scales: Integrating Sentinel-2A and laboratory Vis-NIR spectra. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 203, p. 1–18, 2023. https://doi.org/10.1016/j.isprsjprs.2023.07.020

BARBETTI, R.; PALAZZI, F.; CHIARABAGLIO, P. M.; FONDON, C. L.; RIZZA, D.; ROCCI, A.; GRIGNANI, C.; ZAVATTARO, L.; MORETTI, B.; FANTAPPIÈ, M.; MONACO, S. Can Soil Organic Carbon in Long-Term Experiments be Detected Using Vis-NIR Spectroscopy? **IEEE Transactions on AgriFood Electronics**, 2024. https://doi.org/10.1109/TAFE.2024.3449215

BARTHÈS, B. G.; CHOTTE, J. L. Infrared spectroscopy approaches support soil organic carbon estimations to evaluate land degradation. **Land Degradation & Development**, v. 32, n. 1, p. 310–322, 2021. https://doi.org/10.1002/ldr.3718

BATISTA, P. V.; DAVIES, J.; SILVA, M. L.; QUINTON, J. N. On the evaluation of soil erosion models: Are we doing enough? **Earth-Science Reviews**, v. 197, p. 102898, 2019. https://doi.org/10.1016/j.earscirev.2019.102898

BAUMANN, K.; SCHÖNING, I.; SCHRUMPF, M.; ELLERBROCK, R. H.; LEINWEBER, P. Rapid assessment of soil organic matter: Soil color analysis and Fourier transform infrared spectroscopy. **Geoderma**, v. 278, p. 49–57, 2016. https://doi.org/10.1016/j.geoderma.2016.05.012

BORISOVER, M.; LADO, M.; LEVY, G. J. Modeling soil organic carbon content using mid-infrared absorbance spectra and a nonnegative MCR-ALS analysis. **Soil & Environmental Health**, v. 3, n. 1, p. 100123, 2025. https://doi.org/10.1016/j.seh.2024.100123

BORRELLI, P.; ALEWELL, C.; ALVAREZ, P.; ANACHE, J. A. A.; BAARTMAN, J.; BALLABIO, C.; BEZAK, N.; BIDDOCCU, M.; CERDÀ, A.; CHALISE, D.; CHEN, S.; CHEN, W.; DE GIROLAMO, A. M.; GESSESSE, G. D.; DEUMLICH, D.; DIODATO, N.; EFTHIMIOU, N.; ERPUL, G.; FIENER, P.; FREPPAZ, M.; PANAGOS, P. Soil erosion modelling: A global review and statistical analysis. **Science of the Total Environment**, v. 780, p. 146494, 2021. https://doi.org/10.1016/j.scitotenv.2021.146494

BUTCHER, K.; WICK, A. F.; DESUTTER, T.; CHATTERJEE, A.; HARMON, J. Soil salinity: A threat to global food security. **Agronomy Journal**, v. 108, n. 6, p. 2189–2200, 2016. https://doi.org/10.2134/agronj2016.06.0368

BESTELMEYER, B. T.; OKIN, G. S.; DUNIWAY, M. C.; ARCHER, S. R.; SAYRE, N. F.; WILLIAMSON, J. C.; HERRICK, J. E. Desertification, land use, and the transformation of global drylands. **Frontiers in Ecology and the Environment**, v. 13, n. 1, p. 28–36, 2015. https://doi.org/10.1890/140162

CACHADA, A.; PEREIRA, M. E.; FERREIRA DA SILVA, E.; DUARTE, A. C. Sources of potentially toxic elements and organic pollutants in an urban area subjected to an industrial impact. **Environmental Monitoring and Assessment**, v. 184, p. 15–32, 2012a. https://doi.org/10.1007/s10661-011-1943-8

CACHADA, A.; PATO, P.; ROCHA-SANTOS, T.; DA SILVA, E. F.; DUARTE, A. C. Levels, sources and potential human health risks of organic pollutants in urban soils. **Science of the Total Environment**, v. 430, p. 184–192, 2012b. https://doi.org/10.1016/j.scitotenv.2012.04.075

CAMBOU, A.; HOUSSOUKPÈVI, I. A.; CHEVALLIER, T.; MOULIN, P.; RAKOTONDRAZAFY, N. M.; FONKENG, E. E.; HARMAND, J.-M.; AHOLOUKPÈ, H. N. S.; AMADJI, G. L.; TABI, F. O.; CHAPUIS-LARDY, L.; BARTHÈS, B. G. Quantification of soil organic carbon in particle size fractions using a near-infrared spectral library in West Africa. **Geoderma**, v. 443, p. 116818, 2024. https://doi.org/10.1016/j.geoderma.2024.116818

CASTRO, F. C.; SANTOS, A. M. D. Salinity of the soil and the risk of desertification in the semiarid region. Mercator (Fortaleza), v. 19, e19002, 2020. https://doi.org/10.4215/rm2020.e19002

CHAKRAVARTULA, S. S. N.; MOSCETTI, R.; BEDINI, G.; NARDELLA, M.; MASSANTINI, R. Use of convolutional neural network, CNN combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee. **Food Control**, v. 135, p. 108816, 2022. https://doi.org/10.1016/j.foodcont.2022.108816

CHEN, H.; ZHAO, G.; SUN, L.; WANG, R.; LIU, Y. Prediction of soil salinity using near-infrared reflectance spectroscopy with nonnegative matrix factorization. **Applied Spectroscopy**, v. 70, n. 9, p. 1589–1597, 2016. https://doi.org/10.1177/0003702816662605

CLOETE, W. H.; DU PREEZ, G.; VAN ZIJL, G. M. Developing a near-infrared spectroscopy calibration algorithm for soil organic carbon content in South Africa. **Soil Advances**, v. 3, p. 100039, 2025. https://doi.org/10.1016/j.soilad.2025.100039

CONFORTI, M.; BUTTAFUOCO, G.; LEONE, A. P.; AUCELLI, P. P.; ROBUSTELLI, G.; SCARCIGLIA, F. Studying the relationship between water-induced soil erosion and soil organic matter using Vis–NIR spectroscopy and geomorphological analysis: A case study in southern Italy. **Catena**, v. 110, p. 44–58, 2013. https://doi.org/10.1016/j.catena.2013.06.013

CORWIN, D. L. Climate change impacts on soil salinity in agricultural areas. **European Journal of Soil** Science, v. 72, n. 2, p. 842–862, 2021. https://doi.org/10.1111/ejss.13010

COZZOLINO, D.; MORON, A. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. **The Journal of Agricultural Science**, v. 140, n. 1, p. 65–71, 2003. https://doi.org/10.1017/S0021859602002836

DAI, L.; XUE, J.; LU, R.; WANG, Z.; CHEN, Z.; YU, Q.; SHI, Z.; CHEN, S. In-situ prediction of soil organic carbon contents in wheat-rice rotation fields via visible near-infrared spectroscopy. **Soil & Environmental Health**, v. 2, n. 4, p. 100113, 2024. https://doi.org/10.1016/j.seh.2024.100113

DAI, L.; WANG, Z.; ZHUO, Z.; MA, Y.; SHI, Z.; CHEN, S. Prediction of soil organic carbon fractions in tropical cropland using a regional visible and near-infrared spectral library and machine learning. **Soil and Tillage Research**, v. 245, p. 106297, 2025. https://doi.org/10.1016/j.still.2024.106297

DALIAKOPOULOS, I. N.; TSANIS, I. K.; KOUTROULIS, A.; KOURGIALAS, N. N.; VAROUCHAKIS, A. E.; KARATZAS, G. P.; RITSEMA, C. J. The threat of soil salinity: A European scale review. **Science of the Total Environment**, v. 573, p. 727–739, 2016. https://doi.org/10.1016/j.scitotenv.2016.08.177

DAS, B.; CHAKRABORTY, D.; SINGH, V. K.; DAS, D.; SAHOO, R. N.; AGGARWAL, P.; MURGAOKAR, D.; MONDAL, B. P. Partial least square regression based machine learning models for soil organic carbon prediction using visible–near infrared spectroscopy. **Geoderma Regional**, v. 33, e00628, 2023. https://doi.org/10.1016/j.geodrs.2023.e00628

DEMATTÊ, J. A. M.; DOTTO, A. C.; PAIVA, A. F. S.; SATO, M. V.; DALMOLIN, R. S.; DE ARAÚJO, M. D. S. B.; DA SILVA, E. B.; NANNI, M. R.; TEN CATEN, A.; NORONHA, N. C.; LACERDA, M. P. C.; DE ARAÚJO FILHO, J. C.; RIZZO, R.; BELLINASO, H.; FRANCELINO, M. R.; SCHAEFER, C. E. G. R.; VICENTE, L. E.; DOS SANTOS, U. J.; SAMPAIO, E. V. DE S. B.; MENEZES, R. S. C.; DE SOUZA, J. J. L. L.; ABRAHÃO, W. A. P.; COELHO, R. M.; GREGO, C. R.; LANI, J. L.; FERNANDES, A. R.; GONÇALVES, D. A. M.; SILVA, S. H. G.; DE MENEZES, M. D.; CURI, N.; COUTO, E. G.; DOS ANJOS, L. H. C.; CEDDIA, M. B.; PINHEIRO, É. F. M.; GRUNWALD, S.; VASQUES, G. M.; MARQUES JÚNIOR, J.; DA SILVA, A. J.; DE VASCONCELOS BARRETO, M. C.; NÓBREGA, G. N.; DA SILVA, M. Z.; DE SOUZA, S. F.; VALLADARES, G. S.; VIANA, J. H. M.; TERRA, F. S.; HORÁK-TERRA, I.; FIORIO, P. R.; DA SILVA, R. C.; FRADE JÚNIOR, E. F.; LIMA, R. H. C.; FILIPPINI ALBA, J. M.; DE SOUZA JUNIOR, V. S.; BREFIN, M. DE L. M. S.; RUIVO, M. DE L. P.; FERREIRA, T. O.; BRAIT, M. A.; CAETANO, N. R.; BRINGHENTI, I.; MENDES, W. DE S.; SAFANELLI, J. L.; GUIMARÃES, C. C. B.; POPPIEL, R. R.; SOUZA, A. B.; QUESADA, C. A.; DO COUTO, H. T. Z. The Brazilian soil spectral library (BSSL): A general view, application and challenges. **Geoderma**, v. 354, p. 113793, 2019. https://doi.org/10.1016/j.geoderma.2019.05.043

DHARUMARAJAN, S.; GOMEZ, C.; KUSUMA, C. G.; VASUNDHARA, R.; KALAISSELVI, B.; LALITHA, M.; HEGDE, R. Prediction of soil organic carbon stock along layers and profiles using Vis-NIR laboratory spectroscopy. **Catena**, v. 257, p. 109150, 2025. https://doi.org/10.1016/j.catena.2025.109150

D'ODORICO, P.; BHATTACHAN, A.; DAVIS, K. F.; RAVI, S.; RUNYAN, C. W. Global desertification: Drivers and feedbacks. **Advances in Water Resources**, v. 51, p. 326–344, 2013. https://doi.org/10.1016/j.advwatres.2012.01.013

DORANTES, M. J.; FUENTES, B. A.; MILLER, D. M. Subsetting reduces the error of MIR spectroscopy models for soil organic carbon prediction in the US Great Plains. **Soil Security**, v. 16, p. 100145, 2024. https://doi.org/10.1016/j.soisec.2024.100145

DOUGLAS, R. K.; NAWAR, S.; ALAMAR, M. C.; MOUAZEN, A. M.; COULON, F. Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques. **Science of the Total Environment**, v. 616, p. 147–155, 2018. https://doi.org/10.1016/j.scitotenv.2017.10.323

DU, C.; ZHOU, J. Evaluation of soil fertility using infrared spectroscopy: A review. **Environmental Chemistry Letters**, v. 7, p. 97–113, 2009. https://doi.org/10.1007/s10311-008-0166-x

DUARTE, R. M.; MATOS, J. T.; SENESI, N. Organic pollutants in soils. In: Soil Pollution, p. 103–126, 2018. https://doi.org/10.1016/B978-0-12-849873-6.00005-4

ELLERBROCK, R. H.; GERKE, H. H.; DEUMLICH, D. Soil organic matter composition along a slope in an erosion-affected arable landscape in North East Germany. **Soil and Tillage Research**, v. 156, p. 209–218, 2016. https://doi.org/10.1016/j.still.2015.08.014

EOM, I. C.; RAST, C.; VEBER, A. M.; VASSEUR, P. Ecotoxicity of a polycyclic aromatic hydrocarbon (PAH)-contaminated soil. **Ecotoxicology and Environmental Safety**, v. 67, n. 2, p. 190–205, 2007. https://doi.org/10.1016/j.ecoenv.2006.12.020

FAGNANO, M.; AGRELLI, D.; PASCALE, A.; ADAMO, P.; FIORENTINO, N.; ROCCO, C.; PEPE, O.; VENTORINO, V. Copper accumulation in agricultural soils: Risks for the food chain and soil microbial populations. **Science of the Total Environment**, v. 734, p. 139434, 2020. https://doi.org/10.1016/j.scitotenv.2020.139434

FAN, X.; PEDROLI, B.; LIU, G.; LIU, Q.; LIU, H.; SHU, L. Soil salinity development in the Yellow River Delta in relation to groundwater dynamics. **Land Degradation & Development**, v. 23, n. 2, p. 175–189, 2012. https://doi.org/10.1002/ldr.1071

FARIA, O. C.; TORRES, G. N.; DI RAIMO, L. A.; COUTO, E. G. Estimate of carbon stock in the soil via diffuse reflectance spectroscopy (vis/nir) air and orbital remote sensing. **Revista Caatinga**, v. 36, p. 675–689, 2023. https://doi.org/10.1590/1983-21252023v36n320rc

FARMER, V. C. Infrared spectroscopy in clay mineral studies. **Clay Minerals**, v. 7, n. 4, p. 373–387, 1968. https://doi.org/10.1180/claymin.1968.007.4.01

FERNÁNDEZ-MARTÍNEZ, F.; RUBIANO-SANABRIA, Y. Near-Infrared Spectroscopy: Assessment of Soil Organic Carbon Stock in a Colombian Oxisol. **Ingeniería e Investigación**, v. 43, n. 3, p. 1, 2023. https://doi.org/10.15446/ing.investig.99102

FONSECA, A. A.; PASQUINI, C.; SOARES, E. M. Large-scale measurement of soil organic carbon using compact near-infrared spectrophotometers: Effect of soil sample preparation and the use of local modelling. **Environmental Science: Advances**, v. 2, n. 10, p. 1372–1381, 2023. https://doi.org/10.1039/D3VA00046J

FOSTER, E. J.; YEASMIN, S.; ZUNIGA, L.; BRECHEISEN, Z.; TOMASEK, A.; JIMENEZ CASTAÑEDA, M. E.; LEON-SALAS, W. D.; LOPA BOLÍVAR, J. A.; SCHULZE, D. G.; VILLALTA SOTO, M. J. C.; JOHNSTON, C. T.; FILLEY, T. Establishing Visible–Near Infrared Spectroscopy for the Prediction of Soil Organic Carbon in Arid Croplands of Arequipa, Peru. **Communications in Soil Science and Plant Analysis**, p. 1–22, 2025. https://doi.org/10.1080/00103624.2025.2485416

FREDERICKSON, L. D. Characterization of hydrated aluminas by infrared spectroscopy. **Analytical Chemistry**, v. 26, n. 12, p. 1883–1885, 1954.

GAUTAM, K.; SHARMA, P.; DWIVEDI, S.; SINGH, A.; GAUR, V. K.; VARJANI, S.; SRIVASTAVA, J. K.; PANDEY, A.; CHANG, J.-S.; NGO, H. H. A review on control and abatement of soil pollution by heavy metals: Emphasis on artificial intelligence in recovery of contaminated soil. **Environmental Research**, v. 225, p. 115592, 2023. https://doi.org/10.1016/j.envres.2023.115592

GEORGIOU, K.; JACKSON, R. B.; VINDUŠKOVÁ, O.; ABRAMOFF, R. Z.; AHLSTRÖM, A.; FENG, W.; HARDEN, J. W.; PELLEGRINI, A. F. A.; POLLEY, H. W.; SOONG, J. L.; RILEY, W. J.; TORN, M. S. Global stocks and capacity of mineral-associated soil organic carbon. **Nature Communications**, v. 13, n. 1, p. 3797, 2022. https://doi.org/10.1038/s41467-022-31540-9

GEYER, W.; BRÜGGEMANN, L.; HANSCHMANN, G. Prediction of properties of soil humic substances from FTIR spectra using partial least squares regression. International Journal of Environmental **Analytical Chemistry**, v. 71, n. 2, p. 181–193, 1998. https://doi.org/10.1080/03067319808032626

GHOLIZADEH, A.; BORŮVKA, L.; SABERIOON, M.; VAŠÁT, R. Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. **Applied Spectroscopy**, v. 67, n. 12, p. 1349–1362, 2013. https://doi.org/10.1366/13-07288

GHOLIZADEH, A.; SABERIOON, M.; BEN-DOR, E.; ROSSEL, R. A. V.; BORŮVKA, L. Modelling potentially toxic elements in forest soils with vis–NIR spectra and learning algorithms. **Environmental Pollution**, v. 267, p. 115574, 2020. https://doi.org/10.1016/j.envpol.2020.115574

GIANFREDA, L.; RAO, M. A.; PIOTROWSKA, A.; PALUMBO, G.; COLOMBO, C. Soil enzyme activities as affected by anthropogenic alterations: intensive agricultural practices and organic pollution. **Science of the Total Environment**, v. 341, n. 1-3, p. 265–279, 2005. https://doi.org/10.1016/j.scitotenv.2004.10.005

GOMEZ, C.; CHEVALLIER, T.; MOULIN, P.; BOUFERRA, I.; HMAIDI, K.; ARROUAYS, D.; JOLIVET, C.; BARTHÈS, B. G. Prediction of soil organic and inorganic carbon concentrations in Tunisian samples by mid-infrared reflectance spectroscopy using a French national library. **Geoderma**, v. 375, p. 114469, 2020. https://doi.org/10.1016/j.geoderma.2020.114469

GORE, R. C. Infrared spectroscopy. **Analytical Chemistry**, v. 30, n. 4, p. 570–579, 1958.

HADDAWAY, N. R.; HEDLUND, K.; JACKSON, L. E.; et al. How does tillage intensity affect soil organic carbon? A systematic review. **Environmental Evidence**, v. 6, p. 30, 2017. https://doi.org/10.1186/s13750-017-0108-9

HAJ-AMOR, Z.; ARAYA, T.; KIM, D. G.; BOURI, S.; LEE, J.; GHILOUFI, W.; YANG, Y.; KANG, H.; JHARIYA, M. K.; BANERJEE, A.; LAL, R. Soil salinity and its associated effects on soil microorganisms, greenhouse gas emissions, crop yield, biodiversity and desertification: A review. **Science of The Total Environment**, v. 843, p. 156946, 2022. https://doi.org/10.1016/j.scitotenv.2022.156946

HAN, L.; CHEN, R.; LIU, Z.; CHANG, S.; ZHAO, Y.; LI, L.; LI, R.; XIA, L. Sources of and control measures for PTE pollution in soil at the urban fringe in Weinan, China. **Land**, v. 10, n. 7, p. 762, 2021. https://doi.org/10.3390/land10070762

HARPER, R. J.; DELL, B.; RUPRECHT, J. K.; SOCHACKI, S. J.; SMETTEM, K. R. J. Salinity and the reclamation of salinized lands. In: STANTURF, J. A., & CALLAHAM, M. A. (Eds.). **Soils and landscape restoration**. Academic Press, p. 193–208, 2021. https://doi.org/10.1016/B978-0-12-813193-0.00007-2

HAU, N. X.; TUAN, N. T.; TRUNG, L. Q.; CHI, T. T. Estimation of soil organic carbon content using visible and near-infrared spectroscopy in the Red River Delta, Vietnam. **Chemometrics and Intelligent Laboratory Systems**, v. 255, p. 105253, 2024. https://doi.org/10.1016/j.chemolab.2024.105253

HE, Y.; HUANG, M.; GARCÍA, A.; HERNÁNDEZ, A.; SONG, H. Prediction of soil macronutrients content using near-infrared spectroscopy. **Computers and Electronics in Agriculture**, v. 58, n. 2, p. 144–153, 2007. https://doi.org/10.1016/j.compag.2007.03.011

HERRANZ-LUQUE, J. E.; GONZALEZ-CANALES, J.; MARTÍN-SANZ, J. P.; ANTÓN, O.; MORENO-DELAFUENTE, A.; NAVAS-VÁZQUEZ, M. J.; RAMOS-NIETO, R.; BIENES, R.; GARCÍA-DÍAZ, A.; MARQUES, M. J.; SASTRE, B. Monitoring Land Management Practices Using Vis–NIR Spectroscopy Provides Insights into Predicting Soil Organic Carbon and Limestone Levels in Agricultural Plots. **Agronomy**, v. 14, n. 6, p. 1150, 2024. https://doi.org/10.3390/agronomy14061150

HONG, Y.; MUNNAF, M. A.; GUERRERO, A.; CHEN, S.; LIU, Y.; SHI, Z.; MOUAZEN, A. M. Fusion of visible-to-near-infrared and mid-infrared spectroscopy to estimate soil organic carbon. **Soil and Tillage Research**, v. 217, p. 105284, 2022. https://doi.org/10.1016/j.still.2021.105284

HONG, Y.; CHEN, S.; HU, B.; WANG, N.; XUE, J.; ZHUO, Z.; YANG, Y.; CHEN, Y.; PENG, J.; LIU, Y.; MOUAZEN, A. M.; SHI, Z. Spectral fusion modeling for soil organic carbon by a parallel input-convolutional neural network. **Geoderma**, v. 437, p. 116584, 2023a. https://doi.org/10.1016/j.geoderma.2023.116584

HONG, Y.; CHEN, Y.; CHEN, S.; SHEN, R.; GUO, L.; LIU, Y.; MOUAZEN, A. M.; SHI, Z. Improving spectral estimation of soil inorganic carbon in urban and suburban areas by coupling continuous wavelet transform with geographical stratification. **Geoderma**, v. 430, p. 116284, 2023b. https://doi.org/10.1016/j.geoderma.2022.116284

HOPMANS, J. W.; QURESHI, A. S.; KISEKKA, I.; MUNNS, R.; GRATTAN, S. R.; RENGASAMY, P.; BEN-GAL, A.; ASSOULINE, S.; JAVAUX, M.; MINHAS, P. S.; RAATS, P. A. C.; SKAGGS, T. H.; WANG, G.; DE JONG VAN LIER, Q.; JIAO, H.; LAVADO, R. S.; LAZAROVITCH, N.; LI, B.; TALEISNIK, E. Critical knowledge gaps and research priorities in global soil salinity. **Advances in Agronomy**, v. 169, p. 1–191, 2021. https://doi.org/10.1016/bs.agron.2021.03.001

HORRIGAN, L.; LAWRENCE, R. S.; WALKER, P. How sustainable agriculture can address the environmental and human health harms of industrial agriculture. **Environmental Health Perspectives**, v. 110, n. 5, p. 445–456, 2002.

HUTENGS, C.; SEIDEL, M.; OERTEL, F.; LUDWIG, B.; VOHLAND, M. In situ and laboratory soil spectroscopy with portable visible-to-near-infrared and mid-infrared instruments for the assessment of organic carbon in soils. **Geoderma**, v. 355, p. 113900, 2019. https://doi.org/10.1016/j.geoderma.2019.113900

IVUSHKIN, K.; BARTHOLOMEUS, H.; BREGT, A. K.; PULATOV, A.; KEMPEN, B.; DE SOUSA, L. Global mapping of soil salinity change. **Remote Sensing of Environment**, v. 231, p. 111260, 2019. https://doi.org/10.1016/j.rse.2019.111260

JANIK, L. J.; SKJEMSTAD, J. O. Characterization and analysis of soils using mid-infrared partial least-squares. 2. Correlations with some laboratory data. **Soil Research**, v. 33, n. 4, p. 637–650, 1995. https://doi.org/10.1071/SR9950637

JANIK, L. J.; MERRY, R. H.; SKJEMSTAD, J. O. Can mid infrared diffuse reflectance analysis replace soil extractions?. **Australian Journal of Experimental Agriculture**, v. 38, n. 7, p. 681–696, 1998. https://doi.org/10.1071/EA97144

JIA, X.; XIE, M.; HU, B.; ZHOU, Y.; LI, H.; ZHAO, W.; DENG, W.; SHI, Z. Prediction of soil organic carbon contents in Tibet using a visible near-infrared spectral library. **Eurasian Soil Science**, v. 56, n. 6, p. 727–737, 2023. https://doi.org/10.1134/S1064229322601214

JIA, L.; YANG, F.; CHEN, Y.; PENG, L.; LENG, H.; ZU, W.; ZANG, Y.; GAO, L.; ZHAO, M. Prediction of wetland soil carbon storage based on near infrared hyperspectral imaging and deep learning. **Infrared Physics & Technology**, v. 139, p. 105287, 2024. https://doi.org/10.1016/j.infrared.2024.105287

JIANG, L.; JIAPAER, G.; BAO, A.; KURBAN, A.; GUO, H.; ZHENG, G.; DE MAEYER, P. Monitoring the long-term desertification process and assessing the relative roles of its drivers in Central Asia. **Ecological Indicators**, v. 104, p. 195–208, 2019. https://doi.org/10.1016/j.ecolind.2019.04.067

JIANG, F.; REN, B.; HURSTHOUSE, A.; et al. Evaluating health risk indicators for PTE exposure in the food chain: evidence from a thallium mine area. **Environmental Science and Pollution Research**, v. 27, p. 23686–23694, 2020a. https://doi.org/10.1007/s11356-020-08733-0

JIANG, Q.; CHEN, Y.; HU, J.; LIU, F. Use of visible and near-infrared reflectance spectroscopy models to determine soil erodibility factor (K) in an ecologically restored watershed. **Remote Sensing**, v. 12, n. 18, p. 3103, 2020b. https://doi.org/10.3390/rs12183103

JIANG, Y.; WEN, H.; ZHANG, Q.; et al. Source apportionment and health risk assessment of potentially toxic elements in soil from mining areas in northwestern China. **Environmental Geochemistry and Health**, v. 44, p. 1551–1566, 2022. https://doi.org/10.1007/s10653-021-00907-0

KARIMIAN, F.; AYOUBI, S.; KHALILI, B.; MIREEI, S. A.; AL-MULLA, Y. Using visible and near infrared spectroscopy and machine learning for estimating total petroleum hydrocarbons in contaminated soils. **Journal of Near Infrared Spectroscopy**, v. 32, n. 4–5, p. 152–162, 2024. https://doi.org/10.1177/09670335241269168

KÄSTNER, F.; SUT-LOHMANN, M.; RAMEZANY, S.; RAAB, T.; FEILHAUER, H.; CHABRILLAT, S. Estimating heavy metal concentrations in Technosols with reflectance spectroscopy. **Geoderma**, v. 406, p. 115512, 2022. https://doi.org/10.1016/j.geoderma.2021.115512

KHAN, S.; NAUSHAD, M.; LIMA, E. C.; ZHANG, S.; SHAHEEN, S. M.; RINKLEBE, J. Global soil pollution by toxic elements: Current status and future perspectives on the risk assessment and remediation strategies–A review. **Journal of Hazardous Materials**, v. 417, p. 126039, 2021. https://doi.org/10.1016/j.jhazmat.2021.126039

KHOSRAVI, V.; GHOLIZADEH, A.; KODEŠOVÁ, R.; AGYEMAN, P. C.; SABERIOON, M.; BORŮVKA, L. Visible, near-infrared, and shortwave-infrared spectra as an input variable for digital mapping of soil organic carbon. **International Soil and Water Conservation Research**, v. 13, n. 1, p. 203–214, 2025. https://doi.org/10.1016/j.iswcr.2024.10.002

KINGMA, D. P.; BA, J. Adam: A Method for Stochastic Optimization. **Proceedings of the International Conference on Learning Representations**, May 7–9, 2015, San Diego, CA, USA, 2015.

KISKU, G. C.; BARMAN, S. C.; BHARGAVA, S. K. Contamination of soil and plants with potentially toxic elements irrigated with mixed industrial effluent and its impact on the environment. **Water, Air, & Soil Pollution**, v. 120, p. 121–137, 2000. https://doi.org/10.1023/A:1005202304584

KOPITTKE, P. M.; MINASNY, B.; PENDALL, E.; RUMPEL, C.; MCKENNA, B. A. Healthy soil for healthy humans and a healthy planet. **Critical Reviews in Environmental Science and Technology**, v. 54, n. 3, p. 210–221, 2024. https://doi.org/10.1080/10643389.2023.2228651

LECUN, Y.; BENGIO, Y. Convolutional networks for images, speech, and time series. In: _____. **The handbook of brain theory and neural networks**, 1998, 978-0-262-51102-5.

LI, X.; FAN, P.; QIU, H.; LIU, Y. Optimizing soil carbon content prediction performance by multi-band feature fusion based on visible near-infrared spectroscopy. **Journal of Soils and Sediments**, v. 24, n. 3, p. 1333–1347, 2024a. https://doi.org/10.1007/s11368-024-03724-x

LI, X.; PAN, W.; LI, D.; GAO, W.; ZENG, R.; ZHENG, G.; CAI, K.; ZENG, Y.; JIANG, C. Can fusion of vis-NIR and MIR spectra at three levels improve the prediction accuracy of soil nutrients?. **Geoderma**, v. 441, p. 116754, 2024b. https://doi.org/10.1016/j.geoderma.2023.116754

LI, S.; SHEN, X.; SHEN, X.; CHENG, J.; XU, D.; MAKAR, R. S.; GUO, Y.; HU, B.; CHEN, S.; HONG, Y.; PENG, J.; SHI, Z. Improving the Accuracy of Soil Classification by Using Vis–NIR, MIR, and Their Spectra Fusion. **Remote Sensing**, v. 17, n. 9, p. 1524, 2025. https://doi.org/10.3390/rs17091524

LIU, S.; YANG, Z.; WANG, X.; et al. Effects of Cd and Pb pollution on soil enzymatic activities and soil microbiota. **Frontiers in Agriculture in China**, v. 1, p. 85–89, 2007. https://doi.org/10.1007/s11703-007-0016-9

LIU, S.; WANG, T.; KANG, W.; DAVID, M. Several challenges in monitoring and assessing desertification. **Environmental Earth Sciences**, v. 73, p. 7561–7570, 2015. https://doi.org/10.1007/s12665-014-3926-x

LIU, B.; GUO, B.; ZHUO, R.; DAI, F. Estimation of soil organic carbon content by Vis-NIR spectroscopy combining feature selection algorithm and local regression method. **Revista Brasileira de Ciência do Solo**, v. 47, p. e0230067, 2023a. https://doi.org/10.36783/18069657rbcs20230067

LIU, S.; CHEN, J.; GUO, L.; WANG, J.; ZHOU, Z.; LUO, J.; YANG, R. Prediction of soil organic carbon in soil profiles based on visible–near-infrared hyperspectral imaging spectroscopy. **Soil and Tillage Research**, v. 232, p. 105736, 2023b. https://doi.org/10.1016/j.still.2023.105736

LIU, B.; GUO, B.; ZHUO, R.; DAI, F. Estimation of soil organic carbon in LUCAS soil database using Vis-NIR spectroscopy based on hybrid kernel Gaussian process regression. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 321, p. 124687, 2024a. https://doi.org/10.1016/j.saa.2024.124687

LIU, J.; MUNNAF, M. A.; MOUAZEN, A. M. Micro-near-infrared (micro-NIR) sensor for predicting organic carbon and clay contents in agricultural soil. **Soil and Tillage Research**, v. 242, p. 106155, 2024b. https://doi.org/10.1016/j.still.2024.106155

LONG, M.; YUE, T.; XU, Z.; GUO, J.; LUO, J.; GUO, X.; ZHAO, X. Improved soil organic carbon prediction in a forest area by near-infrared spectroscopy: Spiking of a soil spectral library. **Forests**, v. 14, n. 1, p. 118, 2023. https://doi.org/10.3390/f14010118

LOTFOLLAHI, L.; DELAVAR, M. A.; BISWAS, A.; FATEHI, S.; SCHOLTEN, T. Spectral prediction of soil salinity and alkalinity indicators using visible, near-, and mid-infrared spectroscopy. **Journal of Environmental Management**, v. 345, p. 118854, 2023. https://doi.org/10.1016/j.jenvman.2023.118854

LUO, X.; REN, B.; HURSTHOUSE, A. S.; et al. Potentially toxic elements (PTEs) in crops, soil, and water near Xiangtan manganese mine, China: potential risk to health in the foodchain. **Environmental Geochemistry and Health**, v. 42, p. 1965–1976, 2020. https://doi.org/10.1007/s10653-019-00454-9

MAIA, A. J.; NASCIMENTO, R. C.; DA SILVA, Y. J. A. B.; DO NASCIMENTO, C. W. A.; DE SOUSA MENDES, W.; NETO, J. G. V.; ARAÚJO FILHO, J. C.; TIECHER, T.; DA SILVA, Y. J. A. B. Near-infrared spectroscopy for prediction of potentially toxic elements in soil and sediments from a semiarid and coastal humid tropical transitional river basin. **Microchemical Journal**, v. 179, p. 107544, 2022. https://doi.org/10.1016/j.microc.2022.107544

MALLEY, D. F.; YESMIN, L.; WRAY, D.; EDWARDS, S. Application of near-infrared spectroscopy in analysis of soil mineral nutrients. **Communications in Soil Science and Plant Analysis**, v. 30, n. 7-8, p. 999–1012, 1999. https://doi.org/10.1080/00103629909370263

MAMMADOV, E.; DENK, M.; MAMEDOV, A. I.; GLAESSER, C. Predicting soil properties for agricultural land in the Caucasus Mountains using mid-infrared spectroscopy. **Land**, v. 13, n. 2, p. 154, 2024. https://doi.org/10.3390/land13020154

MANZO, S.; DE NICOLA, F.; PICIONE, F. D. L.; MAISTO, G.; ALFANI, A. Assessment of the effects of soil PAH accumulation by a battery of ecotoxicological tests. **Chemosphere**, v. 71, n. 10, p. 1937–1944, 2008. https://doi.org/10.1016/j.chemosphere.2007.12.026

MAO, D.; WANG, Z.; WU, B.; ZENG, Y.; LUO, L.; ZHANG, B. Land degradation and restoration in the arid and semiarid zones of China: Quantified evidence and implications from satellites. **Land Degradation & Development**, v. 29, n. 11, p. 3841–3851, 2018. https://doi.org/10.1002/ldr.3135

MARTÍNEZ-VALDERRAMA, J.; GUIRADO, E.; MAESTRE, F. T. Desertifying deserts. **Nature Sustainability**, v. 3, n. 8, p. 572–575, 2020. https://doi.org/10.1038/s41893-020-0561-2

MCCARTY, G. W.; REEVES, J. B.; REEVES, V. B.; FOLLETT, R. F.; KIMBLE, J. M. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. **Soil Science Society of America Journal**, v. 66, n. 2, p. 640–646, 2002. https://doi.org/10.2136/sssaj2002.6400a

METZGER, K.; LIEBISCH, F.; HERRERA, J. M.; GUILLAUME, T.; WALDER, F.; BRAGAZZA, L. The use of visible and near-infrared spectroscopy for in-situ characterization of agricultural soil fertility: A proposition of best practice by comparing scanning positions and spectrometers. **Soil Use and Management**, v. 40, n. 1, p. e12952, 2024. https://doi.org/10.1111/sum.12952

MIAO, Y.; WANG, H.; HUANG, X.; LIU, K.; SUN, Q.; MENG, L.; XU, D. Soil organic carbon prediction based on Vis–NIR spectral classification data using GWPCA–FCM algorithm. **Sensors**, v. 24, n. 15, p. 4930, 2024. https://doi.org/10.3390/s24154930

MILOŠ, B.; BENSA, A. Organic carbon estimation in a regional soil Vis-NIR database supported by unsupervised learning and chemometrics techniques. **Soil Advances**, v. 2, p. 100013, 2024. https://doi.org/10.1016/j.soilad.2024.100013

MO, C. H.; CAI, Q. Y.; LI, Y. H.; ZENG, Q. Y. Occurrence of priority organic pollutants in the fertilizers, China. **Journal of Hazardous Materials**, v. 152, n. 3, p. 1208–1213, 2008. https://doi.org/10.1016/j.jhazmat.2007.07.105

MONTANARELLA, L.; PENNOCK, D. J.; MCKENZIE, N.; BADRAOUI, M.; CHUDE, V.; BAPTISTA, I.; MAMO, T.; YEMEFACK, M.; SINGH AULAKH, M.; YAGI, K.; YOUNG HONG, S.; VIJARNSORN, P.; ZHANG, G.-L.; ARROUAYS, D.; BLACK, H.; KRASILNIKOV, P.; SOBOCKÁ, J.; ALEGRE, J.; HENRIQUEZ, C. R.; DE LOURDES MENDONÇA-SANTOS, M.; TABOADA, M.; ESPINOSA-VICTORIA, D.; ALSHANKITI, A.; ALAVIPANAH, S. K.; ELSHEIKH, E. A. E. M.; HEMPEL, J.; CAMPS ARBESTAIN, M.; NACHTERGAELE, F.; VARGAS, R. World's soils are under threat. **Soil**, v. 2, p. 79–82, 2016. https://doi.org/10.5194/soil-2-79-2016

MUNAWAR, A. A.; YUNUS, Y.; SATRIYO, P. Calibration models database of near infrared spectroscopy to predict agricultural soil fertility properties. **Data in Brief**, v. 30, p. 105469, 2020. https://doi.org/10.1016/j.dib.2020.105469

NAWAR, S.; CIPULLO, S.; DOUGLAS, R. K.; COULON, F.; MOUAZEN, A. M. The applicability of spectroscopy methods for estimating potentially toxic elements in soils: state-of-the-art and future trends. **Applied Spectroscopy Reviews**, v. 55, n. 7, p. 525–557, 2020. https://doi.org/10.1080/05704928.2019.1608110

NG, W.; MALONE, B. P.; MINASNY, B. Rapid assessment of petroleum-contaminated soils with infrared spectroscopy. **Geoderma**, v. 289, p. 150–160, 2017. https://doi.org/10.1016/j.geoderma.2016.11.030

NG, W.; MINASNY, B.; MONTAZEROLGHAEM, M.; PADARIAN, J.; FERGUSON, R.; BAILEY, S.; MCBRATNEY, A. B. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. **Geoderma**, v. 352, p. 251–267, 2019. https://doi.org/10.1016/j.geoderma.2019.06.016

NG, W.; MINASNY, B.; MENDES, W. D. S.; DEMATTÊ, J. A. M. The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. **Soil**, v. 6, n. 2, p. 565–578, 2020. https://doi.org/10.5194/soil-6-565-2020

NG, W.; MINASNY, B.; JEON, S. H.; MCBRATNEY, A. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. **Soil Security**, v. 6, p. 100043, 2022. https://doi.org/10.1016/j.soisec.2022.100043

NI, L. S.; FANG, N. F.; SHI, Z. H.; TAN, W. F. Mid-infrared spectroscopy tracing of channel erosion in highly erosive catchments on the Chinese Loess Plateau. **Science of the Total Environment**, v. 687, p. 309–318, 2019. https://doi.org/10.1016/j.scitotenv.2019.06.116

NUNES, M. R.; VEUM, K. S.; PARKER, P. A.; HOLAN, S. H.; KARLEN, D. L.; AMSILI, J. P.; VAN ES, H. M.; WILLS, S. A.; SEYBOLD, C. A.; MOORMAN, T. B. The soil health assessment protocol and evaluation applied to soil organic carbon. **Soil Science Society of America Journal**, v. 85, n. 4, p. 1196–1213, 2021. https://doi.org/10.1002/saj2.20244

OLIVEIRA, K. M. D.; GONÇALVES, J. V. F.; FURLANETTO, R. H.; OLIVEIRA, C. A. D.; MENDONÇA, W. A.; HAUBERT, D. D. F. D. S.; CRUSIOL, L. G. T.; FALCIONI, R.; OLIVEIRA, R. B.; REIS, A. S.; ECKER, A. E. A.; NANNI, M. R. Predicting particle size and soil organic carbon of soil profiles using VIS-NIR-SWIR hyperspectral imaging and machine learning models. **Remote Sensing**, v. 16, n. 16, p. 2869, 2024. https://doi.org/10.3390/rs16162869

OSTOVARI, Y.; GHORBANI-DASHTAKI, S.; BAHRAMI, H. A.; NADERI, M.; DEMATTE, J. A. M.; KERRY, R. Modification of the USLE K factor for soil erodibility assessment on calcareous soils in Iran. **Geomorphology**, v. 273, p. 385–395, 2016. https://doi.org/10.1016/j.geomorph.2016.08.003

PADARIAN, J.; MINASNY, B.; MCBRATNEY, A. B. Using deep learning to predict soil properties from regional spectral data. **Geoderma Regional**, v. 16, p. e00198, 2019. https://doi.org/10.1016/j.geodrs.2018.e00198

PALMBORG, C.; NORDGREN, A. Partitioning the variation of microbial measurements in forest soils into heavy metal and substrate quality dependent parts by use of near infrared spectroscopy and multivariate statistics. **Soil Biology and Biochemistry**, v. 28, n. 6, p. 711–720, 1996. https://doi.org/10.1016/0038-0717(95)00173-5

PANDEY, A.; JAIN, K. A robust deep attention dense convolutional neural network for plant leaf disease identification and classification from smart phone captured real world images. **Ecological Informatics**, v. 70, p. 101725, 2022. https://doi.org/10.1016/j.ecoinf.2022.101725

PINET, P. C.; KAUFMANN, C.; HILL, J. Imaging spectroscopy of changing Earth's surface: a major step toward the quantitative monitoring of land degradation and desertification. **Comptes Rendus Geoscience**, v. 338, n. 14-15, p. 1042–1048, 2006. https://doi.org/10.1016/j.crte.2006.09.012

POESEN, J. Soil erosion in the Anthropocene: Research needs. **Earth Surface Processes and Landforms**, v. 43, n. 1, p. 64–84, 2018. https://doi.org/10.1002/esp.4250

PYO, J.; HONG, S. M.; KWON, Y. S.; KIM, M. S.; CHO, K. H. Estimation of heavy metals using deep neural network with visible and infrared spectroscopy of soil. **Science of the Total Environment**, v. 741, p. 140162, 2020. https://doi.org/10.1016/j.scitotenv.2020.140162

QI, M.; CHEN, S.; WEI, Y.; ZHOU, H.; ZHANG, S.; WANG, M.; ZHENG, J.; VISCARRA ROSSEL, R. A.; CHANG, J.; SHI, Z.; LUO, Z. Using visible-near infrared spectroscopy to estimate whole-profile soil organic carbon and its fractions. **Soil & Environmental Health**, v. 2, n. 3, p. 100100, 2024. https://doi.org/10.1016/j.seh.2024.100100

QUINTON, J.; GOVERS, G.; VAN OOST, K.; et al. The impact of agricultural soil erosion on biogeochemical cycling. **Nature Geoscience**, v. 3, p. 311–314, 2010. https://doi.org/10.1038/ngeo838

RAMIFEHIARIVO, N.; BARTHÈS, B. G.; CAMBOU, A.; CHAPUIS-LARDY, L.; CHEVALLIER, T.; ALBRECHT, A.; RAZAFIMBELO, T. Comparison of near and mid-infrared reflectance spectroscopy for the estimation of soil organic carbon fractions in Madagascar agricultural soils. **Geoderma Regional**, v. 33, p. e00638, 2023. https://doi.org/10.1016/j.geodrs.2023.e00638

RAWLS, W. J.; PACHEPSKY, Y. A.; RITCHIE, J. C.; SOBECKI, T. M.; BLOODWORTH, H. Effect of soil organic carbon on soil water retention. **Geoderma**, v. 116, n. 1-2, p. 61–76, 2003. https://doi.org/10.1016/S0016-7061(03)00094-6

REDDI, S. J.; KALE, S.; KUMAR, S. On the convergence of Adam and beyond. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS. Apr 30-May 3, Vancouver, CA, 2018.

REEVES, J. B.; MCCARTY, G. W.; MEISINGER, J. J. Near infrared reflectance spectroscopy for the analysis of agricultural soils. **Journal of Near Infrared Spectroscopy**, v. 7, n. 3, p. 179–193, 1999.
REEVES III, J. B. Near-versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: where are we and what needs to be done?. **Geoderma**, v. 158, n. 1-2, p. 3–14, 2010. https://doi.org/10.1016/j.geoderma.2009.04.005

RENGASAMY, P. Soil processes affecting crop production in salt-affected soils. **Functional Plant Biology**, v. 37, n. 7, p. 613–620, 2010. https://doi.org/10.1071/FP09249

REYES, J.; LIEß, M. On-the-Go Vis-NIR spectroscopy for field-scale spatial-temporal monitoring of soil organic carbon. **Agriculture**, v. 13, n. 8, p. 1611, 2023. https://doi.org/10.3390/agriculture13081611

RIEFOLO, C.; CASTRIGNANÒ, A.; COLOMBO, C.; CONFORTI, M.; RUGGIERI, S.; VITTI, C.; BUTTAFUOCO, G. Investigation of soil surface organic and inorganic carbon contents in a low-intensity farming system using laboratory visible and near-infrared spectroscopy. **Archives of Agronomy and Soil Science**, v. 66, n. 10, p. 1436–1448, 2020. https://doi.org/10.1080/03650340.2019.1674446

RIVERA-MARIN, D.; DASH, J.; OGUTU, B. The use of remote sensing for desertification studies: A review. **Journal of Arid Environments**, v. 206, p. 104829, 2022. https://doi.org/10.1016/j.jaridenv.2022.104829

SAFANELLI, J. L.; HENGL, T.; PARENTE, L. L.; MINARIK, R.; BLOOM, D. E.; TODD-BROWN, K.; GHOLIZADEH, A.; MENDES, W. S.; SANDERMAN, J. Open Soil Spectral Library (OSSL): Building reproducible soil calibration models through open development and community engagement. **bioRxiv**, 2023-12. https://doi.org/10.1101/2023.12.16.572011

SALEHI-VARNOUSFADERANI, B.; HONARBAKHSH, A.; TAHMOURES, M.; AKBARI, M. Soil erodibility prediction by Vis-NIR spectra and environmental covariates coupled with GIS, regression and PLSR in a watershed scale, Iran. **Geoderma Regional**, v. 28, p. e00470, 2022. https://doi.org/10.1016/j.geodrs.2021.e00470

SANTANA, F. B.; DE SOUZA, A. M.; POPPI, R. J. Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 191, p. 454–462, 2018. https://doi.org/10.1016/j.saa.2017.10.052

SANTASUP, N.; THEANJUMPOL, P.; SANTASUP, C.; KITTIWACHANA, S.; MAWAN, N.; PRANTONG, L.; KHONGDEE, N. Development of near-infrared spectroscopy (NIRS) for estimating organic matter, total carbon, and total nitrogen in agricultural soil. **MethodsX**, v. 13, p. 102798, 2024. https://doi.org/10.1016/j.mex.2024.102798

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, 1964. https://doi.org/10.1021/ac60214a047

SENEVIRATHNE, N. S. L.; AHAMED, T. Rapid Analysis of Soil Organic Carbon in Agricultural Lands: Potential of Integrated Image Processing and Infrared Spectroscopy. **AgriEngineering**, v. 6, n. 3, p. 3001–3015, 2024. https://doi.org/10.3390/agriengineering6030172

SEYBOLD, C. A.; FERGUSON, R.; WYSOCKI, D.; BAILEY, S.; ANDERSON, J.; NESTER, B.; THOMAS, P. Application of mid-infrared spectroscopy in soil survey. **Soil Science Society of America Journal**, v. 83, n. 6, p. 1746–1759, 2019. https://doi.org/10.2136/sssaj2019.06.0205

SHANG, H.; SHANG, L.; WU, J.; XU, Z.; ZHOU, S.; WANG, Z.; WANG, H.; YIN, J. NIR spectroscopy combined with 1D-convolutional neural network for breast cancerization analysis and diagnosis. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 287, p. 121990, 2023. https://doi.org/10.1016/j.saa.2022.121990

SHARIFIFAR, A.; MINASNY, B.; ARROUAYS, D.; BOULONNE, L.; CHEVALLIER, T.; VAN DEVENTER, P.; VAN ZIJL, G. Soil inorganic carbon, the other and equally important soil carbon pool: distribution, controlling factors, and the impact of climate change. **Advances in Agronomy**, v. 178, p. 165–231, 2023. https://doi.org/10.1016/bs.agron.2022.11.005

SHI, T.; CHEN, Y.; LIU, Y.; WU, G. Visible and near-infrared reflectance spectroscopy—An alternative for monitoring soil contamination by heavy metals. **Journal of Hazardous Materials**, v. 265, p. 166–176, 2014. https://doi.org/10.1016/j.jhazmat.2013.11.059

SHI, B.; YANG, X.; LIANG, T.; LIU, S.; YAN, X.; LI, J.; LIU, Z. Source apportionment of soil PTE in a northern industrial county using PMF model: Partitioning strategies and uncertainty analysis. **Environmental Research**, v. 252, p. 118855, 2024. https://doi.org/10.1016/j.envres.2024.118855

SINGH, A. Soil salinity: A global threat to sustainable development. **Soil Use and Management**, v. 38, n. 1, p. 39–67, 2022. https://doi.org/10.1111/sum.12772

SOMMER, R.; BOSSIO, D. Dynamics and climate change mitigation potential of soil organic carbon sequestration. **Journal of Environmental Management**, v. 144, p. 83–87, 2014. https://doi.org/10.1016/j.jenvman.2014.05.017

SONG ZHAOLIANG, S. Z.; LIU CONGQIANG, L. C.; MÜLLER, K.; YANG XIAOMIN, Y. X.; WU YUNTAO, W. Y.; WANG HAILONG, W. H. Silicon regulation of soil organic carbon stabilization and its potential to mitigate climate change. **Earth-Science Reviews**, 2018. https://doi.org/10.1016/j.earscirev.2018.06.020

SØRENSEN, L. K.; DALSGAARD, S. Determination of clay and other soil properties by near infrared spectroscopy. **Soil Science Society of America Journal**, v. 69, n. 1, p. 159–167, 2005. https://doi.org/10.2136/sssaj2005.0159

STENBERG, B.; ROSSEL, R. A. V.; MOUAZEN, A. M.; WETTERLIND, J. Visible and near infrared spectroscopy in soil science. **Advances in Agronomy**, v. 107, p. 163–215, 2010. https://doi.org/10.1016/S0065-2113(10)07005-7

STOCKMANN, U.; PADARIAN, J.; MCBRATNEY, A.; MINASNY, B.; DE BROGNIEZ, D.; MONTANARELLA, L.; HONG, S. Y.; RAWLINS, B. G.; FIELD, D. J. Global soil organic carbon assessment. **Global Food Security**, v. 6, p. 9–16, 2015. https://doi.org/10.1016/j.gfs.2015.07.001

STRONG, J. Experimental infrared spectroscopy. **Physics Today**, v. 4, n. 4, p. 14–21, 1951.

SUN, J.; PAN, L.; TSANG, D. C.; ZHAN, Y.; ZHU, L.; LI, X. Organic contamination and remediation in the agricultural soils of China: A critical review. **Science of the Total Environment**, v. 615, p. 724–740, 2018. https://doi.org/10.1016/j.scitotenv.2017.09.271

SUTHERLAND, G. B. New trends in infrared spectroscopy. **Soviet Physics Uspekhi**, v. 2, n. 5, p. 759, 1959.

TANG, K.; ZHAO, X.; QIN, M.; XU, Z.; SUN, H.; WU, Y. Using convolutional neural network combined with multi-scale channel attention module to predict soil properties from visible and near-infrared spectral data. **Microchemical Journal**, v. 207, p. 111815, 2024a. https://doi.org/10.1016/j.geoderma.2020.114616

TANG, K.; ZHAO, X.; XU, Z.; SUN, H. A stacking ensemble model for predicting soil organic carbon content based on visible and near-infrared spectroscopy. **Infrared Physics & Technology**, p. 105404, 2024b. https://doi.org/10.1016/j.infrared.2024.105404

TAO, J.; RAZA, S.; ZHAO, M.; CUI, J.; WANG, P.; SUI, Y.; ZHOU, J. Vulnerability and driving factors of soil inorganic carbon stocks in Chinese croplands. **Science of the Total Environment**, v. 825, p. 154087, 2022. https://doi.org/10.1016/j.scitotenv.2022.154087

TEIXEIRA, M. P. R.; DA SILVA, Y. J. A. B.; BARBOSA, R. S.; NASCIMENTO, R. C.; LOPES, J. W. B.; GALDINO DE SOUSA, J. C.; SINGH, V. P. Near-infrared spectroscopy as an alternative tool for predicting soil erodibility in a watershed under desertification. **Land Degradation & Development**, v. 35, n. 4, p. 1526–1540, 2024. https://doi.org/10.1002/ldr.500

TENG, H.; ROSSEL, R. A. V.; SHI, Z.; BEHRENS, T.; CHAPPELL, A.; BUI, E. Assimilating satellite imagery and visible–near infrared spectroscopy to model and map soil loss by water erosion in Australia. **Environmental Modelling & Software**, v. 77, p. 156–167, 2016. https://doi.org/10.1016/j.envsoft.2015.11.024

THAKUR, P. S.; CHATURVEDI, S.; KHANNA, P.; SHEOREY, T.; OJHA, A. Vision transformer meets convolutional neural network for plant disease classification. **Ecological Informatics**, v. 77, p. 102245, 2023. https://doi.org/10.1016/j.ecoinf.2023.102245

THOMAS, N. C. The early history of spectroscopy. **Journal of Chemical Education**, v. 68, n. 8, p. 631, 1991.

TIECHER, T.; CANER, L.; MINELLA, J. P. G.; BENDER, M. A.; DOS SANTOS, D. R. Tracing sediment sources in a subtropical rural catchment of southern Brazil by using geochemical tracers and near-infrared spectroscopy. **Soil and Tillage Research**, v. 155, p. 478–491, 2016. https://doi.org/10.1016/j.still.2015.03.001

TINTI, A.; TUGNOLI, V.; BONORA, S.; FRANCIOSO, O. Recent applications of vibrational mid-Infrared (IR) spectroscopy for studying soil components: a review. **Journal of Central European Agriculture**, v. 16, n. 1, p. 0–0, 2015. https://doi.org/10.5513/JCEA01/16.1.1535

TÜMSAVAŞ, Z.; TEKIN, Y.; ULUSOY, Y.; MOUAZEN, A. M. Prediction and mapping of soil clay and sand contents using visible and near-infrared spectroscopy. **Biosystems Engineering**, v. 177, p. 90–100, 2019. https://doi.org/10.1016/j.biosystemseng.2018.06.008

VILLARINO, S. H.; STUDDERT, G. A.; BALDASSINI, P.; CENDOYA, M. G.; CIUFFOLI, L.; MASTRÁNGELO, M.; PIÑEIRO, G. Deforestation impacts on soil organic carbon stocks in the Semiarid Chaco Region, Argentina. **Science of the Total Environment**, v. 575, p. 1056–1065, 2017. https://doi.org/10.1016/j.scitotenv.2016.09.175

VISCARRA ROSSEL, R.; WEBSTER, R. Predicting soil properties from the Australian soil visible–near infrared spectroscopic database. **European Journal of Soil Science**, v. 63, n. 6, p. 848–860, 2012. https://doi.org/10.1111/j.1365-2389.2012.01495.x

VISCARRA ROSSEL, R. A.; HICKS, W. S. Soil organic carbon and its fractions estimated by visible–near infrared transfer functions. **European Journal of Soil Science**, v. 66, n. 3, p. 438–450, 2015. https://doi.org/10.1111/ejss.12237

VISCARRA ROSSEL, R. A.; BEHRENS, T.; BEN-DOR, E.; BROWN, D. J.; DEMATTÊ, J. A. M.; SHEPHERD, K. D.; JI, W. A global spectral library to characterize the world's soil. **Earth-Science Reviews**, v. 155, p. 198–230, 2016. https://doi.org/10.1016/j.earscirev.2016.01.012

WADA, K.; GREENLAND, D. J. Selective dissolution and differential infrared spectroscopy for characterization of 'amorphous' constituents in soil clays. **Clay Minerals**, v. 8, n. 3, p. 241–254, 1970. https://doi.org/10.1180/claymin.1968.007.4.01

WAISER, T. H.; MORGAN, C. L.; BROWN, D. J.; HALLMARK, C. T. In situ characterization of soil clay content with visible near-infrared diffuse reflectance spectroscopy. **Soil Science Society of America Journal**, v. 71, n. 2, p. 389–396, 2007. https://doi.org/10.2136/sssaj2006.0211

WALDEN, L.; SERRANO, O.; SHEN, Z.; ZHANG, M.; LAVERY, P.; LUO, Z.; ROSSEL, R. A. V. Mid-infrared spectroscopy determines the provenance of coastal marine soils and their organic and inorganic carbon content. **Science of The Total Environment**, v. 949, p. 174871, 2024. https://doi.org/10.1016/j.scitotenv.2024.174871

WALDEN, L.; ROSSEL, R. A. V. Combining temperature ramp dry combustion and mid-infrared spectroscopy for enhanced soil organic carbon characterisation. **Geoderma**, v. 458, p. 117316, 2025. https://doi.org/10.1016/j.geoderma.2025.117316

WANG, G.; FANG, Q.; TENG, Y.; YU, J. Determination of the factors governing soil erodibility using hyperspectral visible and near-infrared reflectance spectroscopy. **International Journal of Applied Earth Observation and Geoinformation**, v. 53, p. 48–63, 2016. https://doi.org/10.1016/j.jag.2016.08.006

WANG, J.; DING, J.; ABULIMITI, A.; CAI, L. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (VIS–NIR) spectroscopy, Ebinur Lake Wetland, Northwest China. **PeerJ**, v. 6, e4703, 2018. https://doi.org/10.7717/peerj.4703

WANG, J.; LIU, T.; ZHANG, J.; YUAN, H.; ACQUAH, G. E. Spectral variable selection for estimation of soil organic carbon content using mid-infrared spectroscopy. **European Journal of Soil Science**, v. 73, n. 4, e13267, 2022. https://doi.org/10.1111/ejss.13267

WANG, Z.; MIAO, Z.; YU, X.; HE, F. Vis-NIR spectroscopy coupled with PLSR and multivariate regression models to predict soil salinity under different types of land use. **Infrared Physics & Technology**, v. 133, p. 104826, 2023a. https://doi.org/10.1016/j.infrared.2023.104826

WANG, Y.; YANG, S.; YAN, X.; YANG, C.; FENG, M.; XIAO, L.; WANG, C. Evaluation of data pre-processing and regression models for precise estimation of soil organic carbon using Vis–NIR spectroscopy. **Journal of Soils and Sediments**, v. 23, n. 2, p. 634–645, 2023b. https://doi.org/10.1007/s11368-022-03337-2

WANG, Y.; YIN, K.; HU, B.; HONG, Y.; CHEN, S.; LIU, J.; SHI, Z. Ensemble and transfer learning of soil inorganic carbon with visible near-infrared spectra. **Geoderma**, v. 456, p. 117257, 2025. https://doi.org/10.1016/j.geoderma.2025.117257

WUEPPER, D.; BORRELLI, P.; FINGER, R. Countries and the global rate of soil erosion. **Nature Sustainability**, v. 3, n. 1, p. 51–55, 2020. https://doi.org/10.1038/s41893-019-0438-4

XIONG, J.; LIN, C.; MA, R.; WU, Z.; CHEN, L. Tracing sediment sources in a plain river network area by using optimized experimental design and reflectance spectroscopy. **Water Research**, v. 250, p. 121041, 2024. https://doi.org/10.1016/j.watres.2023.121041

XU, X.; LIU, W.; KIELY, G. Modeling the change in soil organic carbon of grassland in response to climate change: effects of measured versus modelled carbon pools for initializing the Rothamsted Carbon model. **Agriculture, Ecosystems & Environment**, 2011. https://doi.org/10.1016/j.agee.2010.12.018

XU, H.; XU, D.; CHEN, S.; MA, W.; SHI, Z. Rapid determination of soil class based on visible-near infrared, mid-infrared spectroscopy and data fusion. **Remote Sensing**, v. 12, n. 9, p. 1512, 2020. https://doi.org/10.3390/rs12091512

YANG, M.; MOUAZEN, A.; ZHAO, X.; GUO, X. Assessment of a soil fertility index using visible and near-infrared spectroscopy in the rice paddy region of southern China. **European Journal of Soil Science**, v. 71, n. 4, p. 615–626, 2020a. https://doi.org/10.1111/ejss.12907

YANG, J.; WANG, X.; WANG, R.; WANG, H. Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using Vis–NIR spectroscopy. **Geoderma**, v. 380, p. 114616, 2020b. https://doi.org/10.1016/j.microc.2024.111815

YUAN, B.; CAO, H.; DU, P.; REN, J.; CHEN, J.; ZHANG, H.; LUO, H. Source-oriented probabilistic health risk assessment of soil potentially toxic elements in a typical mining city. **Journal of Hazardous Materials**, v. 443, p. 130222, 2023. https://doi.org/10.1016/j.jhazmat.2022.130222

ZAYANI, H.; FOUAD, Y.; MICHOT, D.; KASSOUK, Z.; BAGHDADI, N.; VAUDOUR, E.; WALTER, C. Using machine-learning algorithms to predict soil organic carbon content from combined remote sensing imagery and laboratory vis-NIR spectral datasets. **Remote Sensing**, v. 15, n. 17, p. 4264, 2023a. https://doi.org/10.3390/rs15174264

ZAYANI, H.; FOUAD, Y.; MICHOT, D.; KASSOUK, Z.; LILI-CHABAANE, Z.; WALTER, C. Detecting the temporal trend of cultivated soil organic carbon content using visible near infrared spectroscopy. **Journal of Near Infrared Spectroscopy**, v. 31, n. 5, p. 241–255, 2023b. https://doi.org/10.1177/09670335231193113

ZHANG, Z.; HUISINGH, D. Combating desertification in China: Monitoring, control, management and revegetation. **Journal of Cleaner Production**, v. 182, p. 765–775, 2018. https://doi.org/10.1016/j.jclepro.2018.01.233

ZHANG, T.; LI, Y.; WANG, M. Prediction of soil organic carbon and total nitrogen affected by mine using Vis–NIR spectroscopy coupled with machine learning algorithms in calcareous soils. **Scientific Reports**, v. 14, n. 1, p. 28014, 2024. https://doi.org/10.1038/s41598-024-73761-6

ZHAO, P.; FALLU, D. J.; PEARS, B. R.; ALLONSIUS, C.; LEMBRECHTS, J. J.; VAN DE VONDEL, S.; VAN OOST, K. Quantifying soil properties relevant to soil organic carbon biogeochemical cycles by infrared spectroscopy: The importance of compositional data analysis. **Soil and Tillage Research**, v. 231, p. 105718, 2023a. https://doi.org/10.1016/j.still.2023.105718

ZHAO, W.; WU, Z.; YIN, Z.; LI, D. Reducing moisture effects on soil organic carbon content estimation in vis-NIR spectra with a deep learning algorithm. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 16, p. 7733–7748, 2023b. https://doi.org/10.1109/JSTARS.2023.3287583

ZHOU, Y.; HARTEMINK, A. E.; SHI, Z.; LIANG, Z.; LU, Y. Land use and climate change effects on soil organic carbon in North and Northeast China. **Science of the Total Environment**, v. 647, p. 1230–1238, 2019. https://doi.org/10.1016/j.scitotenv.2018.08.016

ZHOU, Y.; CHEN, S.; HU, B.; JI, W.; LI, S.; HONG, Y.; SHI, Z. Global soil salinity prediction by open soil Vis-NIR spectral library. **Remote Sensing**, v. 14, n. 21, p. 5627, 2022. https://doi.org/10.3390/rs14215627

ZHU, J.; JIN, Y.; ZHU, W.; LEE, D. K. VIS-NIR spectroscopy and environmental factors coupled with PLSR models to predict soil organic carbon and nitrogen. **International Soil and Water Conservation Research**, v. 12, n. 4, p. 844–854, 2024. https://doi.org/10.1016/j.iswcr.2024.02.001

ZIMMERMANN, M.; LEIFELD, J.; FUHRER, J. Quantifying soil organic carbon fractions by infrared-spectroscopy. **Soil Biology and Biochemistry**, v. 39, n. 1, p. 224–231, 2007. https://doi.org/10.1016/j.soilbio.2006.07.010

ZORNOZA, R.; GUERRERO, C.; MATAIX-SOLERA, J.; SCOW, K. M.; ARCENEGUI, V.; MATAIX-BENEYTO, J. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. **Soil Biology and Biochemistry**, v. 40, n. 7, p. 1923–1930, 2008. https://doi.org/10.1016/j.soilbio.2008.04.003

4 PREDICTION OF METAL CONTENTS FROM TECHNOSOLS UNDER DIFFERENT ACCUMULATION TIMES USING NEAR AND MID-INFRARED SPECTROSCOPY

Abstract

Technosols, soils formed from anthropogenic substrates such as mining tailings, represent an opportunity for sustainable land management aligned with circular economy principles. However, the elevated concentrations of potentially toxic elements in these materials raise environmental concerns, underscoring the importance of efficient monitoring. Conventional laboratory analyses for metal quantification are accurate but time-consuming and cost-ineffective, motivating the exploration of near-infrared (NIR) and mid-infrared (MIR) spectroscopy as more efficient alternatives. Despite promising results in other soil types, the application of infrared spectroscopy for predicting metal contents in Technosols remains limited. This study assessed the potential of NIR, MIR, and NIR+MIR fusion spectra to predict total concentrations of Cd, Co, Cr, Cu, Ni, Pb, Sn, Sr, Ti, V, and Zn in Technosols derived from scheelite mining residues in Northeastern Brazil. Surface samples (0-30 cm) were collected from Technosols with 2, 10, and 40 years of accumulation. Spectral data were preprocessed using the Savitzky-Golay second-derivative and modeled using Partial Least Squares regression (PLSR) and Convolutional Neural Network (CNN). Performance was evaluated using root mean squared error (RMSE), coefficient of determination (R^2), and ratio of prediction to deviation (RPD). The results showed that MIR and fusion spectra outperformed NIR, and PLSR delivered more consistent predictions than CNN. The best results were found for Cu (R^2 = 0.69, RPD = 1.81; CNN-Fusion) and Sr (R^2 = 0.67, RPD = 1.76; PLSR-MIR). Reasonable prediction performances (R^2 > 0.50; RPD > 1.4) were also achieved for Co, Sn, and Zn, while Cr was close to this threshold. These elements exhibited strong geochemical associations with potential spectrally active components such as Fe_2O_3T, Al_2O_3, CaO, and MnO. These findings reinforce the promise of infrared spectroscopy as an efficient, cost-effective tool for environmental monitoring of Technosols, supporting their safe transformation into productive and sustainable substrates.

Keywords: Spectral fusion. Trace elements. Soil spectroscopy. Chronosequence.

PREDIÇÃO DOS TEORES DE METAIS EM TECNOSSOLOS SOB DIFERENTES TEMPOS DE ACUMULAÇÃO USANDO ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E MÉDIO

Resumo

Tecnossolos, formados a partir de substratos antrópicos como rejeitos de mineração, representam uma oportunidade para o manejo sustentável do solo alinhado aos princípios da economia circular. No entanto, as elevadas concentrações de elementos potencialmente tóxicos nesses materiais levantam preocupações ambientais, ressaltando a importância de um monitoramento eficiente. As análises laboratoriais convencionais para quantificação de metais são precisas, mas demandam tempo e apresentam um alto custo, o que motiva a adoção da espectroscopia no infravermelho próximo (NIR) e médio (MIR) como alternativas mais eficientes. Apesar de resultados promissores em outros tipos de solos, a aplicação da espectroscopia no infravermelho para predição de teores de metais em Tecnossolos ainda é limitada. Este estudo avaliou o potencial dos espectros NIR, MIR e da fusão NIR+MIR para prever as concentrações totais de Cd, Co, Cr, Cu, Ni, Pb, Sn, Sr, Ti, V e Zn em Technosolos derivados de resíduos de mineração de scheelita no Nordeste do Brasil. Amostras de superfície (0–30 cm) foram coletadas em áreas com 2, 10 e 40 anos de acúmulo. Os dados espectrais foram pré-processados com a derivada de segunda ordem de Savitzky-Golay e modelados por regressão de quadrados mínimos parciais (PLSR) e redes neurais convolucionais (CNN). O desempenho foi avaliado usando a raiz do erro quadrático médio (RMSE), coeficiente de determinação (R^2) e razão da predição para desvio (RPD). Os espectros MIR e fusionados superaram o NIR, e os modelos PLSR apresentaram previsões mais consistentes que os CNN. Os melhores resultados foram obtidos para Cu (R^2 = 0,69; RPD = 1,81; CNN-Fusion) e Sr (R^2 = 0,67; RPD = 1,76; PLSR-MIR). Desempenhos razoáveis (R^2 > 0,50; RPD > 1,4) também foram observados para Co, Sn e Zn, enquanto Cr ficou próximo desse limite. Esses elementos mostraram fortes associações geoquímicas com componentes potencialmente ativos espectralmente, como Fe_2O_3T, Al_2O_3, CaO e MnO. Os resultados reforçam o potencial da espectroscopia no infravermelho como ferramenta eficiente e de baixo custo para o monitoramento ambiental de Technosolos, apoiando sua conversão segura em substratos produtivos e sustentáveis.

Palavras-chave: Fusão espectral. Elementos traço. Espectroscopia do solo. Cronossequência.

4.1 Introduction

Technosols are a distinct Reference Soil Group within the World Reference Base for Soil Resources, defined by the presence of human-made materials (e.g. inputs of mining tailings) with at least 20% mass per mass proportion in the upper 100 cm of the soil profile (ECHEVARRIA; MOREL, 2015). They differ from natural soils because their properties depend primarily on the characteristics of these anthropogenic substrates rather than on long-term pedogenic processes (LEHMANN, 2006; SCHAD, 2018). Technosols may include construction debris, industrial by-products or engineered amendments, and they are classified by the nature and proportion of those materials in the soil profile (LEGUÉDOIS et al., 2016).

Economically, Technosols represent a sustainable path for valorizing urban and industrial wastes, aligning with circular economy principles by converting by-products into fertile substrates. Agronomic modeling confirms that mixtures of construction and demolition waste, when properly amended, can match or even improve the cation exchange capacity, pH range, and water-retention properties of conventional soils (ROKIA et al., 2014). *In situ* monitoring over ten years has showed that Technosols can store from 50 to 500 % more organic carbon than adjacent natural soils at similar depths and can continue to sequester carbon over time under permanent vegetation cover (REES et al., 2019).

Early pedogenic studies revealed that Technosols exhibit a fast evolution, including horizon formation, aggregation, decarbonation, and organic matter incorporation, which provides useful data for modeling soil formation processes (SÉRÉ et al., 2010; LEGUÉDOIS et al., 2016). When developed from construction and demolition waste and amended with organic compost or biofertilizers, they have been shown to enhance plant biomass and improve soil health indicators, making them effective for supporting vegetation and land restoration (AZEVEDO-LOPES et al., 2024). Moreover, integrating Technosols into urban green infrastructure may increase the provision of ecosystem services such as carbon sequestration, pollutant immobilization, and flood regulation, contributing to public health and sustainable development strategies like the European Green Deal (RODRÍGUEZ-ESPINOSA et al., 2021).

However, metal contamination in Technosols represents a relevant environmental concern due to the increased presence of these potentially toxic elements in anthropogenic substrates (NASCIMENTO et al., 2021). Column leaching experiments have shown that

mixtures of sewage sludges and industrial wastes can release elevated concentrations of Cu, Zn, Ni, Pb, Cd, and Cr, particularly under high pH conditions, highlighting the dependence of metal solubility on both waste composition and acidity (YAO et al., 2009). Studies regarding plants cultivated on Technosols from abandoned mine sites have revealed metal concentrations well above regional background limits, demonstrating significant phytoavailability and potential risks to food chains (WAHSHA et al., 2012). In addition, Arbestain et al. (2008) demonstrated that the extractability of Cu, Ni, Cr, and Zn in waste-derived Technosols is influenced by a combination of factors, including organic matter content, pH buffering capacity, and the specific characteristics of the waste materials. Meanwhile, Guerrón et al. (2021) highlighted that the composition and treatment of Technosols can be engineered to influence metal retention capacity, emphasizing the role of material design in managing contaminant dynamics. In the face of this complex scenario, the assessment of metal contents in Technosols arises as an important step to ensure its safety and utility.

Conventional laboratory analyses for soil metal quantification (e.g. optical emission spectrometry, mass spectrometry) are accurate but time-consuming and cost-ineffective, motivating the adoption of near-infrared (NIR) and mid-infrared (MIR) spectroscopy as faster and more cost-effective alternatives. In general soils, deep learning approaches applied to visible-NIR (Vis-NIR) spectra have achieved reliable predictions for As, Cu, and Pb (Pyo et al., 2020), while comprehensive reviews highlight both the strengths and pitfalls of NIR methods for soil contaminant monitoring (GHOLIZADEH et al., 2013; COZZOLINO, 2016). Regarding Technosols, Vis-NIR spectroscopy has successfully predicted organic matter and microbial biomass in volcanic-ash-derived substrates (DI IORIO et al., 2022), and hyperspectral-based reflectance studies coupled with PLSR and random forest models have achieved reasonable results in the quantification of Cr, Cu, Ni, and Zn contents in sewage-derived Technosols (KÄSTNER et al., 2022). However, scientific literature remains scarce on the application of infrared spectroscopy for the prediction of Technosol properties, especially for metal contents.

There is an open gap which underscores the need for focused research to develop infrared-based predictive models specifically tailored for Technosols, including those which are derived from mining tailings. Such models could enable a viable alternative for the assessment of metal concentrations, enhancing the environmental monitoring and land management of such areas. In this context, the present study evaluated the potential of NIR and MIR spectroscopy, combined with partial least squares regression (PLSR) and

convolutional neural networks (CNN), to predict total concentrations of Cd, Co, Cr, Cu, Ni, Pb, Sn, Sr, Ti, V, and Zn in Technosols formed from scheelite mining residues with different accumulation times in northeastern Brazil.

4.2 Materials and Methods

4.2.1 Study area and sampling

The research was carried out in the municipality of Currais Novos, located in the Seridó region of Rio Grande do Norte, within Brazil's semiarid northeast (BEZERRA JÚNIOR; SILVA, 2007). The study site is the Brejuí Mine (6°19'19" S, 36°32'52" W), a historically significant mining area situated 10 kilometers from the town center (Godeiro et al., 2010). Operating since the 1940s, Brejuí is among the largest scheelite ($CaWO_4$) mines in South America and has played a strategic role in Brazil's tungsten production (DANTAS, 2002). The region is characterized by a hot semi-arid climate (BSh), with average annual rainfall around 610 mm, mostly concentrated between February and April. Long dry periods, high evapotranspiration, and irregular precipitation patterns define local environmental conditions (SILVA et al., 2023). Vegetation is characterized by the hyperxerophilous *Caatinga* domain and shallow Neosols are predominant (SILVA et al., 2023).

Mining at Brejuí is conducted via an extensive underground network of tunnels that span about 65 kilometers and reach depths approaching 900 meters (GERAB, 2014). The ore is extracted, crushed, and concentrated through gravity-based separation techniques, leveraging the high density of scheelite in relation to other minerals (GODEIRO et al., 2010). Unlike many other ore processing methods, no chemical reagents are used, which, while environmentally favorable, makes the process more costly and less efficient. In fact, less than 1% of the extracted material is converted into commercial product, with the remaining 99.2% becoming waste (RAMOS FILHO, 2021).

These mining residues are deposited in engineered basins designed to promote sedimentation and facilitate future storage. Basins are excavated in deeper landscape areas, shaped with local materials, and compacted by heavy machinery (DUTRA, 2019). As one basin nears the storage limit, a new structure is constructed. Over time, the accumulation and transformation of these materials give rise to Technosols.

Three sites with different accumulation times were selected to comprise this study: namely 2, 10, and 40 years since deposition. Surface horizon samples (0-30 cm) were collected from these three Technosols areas. For each area, 40 composite samples were

obtained, each composed of five subsamples collected around a central point. Following collection, samples were air-dried, disaggregated, and passed through a 2 mm sieve.

4.2.2 Chemical analyses

Soil pH was measured in both water and 1 mol L^{-1} KCl solution at a 1:2.5 ratio (soil:solution). A saturation paste was prepared for the determination of pH and electrical conductivity, following the protocol described by Teixeira et al. (2017). The oxides of Si, Al, Ti, Fe, Mn, Cu, Zn, Ca, K, Na, Mg, P, and S were quantified using X-ray fluorescence spectrometry (Rigaku ZSX Primus II, equipped with a Rhodium tube and seven analyzing crystals). Prior to analysis, the samples were ground using a ring mill; approximately 10 g of each were then pressed under 30 tons of pressure into aluminum capsules (34 mm diameter, 4 mm thickness) to form pressed powder pellets. Loss on ignition was determined by heating the samples in a muffle furnace at 1000 °C for one hour, with gradual temperature ramping. Results were expressed as weight percentages. Analytical quality control was ensured through the use of the certified reference material SRM 2709 San Joaquin Soil (Baseline Trace Element Concentrations), certified by the National Institute of Standards and Technology (NIST, 2002).

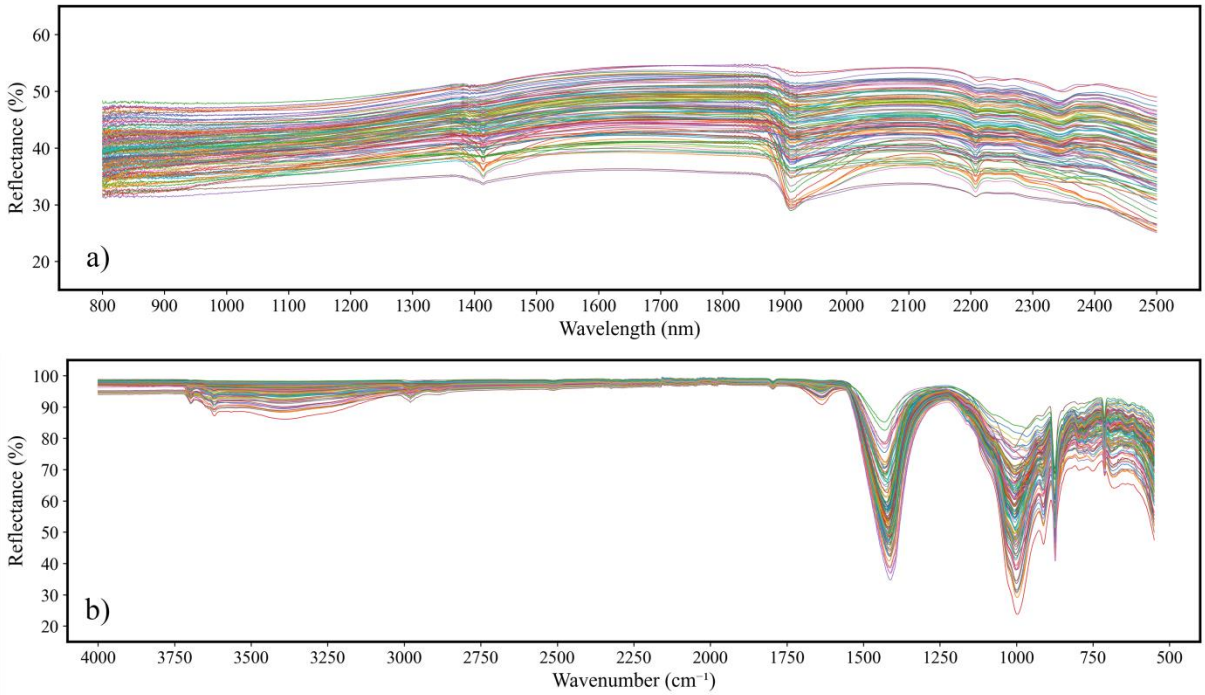
The total concentrations of Cd, Co, Cr, Cu, Ni, Pb, Sn, Sr, Ti, V, and Zn were determined in all samples. Metal extraction was carried out through open acid digestion, employing a combination of nitric, hydrochloric, perchloric, and hydrofluoric acids (ESTÉVEZ ÁLVAREZ et al., 2001). Initially, 1 gram of each sample was placed in open Teflon vessels and treated with 10 ml of HF for 12 hours. This was followed by the addition of 5 ml HNO_3 and 3 ml $HClO_4$, with heating at 180 °C. The digestion step was repeated to ensure complete dissolution of solid material. After digestion, the solutions were treated with 5 ml of HCl and diluted with deionized water in certified 25 ml volumetric flasks. Analytical quality control was maintained with the use of blanks, instrument calibration and recalibration, and high-purity reagents. Elemental concentrations were quantified using inductively coupled plasma optical emission spectroscopy (ICP-OES, Optima DV7000 PerkinElmer), equipped with a cyclonic spray chamber to improve measurement precision.

4.2.3 Infrared spectroscopy

Near-infrared (NIR, 800 – 2500 nm) spectra were collected using a Nicolet 26700 FTIR spectrometer (Waltham, MA, USA) configured for diffuse reflectance, incorporating an

integrating sphere and an InGaAs detector. Each spectral acquisition comprised 100 scans with a resolution of 2 cm^{-1}. Mid-infrared (MIR, 4000 – 550 cm^{-1}) spectra were collected on a Nicolet 510 FTIR instrument (Thermo Electron Scientific, Madison, WI, USA), also using diffuse reflectance mode, with the same number of scans and spectral resolution. To minimize interference from moisture and carbon dioxide during MIR measurements a stream of dry, CO_2-free air was maintained through the spectrometer. The NIR and MIR spectra for all samples can be visualized in Figure 10. In addition to the separate spectral regions, a fusion dataset was built by directly concatenating NIR and MIR spectra at the 2500 nm wavelength (equivalent to 4000 cm^{-1} wavenumber).

Figure 10 - Spectral data for all Technosol samples: a) Near-infrared spectra; b) Mid-infrared spectra.



4.2.4 Modeling and statistical evaluation

Firstly, the full sample set was partitioned into two subsets: 80% for training (n = 96) and 20% for validation (n = 24), stratifying the splits by the accumulation time sites (2-year, 10-year, 40-year) to ensure the representativeness of the three studied areas. All data-related procedures, including importing, partitioning, and processing, were carried out in Python (version 3.11.0) using the libraries pandas (v1.5.3) and scikit-learn (v1.2.2). Prior to modeling, spectral data were preprocessed with the Savitzky-Golay second-derivative, which was applied to both the NIR and MIR spectra, as well as to the fused NIR+MIR spectra. This

preprocessing aims to reduce high-frequency noise, correct baseline variations, and emphasize relevant spectral features (SAVITZKY; GOLAY, 1964). These preprocessed spectra were used exclusively in the partial least squares regression (PLSR) models. For each element, a PLSR model was optimized using a grid search over 2 to 10 components combined with 10-fold cross-validation. The PLSR model achieving the best results during cross-validation was then selected for calibration with the training subset and evaluation on the validation subset, producing the prediction performance metrics.

A Convolutional Neural Network (CNN) was also employed to explore deep learning for elemental prediction. CNNs are particularly effective for structured data arrays due to their use of convolutional filters that traverse the input array, performing point-wise calculations (PADARIAN et al., 2019). While individual convolutional layers are capable of detecting basic patterns, stacking layers enables the network to capture increasingly abstract and complex representations (LECUN; BENGIO, 1995). Although CNNs are typically applied to image-like data (PANDEY; JAIN, 2022; THAKUR et al., 2023), one-dimensional spectral data such as infrared reflectance can also be interpreted in this way (CHAKRAVARTULA et al., 2022; SHANG et al., 2023). Previous research has demonstrated the effectiveness of CNNs for predicting a variety of soil properties using infrared spectra (NG et al., 2019).

In this study, the CNN architecture (Appendix A) was designed using depthwise separable convolutions (YANG et al., 2020). This architecture efficiently captures patterns within the raw data. The model was trained using the Adam optimization algorithm (KINGMA; BA, 2015; REDDI; KALE; KUMAR, 2018), with the default learning rate of 0.001 and no weight decay. Training was conducted over 200 epochs with a batch size of 5. Once training was complete, the performance was assessed on the validation set, generating the corresponding prediction performance metrics. The implementation was carried out in Python using the keras API within the TensorFlow framework (version 2.12.0rc0).

Both PLSR and CNN performances were evaluated for the three spectral datasets (NIR, MIR and NIR+MIR fusion) using well-established statistical indicators commonly employed in spectroscopic modeling. These included the root mean squared error (RMSE), the coefficient of determination (R^2), and the ratio of prediction to deviation (RPD). To interpret the prediction quality, a classification scheme adapted from Wang et al. (2017) was used, relying on R^2 and RPD values. According to this framework, predictions were considered satisfactory when R^2 exceeded 0.70 and RPD was greater than 2.0; predictions were classified as reasonable when R^2 was above 0.50 with RPD values between 1.4 and 2.0, and; predictions were considered poor when R^2 was below 0.50 and RPD fell below 1.4.

In addition to the spectral modeling procedures, descriptive statistical analyses were performed for all measured soil properties to characterize their distribution and variability across the sampled Technosols. Key metrics such as mean, median, standard deviation, minimum, and maximum values were calculated to summarize the central tendency and dispersion of each variable. Furthermore, a Spearman correlation matrix was constructed and visualized as a heatmap to explore the relationships between the total metal concentrations and the other available soil properties. This correlation analysis provided an initial assessment of potential predictive associations to guide the interpretation of the modeling results. Furthermore, a Principal Component Analysis (PCA) was performed, considering the metal concentrations and the other available soil properties, to identify underlying patterns in the chemical composition of the samples. The PCA was used to reveal clusters and trends related to accumulation times and metal contents, also supporting the interpretation of the modeling results.

4.3 Results and discussion

The chemical composition of Technosols formed from scheelite mining residues from Brejuí Mine reflects both the nature of the deposited materials and the transformations they undergo over time. Differences in elemental concentrations, major oxides, and pH across the 2-, 10-, and 40-year accumulation sites provide insights into the geochemistry of these anthropogenic soils. Understanding these patterns is essential for assessing environmental risks. In addition to the conventional analyses, the prediction of Cd, Co, Cr, Cu, Ni, Pb, Sn, Sr, Ti, V, and Zn contents using infrared spectroscopy with PLSR and CNN is discussed. By comparing the performance of PLSR and CNN models across different spectral domains (NIR, MIR and NIR+MIR fusion), we assess the viability of rapid, non-destructive approaches for monitoring these key elements in scheelite mining-derived Technosols.

4.3.1 Geochemistry

The concentrations of metals in the studied Technosols (Table 10) are notably elevated compared to Earth's upper continental crust (UCC) average values (WEDEPOHL, 1995), reflecting the strong influence of mining residues accumulation. Elements such as cadmium (Cd), chromium (Cr), tin (Sn), strontium (Sr), and vanadium (V) exhibit a clear increasing trend over time, with Cd rising from 0.59 mg kg^{-1} at 2 years to 0.91 mg kg^{-1} at 40 years, nearly nine times higher than the UCC average (0.102 mg kg^{-1}). Cr increases from 136.38 to 149.93 mg kg^{-1}, more than four times the UCC value (35 mg kg^{-1}), while Sn and Sr similarly

grow to reach 27.06 mg kg^{-1} and 1082.28 mg kg^{-1} respectively, well above the UCC averages of 2.5 and 316 mg kg^{-1}. V also rises over time from 76.46 to 81.33 mg kg^{-1}, surpassing the UCC value of 53 mg kg^{-1} from early on.

In contrast, cobalt (Co), copper (Cu), nickel (Ni), lead (Pb), titanium (Ti), and zinc (Zn), show a decreasing trend over time (Table 10), despite initially high levels. Cu decreases from 509.43 mg kg^{-1} at 2 years to 146.81 mg kg^{-1} at 40 years, though it still remains 10 times higher than the UCC average of 14.3 mg kg^{-1} (WEDEPOHL, 1995). Zn follows a similar trajectory, dropping from 172.81 to 144.88 mg kg^{-1}, yet still nearly three times the UCC average (52 mg kg^{-1}). Co shows an initial concentration of 15.64 mg kg^{-1} at 2 years (slightly above the UCC average of 11.6 mg kg^{-1}) but decreases to 9.7 mg kg^{-1} at both 10 and 40 years, falling below background levels. Ni also declines slightly, from 32.50 mg kg^{-1} at 2 years to 25.99 mg kg^{-1} at 40 years, but remains above the UCC reference of 18.6 mg kg^{-1}, indicating moderate enrichment despite the downward trend. Pb, in contrast, remains consistently below the UCC average of 17 mg kg^{-1} across all sites, ranging from 11.50 to 11.96 mg kg^{-1}, which may suggest limited enrichment. Ti decreases from 2738.65 mg kg^{-1} to 2332.50 mg kg^{-1} over time, staying below the UCC average of 3117 mg kg^{-1}.

These trends suggest ongoing simultaneous elemental mobilization, retention, or depletion processes occurring within the Technosols over time. Overall, the persistent enrichment, particularly of Cd, Cr, Sn, and Sr, reinforces the environmental concerns and points to the importance of long-term monitoring of metal dynamics at these sites.

Table 10 - Descriptive statistics of total metal contents per accumulation time site.

	Site	Cd	Co	Cr	Cu	Ni	Pb	Sn	Sr	Ti	V	Zn
		mg kg $^{-1}$										
Mean	2y	0.59	15.64	136.38	509.43	32.50	11.96	18.84	767.94	2738.65	76.46	172.81
	10y	0.67	9.70	126.08	212.68	24.79	11.50	22.73	883.81	2332.50	76.71	143.07
	40y	0.91	9.73	149.93	146.81	25.99	11.59	27.06	1082.28	2489.54	81.33	144.88
Median	2y	0.50	14.25	128.50	433.00	32.00	11.50	17.25	716.00	2721.25	75.00	160.88
	10y	0.75	9.75	124.88	184.88	24.88	12.00	21.25	912.63	2323.50	78.13	142.00
	40y	0.88	9.75	144.13	144.88	26.13	10.75	25.88	1079.25	2525.00	77.38	136.50
SD	2y	0.28	6.42	44.31	367.51	9.43	3.36	7.72	204.36	523.12	14.59	42.52
	10y	0.17	2.28	30.27	98.07	4.85	3.84	6.62	203.08	326.09	13.17	22.87
	40y	0.36	1.94	32.52	50.00	5.65	3.17	8.38	165.48	509.78	15.72	31.51
Min.	2y	0.00	7.75	60.00	109.50	18.25	6.00	8.50	481.50	1635.50	43.25	93.25
	10y	0.50	5.75	69.25	98.00	16.75	4.50	14.25	418.25	1664.25	52.50	109.25
	40y	0.50	5.25	103.25	64.25	14.50	5.75	14.25	712.50	1269.00	44.50	88.75

Max.	2y	1.00	37.50	249.00	1703.50	64.75	19.50	44.00	1393.50	3687.50	123.50	285.25
	10y	1.00	14.50	199.50	542.50	34.75	23.00	47.25	1315.75	2937.50	101.50	196.25
	40y	2.00	14.25	236.25	260.75	39.00	17.75	46.25	1422.00	3352.50	112.75	201.25

SD Standard deviation; Min Minimum; Max Maximum; 2y 2-year; 10y 10-year; 40y 40-year

The geochemical composition of major oxides in the Technosols (Table 11) reveals clear trends over time, reflecting both the aging of the deposited materials and the influence of environmental conditions (HOUT et al., 2014; COUSSY et al., 2017). The consistent decline in average Al_2O_3 (from 10.66% to 6.37%) and SiO_2 (from 29.42% to 26.48%) suggests a gradual loss or transformation of aluminosilicates, probably due to pedogenetic processes. TiO_2 and Fe_2O_3T also decrease, although more modestly, indicating slower but consistent changes. In contrast, MnO increases progressively from 0.28% to 0.42%, which may reflect surface accumulation related to environmental exposure and oxidation, especially under semi-arid conditions with intermittent wet and dry periods. The reduction in CuO (from 0.07% to 0.02%) over time suggests active leaching and mirrors the decreasing pattern of its trace metal counterpart, while ZnO remains low and relatively unchanged across the sites.

Table 11 - Descriptive statistics of geochemistry-related major elements per accumulation time site.

	Site	Al_2O_3	SiO_2	TiO_2	MnO	Fe_2O_3T	CuO	ZnO
					%			
Mean	2y	10.66	29.42	0.48	0.28	6.88	0.07	0.02
	10y	8.14	27.39	0.42	0.37	5.72	0.03	0.02
	40y	6.37	26.48	0.39	0.42	5.61	0.02	0.02
Median	2y	10.51	29.53	0.48	0.26	6.75	0.06	0.02
	10y	7.83	26.20	0.40	0.36	5.62	0.03	0.02
	40y	6.40	26.29	0.40	0.43	5.67	0.02	0.02
SD	2y	1.98	3.87	0.07	0.07	1.33	0.05	0.01
	10y	2.15	5.07	0.07	0.06	0.90	0.02	0.00
	40y	1.02	2.50	0.08	0.10	0.62	0.01	0.01
Min.	2y	7.27	22.68	0.38	0.19	4.37	0.02	0.01
	10y	5.21	20.08	0.28	0.28	4.41	0.02	0.01
	40y	4.32	20.83	0.24	0.22	4.40	0.01	0.00
Max.	2y	14.15	43.00	0.62	0.50	9.79	0.24	0.04
	10y	15.07	41.18	0.65	0.55	8.62	0.08	0.03
	40y	8.05	31.75	0.53	0.59	7.64	0.04	0.03

SD Standard deviation; Min Minimum; Max Maximum; 2y 2-year; 10y 10-year; 40y 40-year

The fertility-related properties of the Technosols (Table 12) reveal patterns consistent with their origin from mine tailings rich in base-forming elements and their exposure to weathering over time. The soils show alkaline pH values throughout the chronosequence, with pH in H_2O rising from 8.10 to 8.53 and pH in KCl increasing from 7.49 to 8.07 over the 40-year period. This persistent alkalinity reflects the dominance of basic oxides, particularly CaO, which increases from 26.40% at 2 years to 37.42% at 40 years, likely due to the accumulation or persistence of calcium-rich material in the absence of significant leaching.

Magnesium oxide (MgO) levels remain stable (~3.1%), contributing to the base saturation and buffering capacity. Potassium (K_2O), while generally present at moderate levels, shows a slight decrease over time (from 1.39% to 1.14%), possibly due to depletion through leaching or plant uptake. Phosphorus (P_2O_5) increases slightly from 0.17% to 0.26%, which may be linked to low initial mobility. Sodium (Na_2O) slightly increases from 0.38% to 0.46% over time, while sulfur (SO_3) declines by half in 40 years.

Table 12 - Descriptive statistics of pH and fertility-related major elements per accumulation time site.

	Site	pH H_2O	pH KCl	P_2O_5	K_2O	CaO	MgO	SO_3	Na_2O
				%					
Mean	2y	8.10	7.49	0.17	1.39	26.40	3.25	0.24	0.38
	10y	8.30	7.71	0.24	1.22	34.23	3.05	0.30	0.37
	40y	8.53	8.07	0.26	1.14	37.42	3.11	0.12	0.46
Median	2y	8.05	7.49	0.13	1.37	26.67	3.18	0.21	0.36
	10y	8.40	7.75	0.23	1.08	35.54	3.06	0.16	0.34
	40y	8.60	8.10	0.25	1.23	38.07	3.14	0.11	0.45
SD	2y	0.17	0.20	0.09	0.24	6.00	0.42	0.11	0.15
	10y	0.28	0.21	0.08	0.40	7.08	0.26	0.33	0.13
	40y	0.20	0.18	0.07	0.31	4.60	0.38	0.03	0.13
Min.	2y	7.90	7.15	0.09	1.03	14.68	2.65	0.11	0.16
	10y	7.70	7.20	0.02	0.74	16.24	2.55	0.06	0.19
	40y	8.00	7.70	0.14	0.64	29.37	1.57	0.06	0.09
Max.	2y	8.70	7.83	0.47	2.49	37.21	4.33	0.71	1.02
	10y	8.70	8.10	0.49	2.21	45.06	3.82	1.57	0.78
	40y	8.90	8.50	0.41	1.63	45.94	3.78	0.19	0.78

SD Standard deviation; Min Minimum; Max Maximum; 2y 2-year; 10y 10-year; 40y 40-year

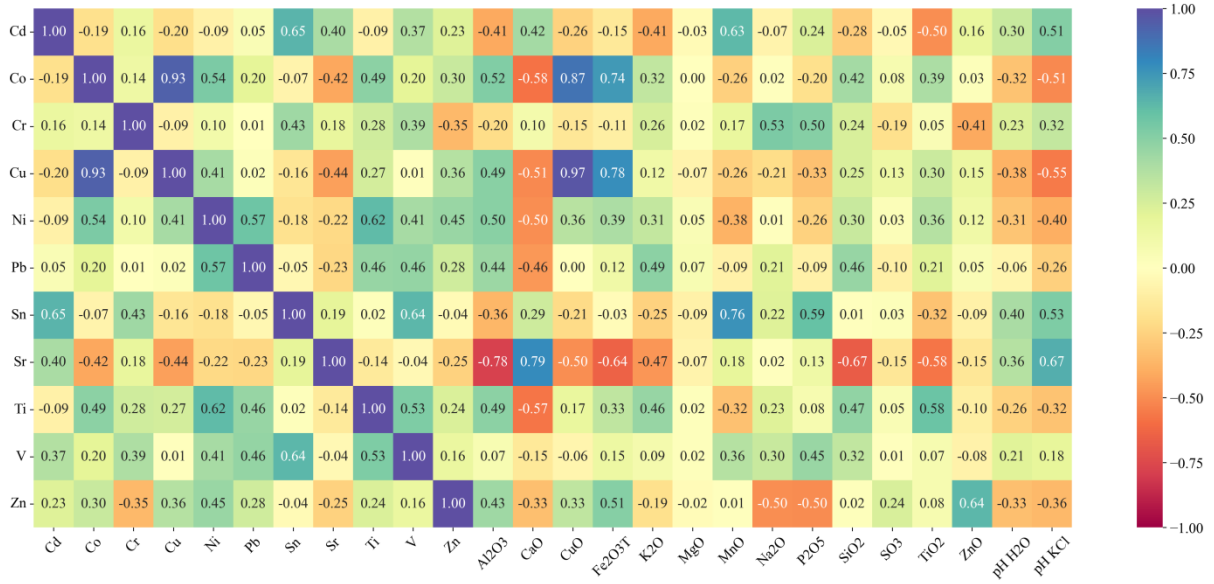
Overall, the data reflect a system with high base cation availability, especially Ca, and strong buffering capacity, but limited levels of plant-available nutrients such as P and K. In addition, the geochemical-related patterns indicate that the Technosols are undergoing

chemical changes over time, shaped by weathering, element mobility, and environmental exposure. Such transformations play a key role in defining the long-term behavior and environmental function of these soils in semi-arid landscapes.

The Spearman correlation matrix (Figure 11) reveals many associations among elemental concentrations and the other soil chemical attributes, indicating shared geochemical behavior and potential common sources. Cadmium (Cd) shows a strong positive correlation with Sn and MnO. Cobalt (Co) is strongly correlated with Cu, CuO, and Fe_2O_3T, implying a common geochemical pathway, possibly associated with iron and copper oxides (FRIEDRICH et al., 2012). Co also exhibits a negative correlation with CaO. Chromium (Cr) exhibits moderate positive correlations with P_2O_5 and Na_2O, but its overall correlations are generally weak, indicating a more dispersed behavior. Copper (Cu) has strong correlations with Co, CuO, and Fe_2O_3T, and moderate negative correlations with pH KCl and CaO. Nickel (Ni) is positively correlated with Co, Pb, Ti, and Al_2O_3, indicating an affinity for titanium-rich minerals.

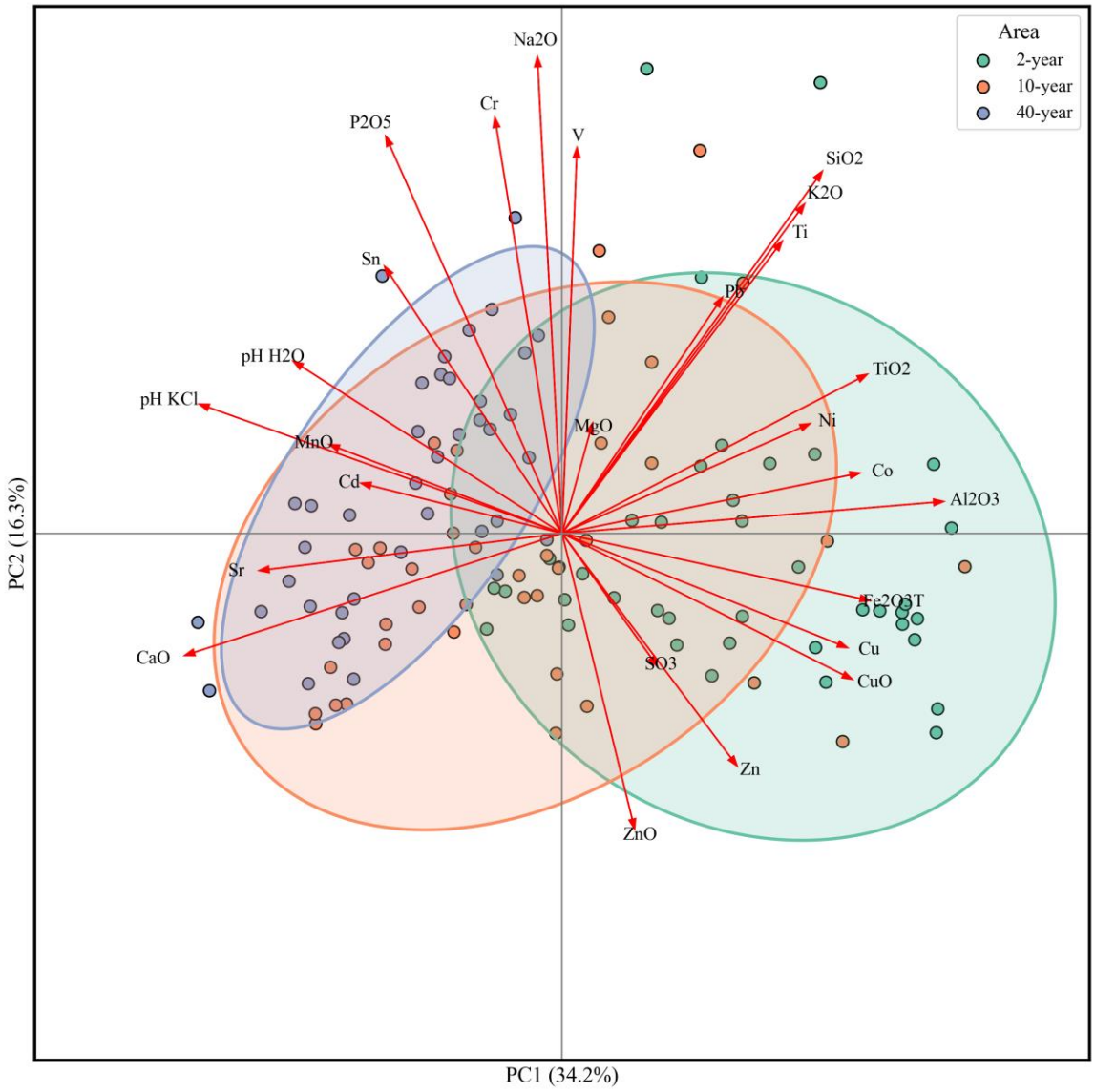
Lead (Pb) shows weak correlations overall, with moderately positive relationships with Ni, Ti, V, K_2O, and SiO_2, suggesting limited geochemical coupling. Tin (Sn) shows positive correlations with MnO, Cd, and V, P_2O_5, and pH KCl. Strontium (Sr) has a strong positive correlation with CaO, consistent with its known geochemical behavior as a calcium analog (BLASCHKO et al., 2013), and strong negative correlations with Al_2O_3, SiO_2, and TiO_2, reflecting inverse distribution with aluminosilicates. Sr is also negatively correlated with Cu, Co, CuO, Fe_2O_3T, and K_2O. Titanium (Ti) is positively correlated with Ni and TiO_2, and negatively correlated with CaO. Vanadium (V) has positive correlations with Sn and Ti, and weak correlations overall. Zinc (Zn) also shows weak correlations overall, with the exception of positive relationships with Fe_2O_3T, ZnO, Na_2O, and P_2O_5.

Figure 11 - Spearman correlation results for the Technosol properties.



The patterns observed in the PCA biplot (Figure 12) further support and contextualize the geochemical trends identified across the Technosol chronosequence. Samples from the 2-year site cluster distinctly on the positive side of PC1, aligning closely with variables such as Cu, CuO, Fe_2O_3T, Zn, ZnO, and Al_2O_3. This association highlights the strong initial influence of newly deposited scheelite mining residues, rich in metals and aluminosilicates. In contrast, 40-year samples are distributed toward the negative side of PC1 and the upper quadrants of PC2, showing proximity to CaO, Sr, MnO, Cd, and pH values, which are attributes indicative of carbonate accumulation and increased alkalinity over time. The intermediate position of 10-year samples reflects a transitional geochemical phase, retaining moderate levels of metals while gradually incorporating characteristics of a more developed pedogenesis. The separation among the three sample clusters of Technosol accumulation times underscores the progressive transformation of the soil system, likely driven by weathering, element mobility, and sorption dynamics. Moreover, the opposing vectors of CaO and Cu-related variables reinforce the antagonistic relationships observed in the correlation matrix.

Figure 12 - Principal component analysis biplot considering all available Technosol attributes.



4.3.2 Spectral modeling

The predictive performance of PLSR and CNN models varied considerably across spectral domains and target elements (Table 13). In general, mid-infrared (MIR) spectra yielded better results than near-infrared (NIR), showing that this region yielded higher quality spectral information, while fusion models (NIR+MIR) enhanced predictions only for some metals. Model performance also differed significantly between PLSR and CNN, with PLSR showing more stable results, whereas CNN performance was more variable, with some outstanding cases and many underperforming ones.

Table 13 - Prediction performance summary for all elements, considering both models and the three spectral datasets.

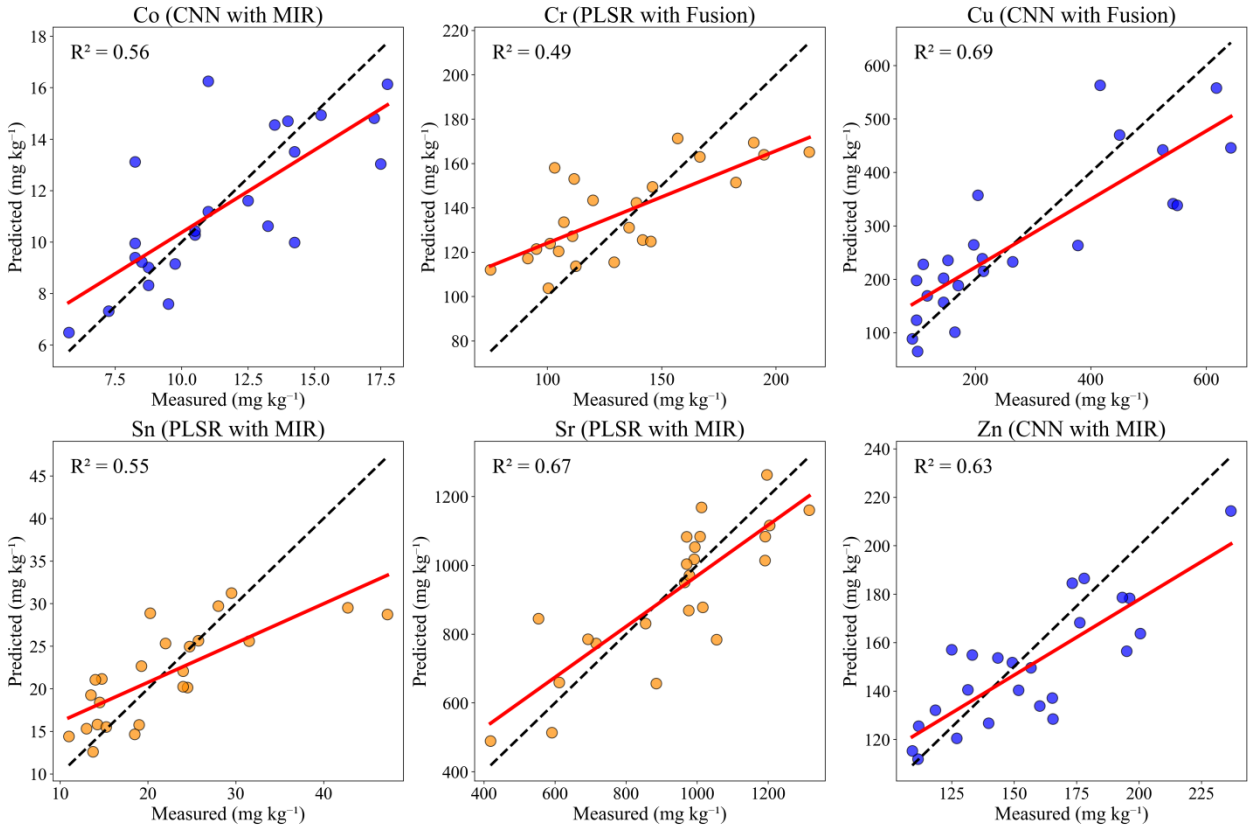
Model	Metric	Cd	Co	Cr	Cu	Ni	Pb	Sn	Sr	Ti	V	Zn
						NIR						
PLSR	RMSE	0.26	3.08	33.30	152.62	7.40	3.28	8.61	221.70	544.70	17.94	35.49
	R^2	-0.27	0.15	0.13	0.30	-0.25	-0.28	0.08	0.03	-0.36	-0.70	-0.19
	RPD	0.96	1.15	1.12	1.26	0.98	0.88	1.04	1.09	0.87	0.78	0.98
CNN	RMSE	0.23	3.06	30.76	181.78	6.12	2.97	8.12	210.07	455.53	13.39	32.67
	R^2	-0.01	0.16	0.26	0.00	0.14	-0.06	0.18	0.13	0.05	0.05	-0.01
	RPD	1.07	1.09	1.20	1.00	1.09	0.98	1.11	1.07	1.03	1.03	1.02
						MIR						
PLSR	RMSE	0.23	2.31	26.14	154.31	5.38	3.10	6.05	128.33	361.63	13.09	23.98
	R^2	0.01	0.52	0.47	0.28	0.34	-0.15	0.55	0.67	0.40	0.09	0.46
	RPD	1.05	1.45	1.38	1.25	1.24	0.94	1.49	1.76	1.41	1.08	1.37
CNN	RMSE	0.23	2.21	33.26	170.60	6.98	3.10	6.67	159.64	404.49	11.05	19.90
	R^2	-0.02	0.56	0.14	0.12	-0.11	-0.15	0.45	0.50	0.25	0.35	0.63
	RPD	1.11	1.51	1.08	1.07	1.01	0.94	1.35	1.43	1.18	1.26	1.71
						NIR + MIR Fusion						
PLSR	RMSE	0.20	2.36	25.48	150.22	5.67	2.90	6.87	137.60	379.42	12.55	24.58
	R^2	0.19	0.50	0.49	0.32	0.26	-0.01	0.42	0.63	0.34	0.17	0.43
	RPD	1.16	1.41	1.43	1.22	1.17	1.00	1.32	1.64	1.24	1.12	1.36
CNN	RMSE	0.32	2.58	30.98	100.46	6.59	3.50	7.82	146.95	472.83	13.96	27.37
	R^2	-1.01	0.40	0.25	0.69	0.01	-0.46	0.24	0.57	-0.02	-0.03	0.29
	RPD	0.88	1.48	1.16	1.81	1.12	0.83	1.15	1.53	1.14	1.06	1.35

The best overall prediction was obtained for Cu (Figure 13) using the CNN-Fusion model, which reached an R^2 of 0.69 and RPD of 1.81, setting it at the threshold of a satisfactory prediction, according to the adopted classification (WANG; ZHANG; PAN, 2017). This result reflects Cu strong geochemical association with oxides such as Fe_2O_3T and CuO, as supported by the correlation matrix and PCA. Additionally, Sr exhibited similar strong results (Figure 13), with an R^2 of 0.67 and RPD of 1.76 using the PLSR model with MIR spectra, and similarly favorable outcomes with the fusion dataset (R^2 = 0.63, RPD = 1.64), qualifying it as a reasonable prediction (WANG; ZHANG; PAN, 2017).

Furthermore, several elements achieved reasonable prediction performances under specific spectral and modeling configurations. Zn was predicted with relatively high accuracy (Figure 13) using CNN-MIR (R^2 = 0.63, RPD = 1.71). Co and Sn also demonstrated consistent predictive capacity across MIR and fusion inputs with both PLSR and CNN. The best performance for Co (Figure 13) was obtained using CNN with MIR spectra (R^2 = 0.56,

RPD = 1.51), while Sn showed optimal results (Figure 13) with PLSR with MIR spectra (R^2 = 0.55, RPD = 1.49). In the case of Cr (Figure 13), the fusion dataset combined with PLSR yielded an R^2 of 0.49 and an RPD of 1.43, placing it at the threshold for reasonable prediction (WANG; ZHANG; PAN, 2017).

Figure 13 - Predicted *versus* measured scatter plots for the six best predicted elements from Technosols.



For the remaining elements, including Cd, Ni, Pb, Ti, and V, predictions were generally poor, with R^2 values below 0.50 and RPD below 1.4 across all configurations. Notably, Cd, the element with the lowest concentrations, had R^2 <0 in most cases, with the best result being PLSR-Fusion (R^2 = 0.19, RPD = 1.16), still far from a reliable prediction. Similarly, Ti and V, which showed diffuse geochemical patterns, were not effectively predicted by any model. It is worth noting that NIR alone produced the lowest predictive performances overall, confirming the advantage of MIR and fusion spectral datasets. From a model perspective, PLSR tended to perform more consistently across elements, particularly in the MIR and fusion domains. CNN, while powerful, sometimes underperformed, as seen with CNN-Fusion for Cd, which resulted in an R^2 of -1.01. This instability may reflect overfitting

or difficulties in learning from small sample sizes, a common flaw exhibited by neural networks in general (DOMINGOS, 2012).

Overall, infrared spectroscopy modeling showed promising results for Co, Cr, Cu, Sn, Sr, and Zn, particularly with MIR and NIR+MIR fusion spectra using PLSR or CNN. These elements showed distinct geochemical behavior, as revealed by the Spearman correlation analysis (Figure 11) and PCA biplot (Figure 12). For instance, Cu, Co, and Zn, which were among the most accurately predicted, showed high initial concentrations and positive correlations with Fe_2O_3T and Al_2O_3, suggesting affinity with oxides that may enhance spectral detectability in the MIR region. Likewise, Sr showed strong positive association with CaO and negative with Fe_2O_3T and Al_2O_3, supported by both Spearman correlation (Figure 11) and PCA analysis (Figure 12), likely contributing to its high prediction performance, which was notable under MIR and NIR+MIR fusion spectra. Sn, which also reached reasonable prediction levels, demonstrated consistent relationships with MnO and P_2O_5. Such findings reinforce the idea that predictability is enhanced when metals are geochemically coupled with spectrally active components such as these oxides.

On the other hand, elements such as Cd, Ni, Pb, Ti, and V, which were poorly predicted, tended to exhibit weaker or more diffuse geochemical associations, as shown by low Spearman correlation values (Figure 11) and scattered directions in the PCA biplot (Figure 12). The combination of diffuse geochemical behavior and low concentrations likely resulted in missing spectral features, contributing to their poor predictive performance. These limitations highlight the need for further refinement in model calibration, possibly including a larger set of samples to better cover the variability of the geochemical features.

4.4 Conclusions

Technosols derived from scheelite mining residues at Brejuí Mine exhibit distinct geochemical patterns at the sites with 2, 10, and 40 years of accumulation. For all sites, most metals have average values much higher than those found for the Earth's upper continental crust, with the exception of Pb for all sites and Co for the 10- and 40-year sites. The geochemical evolution of the Technosols over the chronosequence is characterized by the progressive enrichment of Cd, Cr, Sn, Sr, V, MnO, P_2O_5, CaO, and Na_2O, and an increase in pH, alongside the depletion of Co, Cu, Ni, Pb, Ti, Zn, Al_2O_3, SiO_2, Fe_2O_3T, TiO_2, CuO, K_2O, and SO_3, while ZnO and MgO remained relatively stable. These trends reflect the occurrence of broad chemical transformations which drive the redistribution and stabilization of metals over time.

Spearman correlation analyses disclosed some notable underlying relationships governing element distributions, such as strong positive associations of Cu, Co, and Zn with Fe and Al oxides, Sr with CaO, and Sn with MnO. The principal component analysis (PCA) revealed distinct chronosequence clusters, higlighting the transition from 2-year-old fresh tailings to the 40-year-old weathered substrates, reinforcing the geochemical relationships previously found between metals and oxides.

Infrared spectroscopy modeling, combining NIR, MIR, and NIR+MIR fusion spectral datasets with Partial Least Squares regression (PLSR) and Convolutional Neural Network (CNN), demonstrated promising but variable performance for the prediction of metals. MIR spectra, alone or fused with NIR, yielded reasonable predictions for Co (R^2 = 0.56, RPD = 1.51; CNN-MIR), Cr (R^2 = 0.49, RPD = 1.46; PLSR-Fusion), Cu (R^2 = 0.69, RPD = 1.81; CNN-Fusion), Sn (R^2 = 0.55, RPD = 1.49; PLSR-MIR), Sr (R^2 = 0.67, RPD = 1.76; PLSR-MIR), and Zn (R^2 = 0.63, RPD = 1.71; CNN-MIR), reflecting the strong geochemical coupling of these elements with potential spectrally active components such as Ca, Al, Fe, and Mn oxides. Conversely, poor model accuracy for Cd, Ni, Pb, Ti, and V underscores limitations imposed by either low concentrations or weak geochemical relationships with spectrally active components.

The positive results obtained for Co, Cr, Cu, Sn, Sr, and Zn support the integration of MIR-based predictive systems within routine monitoring protocols of Technosols. Collectively, our findings strengthen the promise of infrared spectroscopy as an efficient alternative tool for the environmental monitoring of Technosols, while considering its limitations, paving the way for continuous assessment to safeguard soil health and support the conversion of mining tailings into sustainable substrates.

References

ARBESTAIN, M. C.; MADINABEITIA, Z.; HORTALÀ, M. A.; MACÍAS-GARCÍA, F.; VIRGEL, S.; MACÍAS, F. Extractability and leachability of heavy metals in Technosols prepared from mixtures of unconsolidated wastes. **Waste Management**, v. 28, n. 12, p. 2653–2666, 2008. https://doi.org/10.1016/j.wasman.2008.01.008

AZEVEDO-LOPES, T.; QUEIROZ, H. M.; RUIZ, F.; ASENSIO, V.; FERREIRA, A. D.; CHERUBIN, M. R.; FERREIRA, T. O. From waste to soil: Technosols made with construction and demolition waste as a nature-based solution for land reclamation. **Waste Management**, v. 186, p. 153–165, 2024. https://doi.org/10.1016/j.wasman.2024.06.010

BEZERRA JÚNIOR, J. G. O.; DA SILVA, N. M. Caracterização geoambiental da microrregião do seridó oriental do rio grande do norte. **Holos**, v. 2, p. 78–91, 2008. https://doi.org/10.15628/holos.2007.102

BLASCHKO, S. D.; CHI, T.; MILLER, J.; FLECHNER, L.; FAKRA, S.; KAPAHI, P.; STOLLER, M. L. Strontium substitution for calcium in lithogenesis. **The Journal of Urology**, v. 189, n. 2, p. 735–739, 2013. https://doi.org/10.1016/j.juro.2012.08.199

CHAKRAVARTULA, S. S. N.; MOSCETTI, R.; BEDINI, G.; NARDELLA, M.; MASSANTINI, R. Use of convolutional neural network, CNN combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee. **Food Control**, v. 135, p. 108816, 2022. https://doi.org/10.1016/j.foodcont.2022.108816

COUSSY, S.; GRANGEON, S.; BATAILLARD, P.; KHODJA, H.; MAUBEC, N.; FAURE, P.; SCHWARTZ, C.; DAGOIS, R. Evolution of iron minerals in a 100 years-old Technosol. Consequences on Zn mobility. **Geoderma**, v. 290, p. 19–32, 2017. https://doi.org/10.1016/j.geoderma.2016.12.009

COZZOLINO, D. Near infrared spectroscopy as a tool to monitor contaminants in soil, sediments and water—State of the art, advantages and pitfalls. **Trends in Environmental Analytical Chemistry**, v. 9, p. 1–7, 2016. https://doi.org/10.1016/j.teac.2015.10.001

DANTAS, J. R. A. Distritos mineiros do nordeste oriental. Departamento Nacional de Produção Mineral – DNPM, 2002.

DI IORIO, E.; NAPOLETANO, P.; CIRCELLI, L.; MEMOLI, V.; SANTORUFO, L.; DE MARCO, A.; COLOMBO, C. Comparison of natural and technogenic soils developed on volcanic ash by Vis-NIR spectroscopy. **Catena**, v. 216, p. 106369, 2022. https://doi.org/10.1016/j.catena.2022.106369

DOMINGOS, P. A few useful things to know about machine learning. **Communications of the ACM**, v. 55, n. 10, p. 78–87, 2012. https://doi.org/10.1145/2347736.2347755

DUTRA, A. S. F. **Potencial fitorremediador da algaroba (Prosopis juliflora (Sw.) DC) e percepção quanto aos riscos toxicológicos associados a rejeitos de mineração no semiárido brasileiro**. Dissertação (Mestrado em Desenvolvimento e Meio Ambiente) - Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal, 2019.

ECHEVARRIA, G.; MOREL, J. L. Technosols of mining areas. **Tópicos em ciência do solo**, Viçosa, MG: Sociedade Brasileira de Ciência do Solo, v. 9, p. 92–111, 2015.

ESTÉVEZ ALVAREZ, J. R.; MONTERO, A. A.; JIMÉNEZ, N. H.; et al. Nuclear and related analytical methods applied to the determination of Cr, Ni, Cu, Zn, Cd and Pb in a red ferralitic soil and Sorghum samples. **Journal of Radioanalytical and Nuclear Chemistry**, v. 247, p. 479–486, 2001. https://doi.org/10.1023/A:1010640009301

FABBRI, D.; PIZZOL, R.; CALZA, P.; MALANDRINO, M.; GAGGERO, E.; PADOAN, E.; AJMONE-MARSAN, F. Constructed technosols: A strategy toward a circular economy. **Applied Sciences**, v. 11, n. 8, p. 3432, 2021. https://doi.org/10.3390/app11083432

FRIERDICH, A. J.; CATALANO, J. G. Fe (II)-mediated reduction and repartitioning of structurally incorporated Cu, Co, and Mn in iron oxides. **Environmental Science & Technology**, v. 46, n. 20, p. 11070–11077, 2012. https://doi.org/10.1021/es302236v

GERAB, A. T. F. de S. C. **Utilização do resíduo grosso do beneficiamento da scheelita em aplicações rodoviárias**. Dissertação (Mestrado em Engenharia Civil) - Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, p. 123, 2014.

GHOLIZADEH, A.; BORŮVKA, L.; SABERIOON, M.; VAŠÁT, R. Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. **Applied Spectroscopy**, v. 67, n. 12, p. 1349–1362, 2013. https://doi.org/10.1366/13-07288

GODEIRO, M. L. da S.; BORGES JÚNIOR, J. P.; FERNANDES, B. R. B.; LEITE, J. Y. P. Caracterização de pré-concentrado do rejeito de scheelita da mina brejuí em concentrador centrífugo. **Holos**, v. 5, p. 81–90, 2011. https://doi.org/10.15628/holos.2010.557

GUERRÓN, D. B.; CAPA, J.; FLORES, L. C. Retention of heavy metals from mine tailings using Technosols prepared with native soils and nanoparticles. **Heliyon**, v. 7, n. 7, 2021. https://doi.org/10.1016/j.heliyon.2021.e07631

HUOT, H.; SIMONNOT, M. O.; WATTEAU, F.; MARION, P.; YVON, J.; DE DONATO, P.; MOREL, J. L. Early transformation and transfer processes in a Technosol developing on iron industry deposits. **European Journal of Soil Science**, v. 65, n. 4, p. 470–484, 2014. https://doi.org/10.1111/ejss.12106

KÄSTNER, F.; SUT-LOHMANN, M.; RAMEZANY, S.; RAAB, T.; FEILHAUER, H.; CHABRILLAT, S. Estimating heavy metal concentrations in Technosols with reflectance spectroscopy. **Geoderma**, v. 406, p. 115512, 2022. https://doi.org/10.1016/j.geoderma.2021.115512

KINGMA, D. P.; BA, J. Adam: A Method for Stochastic Optimization. In: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, May 7-9, 2015, San Diego, CA, USA, 2015.

LECUN, Y.; BENGIO, Y. Convolutional networks for images, speech, and time series. **The handbook of brain theory and neural networks**, v. 3361(10), 1995.

LEGUÉDOIS, S.; SÉRÉ, G.; AUCLERC, A.; CORTEt, J.; HUOT, H.; OUVRARD, S.; MOREL, J. L. Modelling pedogenesis of Technosols. **Geoderma**, v. 262, p. 199–212, 2016. https://doi.org/10.1016/j.geoderma.2015.08.008

LEHMANN, A. Technosols and other proposals on urban soils for the WRB [World Reference Base for Soil Resources]. **International Agrophysics**, v. 20, n. 2, 2006.

NASCIMENTO, A. R. V. J.; CUNHA, G. K. G.; DO NASCIMENTO, C. W. A.; et al. Assessing Soil Quality and Heavy Metal Contamination on Scheelite Mining Sites in a Tropical Semi-arid Setting. **Water, Air, and Soil Pollution**, v. 232, p. 375, 2021. https://doi.org/10.1007/s11270-021-05299-6

NG, W.; MINASNY, B.; MONTAZEROLGHAEM, M.; PADARIAN, J.; FERGUSON, R.; BAILEY, S.; MCBRATNEY, A. B. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. **Geoderma**, v. 352, p. 251–267, 2019. https://doi.org/10.1016/j.geoderma.2019.06.016

NIST - National Institute of Standards and Technology. Standard Reference Materials - SRM 2709, p. 2710-2711, 2002.

PYO, J.; HONG, S. M.; KWON, Y. S.; KIM, M. S.; CHO, K. H. Estimation of heavy metals using deep neural network with visible and infrared spectroscopy of soil. **Science of the Total Environment**, v. 741, p. 140162, 2020. https://doi.org/10.1016/j.scitotenv.2020.140162

REDDI, S. J.; KALE, S.; KUMAR, S. On the convergence of Adam and beyond. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, Apr 30–May 3, Vancouver, CA, 2018.
PADARIAN, J.; MINASNY, B.; MCBRATNEY, A. B. Using deep learning to predict soil properties from regional spectral data. **Geoderma Regional**, v. 16, p. e00198, 2019. https://doi.org/10.1016/j.geodrs.2018.e00198

PANDEY, A.; JAIN, K. A robust deep attention dense convolutional neural network for plant leaf disease identification and classification from smart phone captured real world images. **Ecological Informatics**, v. 70, p. 101725, 2022. https://doi.org/10.1016/j.ecoinf.2022.101725

RAMOS FILHO, R. E. B. **Estudo da combinação de resíduo de scheelita, pó de pedra e manipueira com aglomerantes para produção de tijolos ecológicos**. Tese (Doutorado em Ciência e Engenharia de Materiais) - Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Natal / RN, 2021.

REES, F.; DAGOIS, R.; DERRIEN, D.; FIORELLI, J. L.; WATTEAU, F.; MOREL, J. L.; SÉRÉ, G.; et al. Storage of carbon in constructed technosols: in situ monitoring over a decade. **Geoderma**, v. 337, p. 641–648, 2019. https://doi.org/10.1016/j.geoderma.2018.10.009

ROKIA, S.; SÉRÉ, G.; SCHWARTZ, C.; DEEB, M.; FOURNIER, F.; NEHLS, T.; VIDAL-BEAUDET, L.; et al. Modelling agronomic properties of Technosols constructed with urban wastes. **Waste Management**, v. 34, n. 11, p. 2155–2162, 2014. https://doi.org/10.1016/j.wasman.2013.12.016

RODRÍGUEZ-ESPINOSA, T.; NAVARRO-PEDREÑO, J.; GÓMEZ-LUCAS, I.; et al. Urban areas, human health and technosols for the green deal. **Environmental Geochemistry and Health**, v. 43, p. 5065–5086, 2021. https://doi.org/10.1007/s10653-021-00953-8

SAVITZKY, A.; GOLAY, M. J. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, 1964. https://doi.org/10.1021/ac60214a047

SCHAD, P. Technosols in the World Reference Base for Soil Resources – history and definitions. **Soil Science and Plant Nutrition**, v. 64, n. 2, p. 138–144, 2018. https://doi.org/10.1080/00380768.2018.1432975

SÉRÉ, G.; SCHWARTZ, C.; OUVRARD, S.; et al. Early pedogenic evolution of constructed Technosols. **Journal of Soils and Sediments**, v. 10, p. 1246–1254, 2010. https://doi.org/10.1007/s11368-010-0206-6

SHANG, H.; SHANG, L.; WU, J.; XU, Z.; ZHOU, S.; WANG, Z.; WANG, H.; YIN, J. NIR spectroscopy combined with 1D-convolutional neural network for breast cancerization analysis and diagnosis. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 287, p. 121990, 2023. https://doi.org/10.1016/j.saa.2022.121990

SILVA, R. C.; CORRÊA, G. R.; ARRUDA, D. M.; VELOSO, G. V.; FERNANDES-FILHO, E. I.; CÂNDIDO, H. G.; SCHAEFER, C. E. G. R.; et al. The Brazilian semiarid region over the past 21,000 years: Vegetation dynamics in small pulses of higher humidity. **Ecological Informatics**, v. 77, p. 102259, 2023. https://doi.org/10.1016/j.ecoinf.2023.102259

TEIXEIRA, P. C.; DONAGEMMA, G. K.; FONTANA, A.; TEIXEIRA, W. G.; editores técnicos. **Manual de Métodos de Análise de Solo** - 3. ed. rev. e ampl. – Brasília, DF: Embrapa, 2017.

THAKUR, P. S.; CHATURVEDI, S.; KHANNA, P.; SHEOREY, T.; OJHA, A. Vision transformer meets convolutional neural network for plant disease classification. **Ecological Informatics**, v. 77, p. 102245, 2023. https://doi.org/10.1016/j.ecoinf.2023.102245

WAHSHA, M.; BINI, C.; ARGESE, E.; MINELLO, F.; FONTANA, S.; WAHSHEH, H. Heavy metals accumulation in willows growing on Spolic Technosols from the abandoned Imperina Valley mine in Italy. **Journal of Geochemical Exploration**, v. 123, p. 19–24, 2012. https://doi.org/10.1016/j.gexplo.2012.07.004

WANG, C.; ZHANG, T.; PAN, X. Potential of visible and near-infrared reflectance spectroscopy for the determination of rare earth elements in soil. **Geoderma**, v. 306, p. 120–126, 2017. https://doi.org/10.1016/j.geoderma.2017.07.016

WEDEPOHL, K. H. The composition of the continental crust. **Geochimica et Cosmochimica Acta**, v. 59, n. 7, p. 1217–1232, 1995. https://doi.org/10.1016/0016-7037(95)00038-2

YAO, F. X.; MACÍAS, F.; VIRGEL, S.; BLANCO, F.; JIANG, X.; ARBESTAIN, M. C. Chemical changes in heavy metals in the leachates from Technosols. **Chemosphere**, v. 77, n. 1, p. 29–35, 2009. https://doi.org/10.1016/j.chemosphere.2009.06.012

5 CONVOLUTIONAL NEURAL NETWORKS FOR PREDICTION OF Ba, Cd, Cu, AND Pb CONTENTS IN SOIL USING A MID-INFRARED OPEN SPECTRAL LIBRARY

Abstract

Soil contamination by heavy metals poses risks to ecosystems, agriculture, and human health, making concentration monitoring essential. Regular assessments help track concentration changes, identify pollution sources, and guide remediation. However, traditional analytical methods are costly and complex, while infrared spectroscopy offers a rapid, non-destructive alternative for more efficient soil analysis. Nonetheless, this method faces challenges in predicting heavy metal concentrations accurately, relying on indirect detection methods. While soil spectral libraries have been applied extensively to estimate general soil properties, the prediction of heavy metal concentrations remains underexplored. Therefore, we focused on the prediction of Ba, Cd, Cu, and Pb concentrations in soil using mid-infrared (MIR) data available in the Open Soil Spectral Library (OSSL). The dataset consisted of a total of 773 soil samples presenting elemental concentrations (mg kg^{-1}) and MIR absorbance spectra (4000 – 600 cm^{-1}). Prediction models were built using Partial Least Squares (PLS) regression and a Convolutional Neural Network (CNN). Both raw and preprocessed spectra were used in the modeling. Taking into account the diverse and highly heterogeneous nature of the soil data, positive results were found. CNN generated predictions that were closely aligned with the measured values, while the PLS model exhibited a general tendency to underestimate, indicating lower accuracy. The prediction performance with CNN for Cd (R^2 = 0.81, RPD = 2.32), and for Ba, Cu, and Pb (R^2 > 0.70, RPD > 1.8) suggests the potential for developing a MIR-based monitoring system to quickly and cost-effectively screen soil contamination levels for these metals.

Keywords: Soil pollution. Trace elements. Soil health. Soil spectral data.

REDES NEURAIS CONVOLUCIONAIS PARA PREDIÇÃO DOS TEORES DE Ba, Cd, Cu E Pb EM SOLOS UTILIZANDO UMA BIBLIOTECA ESPECTRAL ABERTA DE INFRAVERMELHO MÉDIO

Resumo

A contaminação do solo por metais pesados representa riscos aos ecossistemas, à agricultura e à saúde humana, tornando o monitoramento das concentrações essencial. Avaliações regulares ajudam a acompanhar variações nos teores, identificar fontes de poluição e orientar ações de remediação. No entanto, os métodos analíticos tradicionais são custosos e complexos, enquanto a espectroscopia no infravermelho surge como alternativa rápida e não destrutiva para uma análise mais eficiente do solo. Ainda assim, esse método enfrenta desafios na predição precisa das concentrações de metais pesados, por depender de mecanismos de detecção indiretos. Embora bibliotecas espectrais de solos venham sendo amplamente utilizadas para estimativas de propriedades gerais do solo, a predição de teores de metais pesados permanece pouco explorada. Diante disso, este estudo teve como foco a predição das concentrações de Ba, Cd, Cu e Pb em solos, utilizando dados no infravermelho médio (MIR) disponibilizados pela Open Soil Spectral Library (OSSL). O conjunto de dados foi composto por 773 amostras de solo, contendo valores de concentração elementar (mg kg^{-1}) e espectros de absorbância no MIR (4000 – 600 cm^{-1}). Modelos de predição foram construídos utilizando regressão por mínimos quadrados parciais (PLS) e redes neurais convolucionais (CNN), empregando tanto espectros brutos quanto pré-processados. Considerando a natureza diversa e altamente heterogênea dos dados, os resultados foram positivos. As predições geradas por CNN mostraram-se próximas dos valores medidos, enquanto o modelo PLS apresentou tendência à subestimação, indicando menor acurácia. O desempenho das predições por CNN para Cd (R^2 = 0,81; RPD = 2,32) e para Ba, Cu e Pb (R^2 > 0,70; RPD > 1,8) demonstra o potencial para o desenvolvimento de um sistema de monitoramento baseado em MIR para triagem rápida e de baixo custo da contaminação do solo por esses metais.

Palavras-chave: Poluição do solo. Elementos traço. Saúde do solo. Dados espectrais do solo.

5.1 Introduction

Soil contamination by heavy metals is a critical environmental issue (MOTUZOVA et al., 2014) that arises from the accumulation of these potentially toxic elements in the soil profile, where it may lead to detrimental effects on ecosystems (SINGH; PRASAD, 2015), agricultural productivity (ELANGO et al., 2022) and human health (ALEGEBAWY et al., 2021). A clear understanding of this problem is essential for improved environmental management. Furthermore, the most important requirement for the mitigation and control of soil contamination by heavy metals is the monitoring of the elemental concentrations.

Regular assessments of heavy metal contents in soils help track changes in concentrations over time (REN et al., 2022), identify pollution sources (LI et al., 2014), assess ecological and contamination risks (KUMAR et al., 2019; XIANG et al., 2021), and inform remediation efforts (WANG et al., 2022a). However, there are limitations involved in the effective assessment of these elements. The most notable limitation is the reliance on traditional laboratory methods that involve complex sample preparation and analysis (COZZOLINO, 2016; DEMATTÊ et al., 2019a), demanding highly skilled labour and hazardous reagents. While these methods are accurate and precise, they can be time-consuming and expensive, making them less feasible for large-scale or real-time monitoring. In this context, infrared spectroscopy offers a non-destructive and rapid means of analyzing soil samples (SHI et al., 2014; CAPUANO; VAN RUTH, 2015). This technique functions by measuring the response of the interaction between matter and infrared radiation, resulting in spectral data that can be processed and analyzed for the estimation of multiple soil properties (VISCARRA ROSSEL et al., 2006). Infrared spectroscopy has the potential to revolutionize the field of soil monitoring by permitting faster and more cost-effective data collection.

However, further development of soil infrared spectroscopy requires extensive and diversified data for testing and exploring novel approaches. The availability of open soil spectral data has greatly facilitated research, exemplified by initiatives like the Open Soil Spectral Library (OSSL) (SAFANELLI et al., 2021), the Global Soil Spectral Calibration Library and Estimation Service (SHEPHERD et al., 2022), the Brazilian Soil Spectral Library (Demattê et al., 2019b), the European Land Use/Land Cover area frame Survey 2009 (LUCAS) (ORGIAZZI et al., 2018) and the Soil Spectral Library of New Zealand (MA et al., 2023). These resources provide valuable information and research materials, promoting collaboration, transparency, and open access data that can contribute to more robust scientific investigations in regard of soil infrared spectroscopy.

Despite its wide-ranging promise, infrared spectroscopy does, however, present analytical limitations, particularly when it comes to predicting soil heavy metal concentrations accurately. While there has been substantial research on the prediction of heavy metal concentrations in soils based on infrared spectroscopy (SONG et al., 2012; LIU et al., 2018; ZHANG et al., 2019; LIU et al., 2020; KÄSTNER et al., 2022; MAIA et al., 2022; NYARKO; TACK; MOUAZEN, 2022), most of these are focused on specific environmental contexts or limited to predefined pedological settings. It is important to note that the estimation of heavy metal concentrations from soil infrared spectra is mainly achieved by indirect detection (HORTA et al., 2015; COZZOLINO, 2016), therefore depending on the geochemical features which correlates the elements of interest with other soil properties.

While there has been extensive research on the application of soil spectral libraries for the estimation of general soil properties such as particle size distribution (ZHAO et al., 2018; JANIK; SORIANO-DISLA; FORRESTER, 2020; SANTANA et al., 2021) and carbon content (MOURA-BUENO et al., 2019; LUCE; ZIADI; ROSSEL, 2022; WANG et al., 2022b), there is an evident scarcity when it comes to the estimation of heavy metal concentrations. In this context, we have focused on the estimation of Ba, Cd, Cu, and Pb concentrations using data from a large soil spectral library. We utilized mid-infrared (MIR) and wet-chemistry data from the Kellogg Soil Survey Laboratory (KSSL) of the Natural Resources Conservation Service of the United States Department of Agriculture (USDA), available in the Open Soil Spectral Library (OSSL). The dataset was filtered, processed, and exploratory data analysis undertaken. The MIR spectra were preprocessed using Savitzky-Golay (SG) derivative and Standard Normal Variate (SNV). Then, prediction models with raw and preprocessed spectra were built for Ba, Cd, Cu, and Pb concentrations using Partial Least Squares (PLS) regression and a Convolutional Neural Network (CNN) with depthwise separable convolution.

5.2 Materials and Methods

5.2.1 Data acquisition

The dataset was obtained from the Open Soil Spectral Library (OSSL) (SAFANELLI et al., 2021), containing >20,000 samples from the Kellogg Soil Survey Laboratory (KSSL) of the Natural Resources Conservation Service of the United States Department of Agriculture (USDA). All analyses and laboratory procedures conducted for generating this dataset are described in the Kellogg Soil Survey Laboratory methods manual (SOIL SURVEY STAFF,

2022). The KSSL spectral library is credited to Wijewardane et al. (2016) and Sanderman et al. (2020). All data made public by the OSSL is licensed by Creative Commons Attribution 4.0 International CC-BY and are therefore open for use, sharing, and adaptation.

Taking the entire dataset, only a limited amount of samples presented both MIR spectra (absorbance at 600 – 4000 cm^{-1} with 2 cm^{-1} spectral resolution) and elemental concentrations for heavy metals (mg kg^{-1}). Hence, the dataset was filtered for the availability and data consistency for both MIR spectra and the concentrations of Ba, Cd, Cu, and Pb, resulting in a total of 773 samples. The locations of the selected samples span across most of North America. KSSL MIR data were generated by means of Diffuse Reflectance Fourier Transform Spectroscopy (DRIFTS, Bruker Vertex-70 Spectrometer with HTS-XT) from air-dried <2 mm soil samples ground using an 80 mesh (<177 μm). The resulting spectral interferograms were then processed by Fourier transform and converted to absorbance spectra (SOIL SURVEY STAFF, 2022).

The extraction method for the KSSL metals was Mehlich-3, using 2.5 g of air-dried, crushed and sieved (<2 mm) soil samples. The metal contents in the extracts were measured using an Inductively Coupled Plasma Atomic Emission Spectrophotometer (ICP-AES) in axial mode with a cross-flow nebulizer. Quality assurance and control of the analyses are available for consultation in KSSL LIMS (SOIL SURVEY STAFF, 2022). It is important to note that while most of the legislation limits for risk elements are based on aqua regia digestion, Mehlich-3 has been tested as a surrogate method with good results for screening heavy metal contents in soil (MALÝ; ZBÍRAL; ČIŽMÁROVÁ, 2021).

Additional soil properties data were also available for the selected samples, comprising absolute particle size distribution (% of sand, silt, and clay; pipette method), pH (H_2O), sum of bases (cmolc kg^{-1}; NH_4O_{Ac} method), cation exchange capacity (cmolc kg^{-1}; displacement after washing, NH_4O_{Ac}, pH 7), organic carbon (g kg^{-1}; dry combustion method), total carbon (g kg-1; dry combustion method), and total nitrogen (g kg^{-1}; combustion method).

5.2.2 Exploratory data analysis and dataset splitting

With the filtered dataset of 773 samples, the distributions of elemental concentrations for Ba, Cd, Cu, and Pb were analyzed by means of descriptive statistics. Since these variables did not exhibit a normal distribution, Spearman correlation analyses were conducted between the elemental concentrations and the following soil properties: sand, silt, clay, pH, sum of bases (SB), cation exchange capacity (CEC), organic carbon (OC), total carbon (TC), and

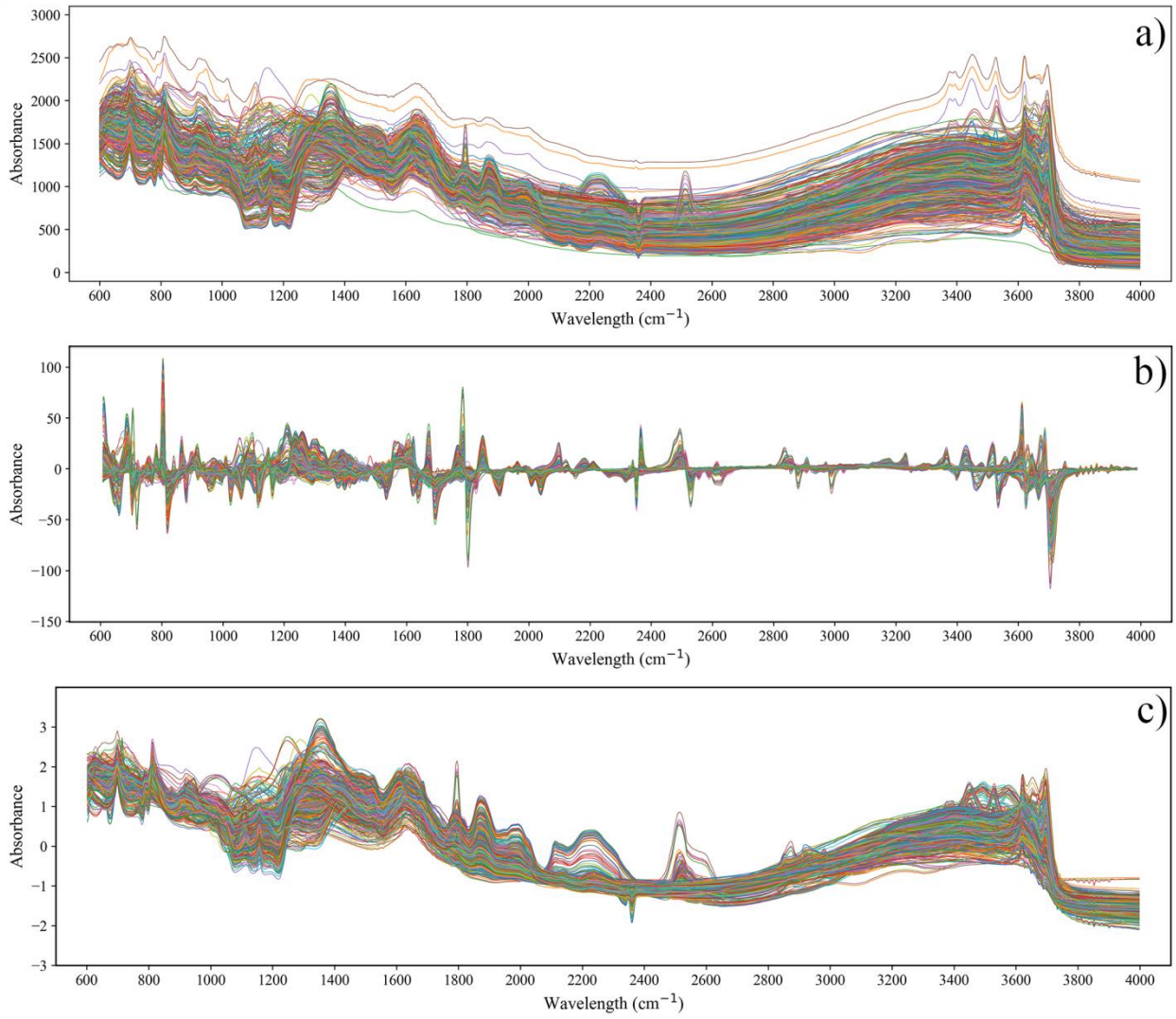
total nitrogen (TN). MIR data comprised 1701 data points per sample, consisting of absorbance values measured between 600 and 4000 cm^{-1}, with a 2 cm^{-1} spectral resolution.

To prepare the data for building the prediction models, the dataset was split for each element with 80% of the samples for training and 20% for validation. To ensure a representative split between training and validation sets, the distributions of the dependent variables (the concentration of each element) were taken into account by the splitting algorithm. All steps comprising data analysis, transformation, and splitting were conducted with Python (version 3.11.0) using the libraries ‘pandas’ (version 1.5.3) and ‘scikit-learn’ (version 1.2.2).

5.2.3 Spectral preprocessing

Preprocessing of spectral data is essential for reducing noise, correcting baseline shifts, and normalizing the measurements. In this context, MIR spectra from all samples (n = 773) were preprocessed using two well-established techniques: the Savitzky-Golay (SG) derivative and Standard Normal Variate (SNV). SG aims to smooth the raw spectra by computing the second derivative. This method reduces high-frequency noise while preserving key spectral features, making subtle variations more detectable and improving the overall quality of the data for further analysis (SAVITZKY; GOLAY, 1964). Meanwhile, SNV aims to normalize the spectra and mitigate scatter effects (BARNES; DHANOA; LISTER, 1989). SNV adjusts each spectrum to have a mean of zero and a standard deviation of one, effectively compensating for variations due to differences in particle size, common in spectroscopic data. The raw and preprocessed MIR spectral signatures can be visualized in Figure 14.

Figure 14 - Mid-infrared spectral signatures from the soil spectral library samples (n = 773): a) Raw spectra; b) Spectra preprocessed by the Savitzky-Golay derivative (SG); c) Spectra preprocessed by the Standard Normal Variate (SNV).



5.2.4 Prediction modeling

With the specific aim of achieving the optimal development of robust models for the prediction of soil heavy metals concentrations, Partial Least Squares (PLS) regression and Convolutional Neural Network (CNN) were used. These algorithms are both well-established in scientific research (GARTHWAITE, 1994; FALEH; KACHOURI, 2023), and have also been successfully tested for the prediction of heavy metals in soils based on infrared spectra (PANDIT; FILIPPELLI; LI, 2010; TODOROVA et al., 2014; LIU et al., 2018; PYO et al., 2020).

5.2.4.1 Partial Least Squares (PLS)

PLS regression is a classical method for the prediction of soil properties using infrared spectroscopy (UDELHOVEN; EMMERLING; JARMER, 2003; JANIK et al., 2007; PANDIT; FILIPPELLI; LI, 2010; WANG et al., 2014; LI et al., 2021). A PLS regression model was built for the prediction of the elemental concentrations of each element (Ba, Cd, Cu, and Pb). The models were selected with a 10-fold cross-validation with grid search for the best number of components ('n_components', 2, 4, 6, 8). Then, the best cross-validated model for each element was calibrated with the training set and validated with the test set. The PLS models were built using Python (version 3.11.0) with the 'PLSRegression' module from the 'scikit-learn' library (version 1.2.2).

5.2.4.2 Convolutional Neural Network (CNN)

To harness the potential of deep learning for the prediction of the elemental concentrations, we employed a Convolutional Neural Network (CNN). CNNs are able to process multidimensional data arrays by applying a mathematical operation known as convolution (PADARIAN et al., 2019), in which a kernel or filter is slid across the input data, and the integral of the pointwise multiplication at each position is computed. A single convolutional layer discerns basic features, and the addition of more layers enables the extraction of progressively abstract and complex features (LECUN; BENGIO, 1995). CNNs are largely used for tasks involving image-like representations (Pandey and Jain, 2022; THAKUR et al., 2023) such as infrared spectroscopy data (CHAKRAVARTULA et al., 2022; SHANG et al., 2023), which can be considered a one-dimensional (1D) image in this context. Soil properties have previously been simultaneously predicted with high accuracy from infrared spectra using CNN (NG et al., 2019).

The CNN architecture utilized in this study (Appendix A) is the depthwise separable convolution proposed by Yang et al. (2020) as an adaptation of the Xception structure created by Chollet (2017). This architecture was tailored to extract local and abstract features efficiently from the raw spectral data. The optimization algorithm chosen for training the model was the 'Adam' optimizer (KINGMA; BA, 2015; REDDI; KALE; KUMAR, 2018) with the default learning rate of 0.001 and no weight decay. The training was undertaken for 200 epochs with a 5-samples batch size. The CNN was implemented in Python (version 3.11.0) using the 'keras' module from the 'TensorFlow' framework (version 2.12.0rc0).

5.2.5 Model Evaluation

The performance of the regression models was assessed using conventional statistical metrics. These metrics include the Root Mean Squared Error (RMSE), Normalized Root Mean Squared Error (NRMSE), Coefficient of Determination (R^2), and the Ratio of Prediction to Deviation (RPD). The RMSE (Eq. 1) measures the average magnitude of the errors between predicted and observed values, providing an indication of the model's prediction accuracy (CHAI; DRAXLER, 2014). It is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad \text{Eq. 1}$$

where: y_i is the observed value, $\hat{y}_i$ is the predicted value, and n is the total number of observations.

The NRMSE (Eq. 2) normalizes the RMSE dividing it by the mean of the observed values, allowing for the comparison between variables with different scales and units of measurement. It is calculated as:

$$NRMSE = RMSE / \bar{y} \quad \text{Eq. 2}$$

where: $\bar{y}$ is the mean of the observed values.

The R^2 (Eq. 3) represents the proportion of the variance in the observed values that is explained by the model (ZHANG, 2017). It ranges from 0 to 1, with higher values indicating better model performance. It is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \text{Eq. 3}$$

where: y_i is the observed value, $\hat{y}_i$ is the predicted value, n is the total number of observations, and $\bar{y}$ is the mean of the observed values.

The RPD (Eq. 4) measures the ratio between the standard deviation of the observed values and the RMSE. It provides insight into the predictive ability of the model relative to the variability in the data. It is calculated as:

$$RPD = \sigma_y / RMSE \quad \text{Eq. 4}$$

where: σ_y is the standard deviation of the observed values.

Both RMSE and NRMSE quantify the prediction accuracy by measuring the deviation of predicted values from observed data. R^2 evaluates the proportion of variance explained by the model, thus indicating its explanatory power. RPD offers insight into the model's predictive capacity relative to the inherent variability of the dataset. Therefore, the combination of these metrics provides a rigorous assessment of model performance.

5.3 Results and discussion

5.3.1 Soil properties and elemental contents

The descriptive statistics for the soil attributes can be found in Table 14. It is notable that all variables exhibited variability. This confirms that the KSSL dataset brings a high diversity of environmental and pedological settings, ranging from sandy to clayey soils, with acidic (pH < 4) to alkaline (pH > 10) profiles. This may pose challenges for the feasibility of developing models that are able to capture this variability on the basis of a limited number of samples. On the other hand, this high variability may lead to a greater generalization capacity for the trained models (RAVIV; LUPYAN; GREEN, 2022).

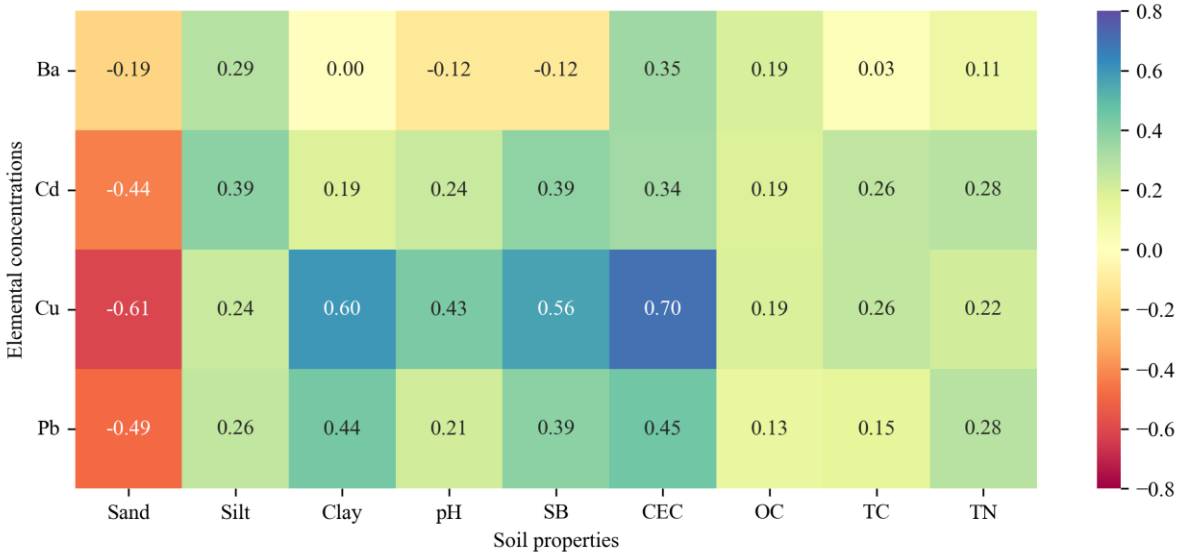
Table 14 - Descriptive statistics for the soil properties from the soil spectral library samples (n = 773).

	Sand (%)	Silt (%)	Clay (%)	pH (H_2O)	SB (cmolc kg^{-1})	CEC (cmolc kg^{-1})	OC (g kg^{-1})	TC (g kg^{-1})	TN (g kg^{-1})
Min.	1.10	1.80	0.23	3.70	0.10	0.61	0.01	0.02	0.00
Median	22.90	43.80	31.19	6.57	22.30	19.66	0.88	1.26	0.10
Mean	29.17	43.27	27.58	6.67	46.44	19.61	1.63	2.03	0.14
Max.	97.40	76.80	76.36	10.42	671.30	50.27	14.22	14.22	0.64
SD	22.44	16.00	16.31	1.16	75.09	11.40	2.02	2.27	0.13
CV	0.77	0.37	0.59	0.17	1.62	0.58	1.24	1.12	0.93

SB sum of bases, CEC cation exchange capacity, OC organic carbon, TC total carbon, TN total nitrogen, SD standard deviation, CV Coefficient of variation

The Spearman correlations (Figure 15) between the soil properties and the elemental concentrations of Ba, Cd, Cu, and Pb, revealed some interesting patterns. All elements are negatively correlated to sand (%), which was expected, since the metals are mostly present in soil bound to the negative charges associated with clay particles. This pattern can be confirmed by observing that all metals are positively correlated with the Cation Exchange Capacity (CEC). However, not all metals are positively correlated with clay (%): Ba exhibited a neutral correlation (0.00). Ba also presented the weakest negative correlation with sand (-0.11) and the only negative correlation with pH (-0.12). The overall correlations of all elements with carbon (OC and TC) were weak (<0.3). All elements exhibited positive correlations with sum of bases (SB), and CEC, with the strongest for Cd and Cu. Among all the elements, Cu exhibited the strongest negative (with sand) and positive (with clay, SB, and CEC) correlations.

Figure 15 - Spearman correlations between the elemental concentrations and soil properties from the soil spectral library.



Considering the elemental concentrations (Table 15), it is notable that the overall dataset is comprised of samples with low concentrations of the selected metals. These low concentrations result from the Mehlich-3 extraction, which is a weak digestion method that leaves a large fraction of these elements non-extracted (MALÝ; ZBÍRAL; ČIŽMÁROVÁ, 2021). However, the contents extracted by the Mehlich-3 method can be considered to represent the bioavailable fraction of these metals in the soil (Hong et al., 2021), thus carrying an environmental relevance worthy of consideration. It is also important to note that there are ongoing studies in regard of the use of Mehlich-3 as a surrogate method for the *aqua regia* extract to screen heavy metal levels in different soils (MALÝ; ZBÍRAL; ČIŽMÁROVÁ, 2021).

Besides *aqua regia*, extraction methods standardized by the United States Environmental Protection Agency (USEPA) are also common in the evaluation of metal levels in soils (SANTOS; ALLEONI, 2013), such as EPA 3050B, EPA 3051, and EPA 3052. In their investigation of soils amended with sewage sludge, Nogueirol et al. (2013) found a strong correlation between *aqua regia*, EPA 3052 and Mehlich-3 extractions for Cu, Fe, Mn, and Zn. Furthermore, Mehlich-3 extractions have also exhibited a good agreement with standard USEPA methods for screening total and bioaccessible Pb contents in soils (MINCA; BASTA; SCHEKEL, 2013). Such findings strengthen the idea that metal contents extracted using Mehlich-3 can be trusted for the purpose of our study, since there are clear correlations

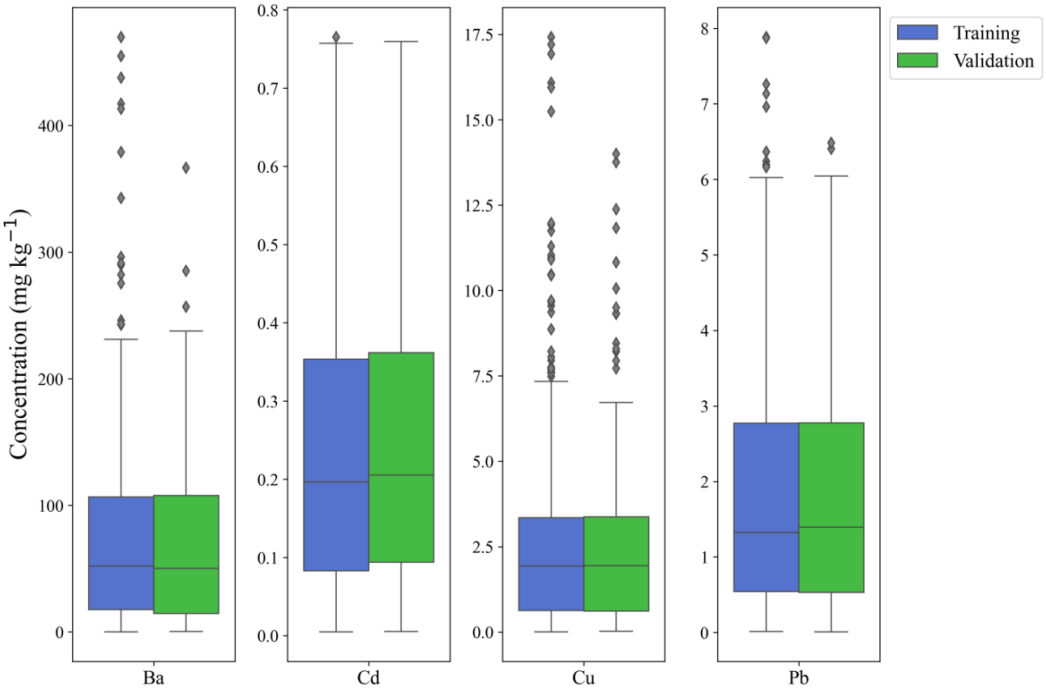
found in the literature for this extraction method with more widely used methods for different elements.

Table 15 - Descriptive statistics of the elemental concentrations from the soil spectral library samples.

Elements	Ba	Cd	Cu	Pb
	mg kg^{-1}			
Minimum	0.20	0.01	0.01	0.01
Median	51.66	0.20	1.94	1.33
Mean	71.07	0.24	2.57	1.76
Maximum	473.59	0.77	18.09	8.01
Standard deviation	69.45	0.18	2.72	1.50

With the objective of visualizing the differences in the distributions of the elemental concentrations between the training and validation sets, boxplots were generated for each element (Figure 16). Considering both training and validation sets and all elements, a clear asymmetrical distribution pattern can be observed in which the majority of the samples present low concentrations. However, both training (n = 618) and validation (n = 155) sets are representative of the total distribution of all elements, despite a few outliers with higher concentrations present in the former. Thus, the split between training and validation sets can be considered adequate for the development of the prediction models.

Figure 16 - Distributions of elements for the training (n = 618) and validation (n = 155) sets.



5.3.2 Modeling results

The results of the prediction models for elemental concentrations from MIR spectra demonstrated varying performances across different elements and preprocessing methods (Table 16). Firstly, it is important to examine the outcomes of the PLS regression, which is considered a baseline approach in the context of the work herein. The best PLS models exhibited either R^2 values <0.50 or RPD values <1.4. These findings indicate that all PLS models in this study demonstrated poor quality, as the evaluation metrics fell below the established thresholds for considering regression models acceptable in the context of soil infrared spectroscopy (WANG; ZHANG; PAN, 2017). Such results underscore the difficulty associated with predicting low concentration heavy metals from mid-infrared spectra by means of traditional regression approaches such as PLS.

On the other hand, the outcomes of the CNN prediction (Table 16) demonstrated acceptable results for all elements, even when using raw spectra. The prediction using raw spectral data performed reasonably well, particularly for Cd (R^2 = 0.81 and RPD = 2.32). Considering Ba, Cu, and Pb, the results with raw spectra can be considered acceptable (R^2 > 0.7 and RPD > 1.8).

Considering the performance of CNN using preprocessed spectra, SG yielded better results for Cu (R^2 = 0.78 and RPD = 2.15) compared to the raw data, indicating that the SG filter enhances spectral features that were better captured by the CNN for this element. However, for Ba, Cd, and Pb, the evaluation metrics were lower using SG. This may indicate that the process of smoothing the spectra leads to a loss of relevant spectral information related to the contents of these elements. Meanwhile, SNV underperformed for all elements. This result is probably a consequence of the normalization promoted by the SNV, which may have also caused loss of relevant spectral information for the CNN algorithm.

Overall, the results suggest that CNN models are much more effective at predicting elemental concentrations from MIR spectra than the PLS model. This is especially the case for Cd Meanwhile, for Ba, Cu, and Pb, the models exhibited acceptable predictive capabilities, indicating that there is potential for the development of robust prediction models for these elements using CNN.

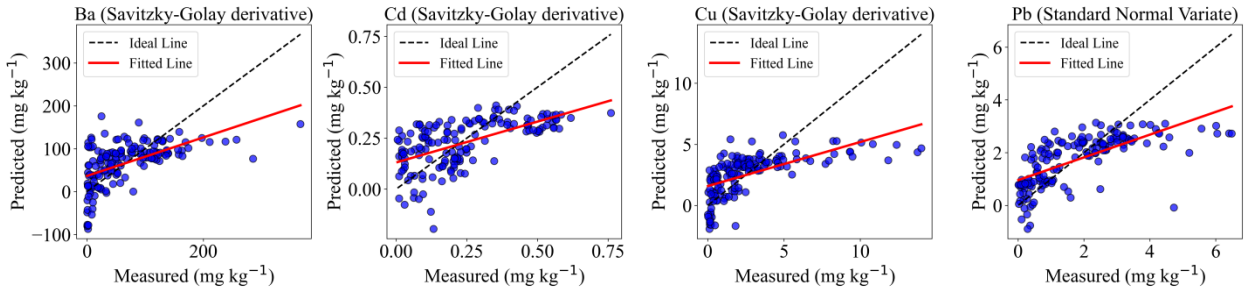
Table 16 - Evaluation metrics for all modeling results using the validation set (n = 155).

Model	Preprocessing	Metric	Ba	Cd	Cu	Pb
PLS	Raw Spectra	RMSE	57.22	0.15	2.43	1.20
		NRMSE	0.86	0.64	0.91	0.67
		R^2	0.16	0.24	0.27	0.35
		RPD	1.09	1.15	1.17	1.25
	Savitzky-Golay derivative	RMSE	54.68	0.14	2.27	1.10
		NRMSE	0.82	0.59	0.85	0.61
		R^2	0.23	0.34	0.37	0.45
		RPD	1.14	1.24	1.26	1.35
	Standard Normal Variate	RMSE	57.15	0.15	2.34	1.10
		NRMSE	0.86	0.62	0.87	0.61
		R^2	0.16	0.28	0.33	0.46
		RPD	1.09	1.18	1.22	1.36
CNN	Raw Spectra	RMSE	33.44	0.08	1.46	0.71
		NRMSE	0.50	0.32	0.54	0.40
		R^2	0.71	0.81	0.74	0.77
		RPD	1.87	2.32	1.96	2.13
	Savitzky-Golay derivative	RMSE	35.96	0.11	1.33	0.78
		NRMSE	0.54	0.46	0.49	0.43
		R^2	0.67	0.61	0.78	0.73
		RPD	1.76	2.16	2.15	2.02
	Standard Normal Variate	RMSE	37.37	0.10	1.62	0.92
		NRMSE	0.56	0.42	0.60	0.52
		R^2	0.64	0.66	0.68	0.61
		RPD	1.68	1.88	1.79	1.65

PLS Partial least squares, CNN Convolutional neural network, RAW Raw spectra, SG Savitzky-golay derivative, SNV Standard normal variate, RMSE Root mean squared error, NRMSE Normalized root mean squared error, R^2 Coefficient of determination, RPD Ratio of prediction to deviation

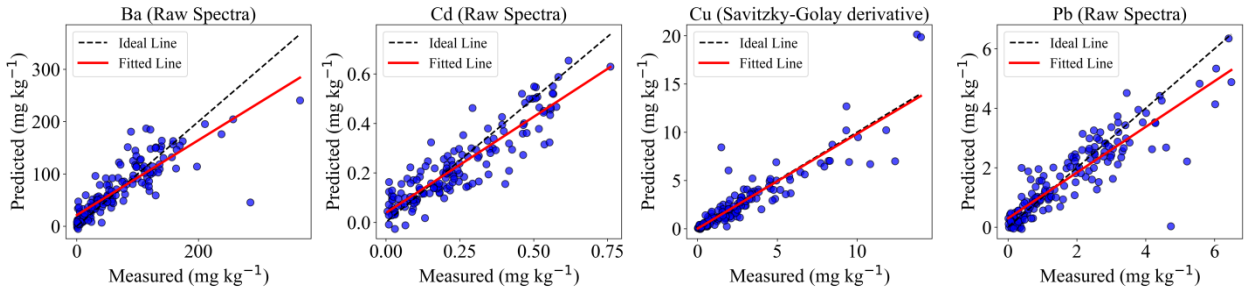
The predicted *versus* measured results can be visualized for both PLS (Figure 17) and CNN (Figure 18). The PLS model, despite its ability to manage high-dimensional datasets with multicollinearity (GARG; TAI, 2013), exhibited poor prediction performances, a pattern that can be clearly observed when comparing the predicted values to the actual measured values. These predictions deviated significantly from the observed data, presenting a sharp underestimation for all elements. Such results highlight the limitations of PLS in capturing the complex, non-linear relationships inherent in the data. Thus, it can be noted that PLS, with its linear assumptions, struggles to generalize the intricate patterns present in this highly variable and heterogeneous dataset, leading to its underperformance.

Figure 17 - Partial Least Squares (PLS) prediction results for the validation set (n = 155) with the best preprocessing for each element from the soil spectral library.



The CNN model presented an evidently better prediction performance than PLS (Figure 18), as its predicted values closely align with the measured values. It appears that the ability to capture non-linear relationships through convolutional layers effectively represented the underlying complexities within the data. It may be assumed that the CNN architecture allowed for the models to learn abstract features related to the elemental concentrations within the MIR spectral data, leading to a higher level of generalization and precision. Therefore, it is possible to recommend that CNN can be a powerful tool for the prediction of heavy metal elemental concentrations from MIR data in the context of a diverse and heterogeneous soil spectral library.

Figure 18 - Convolutional Neural Network (CNN) prediction results for the validation set (n = 155) with the best preprocessing for each element from the soil spectral library.



Our results for Ba, Cd, Cu, and Pb models are promising. All of these elements are problematic in regard of environmental quality. Barium (Ba) poses risks to animals and plants, with its exposure being related to multiple deleterious effects (Lu et al., 2019). Cadmium (Cd) is a potentially toxic element of high environmental importance (CUNHA; NASCIMENTO, 2009; KHAN et al., 2010; KHANAM et al., 2020; ULLAH et al., 2021), ranked as the 11th most relevant substance in the list of priority pollutants (USEPA, 2022).

Copper (Cu), which is a nutrient for plants and animals, is also a known soil pollutant (HAQUE; SUBRAMARIAN; GIBBS, 1982) that generates several negative impacts when excessively present in soil (XIONG et al., 2010). Lead (Pb) is also present in the list of priority pollutants (USEPA, 2022), with its presence in soil leading to a diverse set of problems (DUDKA; MILLER, 1999; FELEAFEL; MIRDAD, 2012), and its chronic exposure possibly leading to fatal outcomes (MAGRISSO; BELKIN; EREL, 2009). In this context, the prospect of screening Ba, Cd, Cu, and Pb levels in soil by means of a quick and cost-effective method, such as a MIR-based monitoring system, may greatly improve the environmental control and management of these elements.

5.4 Conclusions

The aim of this work was to build models for the prediction of Ba, Cd, Cu, and Pb concentrations from MIR spectral signatures. The dataset consisted of highly diverse and heterogeneous soil samples from North America made available by the Kellogg Soil Survey Laboratory (KSSL) in the Open Soil Spectral Library (OSSL). Both Partial Least Squares (PLS) regression and a Convolutional Neural Network (CNN) with depthwise separable convolution architecture were used in combination with raw and preprocessed MIR spectra. CNN produced predictions that closely matched the measured values. In contrast, the PLS model showed an overall underestimation for the predicted values, indicating less accurate predictions. The prediction performances achieved with CNN for , Ba, Cu, and Pb suggest the possibility of developing a MIR-based monitoring system capable of screening levels of soil contamination by these metals in a quick and cost-effective manner.

Despite the promising results, some limitations need to be underscored to provide guidance for future research. One notable limitation arises from the lack of information regarding the geological and pedological settings in which the samples were collected, leaving a gap in our detailed understanding of the environmental context. It is also noteworthy that the low concentrations of the studied metals, a main limitation in this study, are directly attributed to the Mehlich-3 extraction method employed. The *aqua regia* extraction method is the standard for screening heavy metal levels in soil in the context of environmental legislation. Hence, improved availability of open soil spectral data with metal contents extracted by *aqua regia* would significantly enhance the assessment and reproducibility of similar approaches.

Nevertheless, we recommend further research in order to explore the response of these elements in a broader range of environmental settings. Increasing the number of samples would also contribute to more insightful findings for similar research. It is important to

emphasize that data from the Open Soil Spectral Library played a pivotal role in this work, and similar initiatives are essential for advancing scientific development in this field. Notably, data containing concentrations of potentially toxic elements are particularly scarce within currently available soil spectral libraries. Ongoing research on soil infrared spectroscopy would greatly benefit from an augmentation of such specific data resources.

References

ALENGEBAWY, A.; ABDELKHALEK, S. T.; QURESHI, S. R.; WANG, M. Q. Heavy metals and pesticides toxicity in agricultural soil and plants: Ecological risks and human health implications. **Toxics**, v. 9, n. 3, p. 42, 2021. https://doi.org/10.3390/toxics9030042

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. **Applied Spectroscopy**, v. 43, n. 5, p. 772–777, 1989. https://doi.org/10.1366/0003702894202201

CAPUANO, E.; VAN RUTH, S. M. Infrared spectroscopy: applications. In: **Encyclopedia of Food and Health**, pp. 424–431. Elsevier Inc. Academic Press, 2015. https://doi.org/10.1016/B978-0-12-384947-2.00644-9

CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE). **Geoscientific Model Development Discussions**, v. 7, n. 1, p. 1525–1534, 2014. https://doi.org/10.5194/gmdd-7-1525-2014

CHAKRAVARTULA, S. S. N.; MOSCETTI, R.; BEDINI, G.; NARDELLA, M.; MASSANTINI, R. Use of convolutional neural network, CNN combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee. **Food Control**, v. 135, p. 108816, 2022. https://doi.org/10.1016/j.foodcont.2022.108816

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, pp. 1251–1258, 2017. https://doi.org/10.1109/CVPR.2017.195

CONAMA. Conselho Nacional do Meio Ambiente. Resolução CONAMA nº 420, de 28 de dezembro de 2009. **Diário Oficial da União**, Brasília, BRASIL, 28 dez. 2009.

COZZOLINO, D. Near infrared spectroscopy as a tool to monitor contaminants in soil, sediments and water—State of the art, advantages and pitfalls. **Trends in Environmental Analytical Chemistry**, v. 9, p. 1–7, 2016. https://doi.org/10.1016/j.teac.2015.10.001

CUNHA, K. P. V.; NASCIMENTO, C. W. A. Silicon effects on metal tolerance and structural changes in maize, Zea mays L. grown on a cadmium and zinc enriched soil. **Water, Air, and Soil Pollution**, v. 197, p. 323–330, 2009. https://doi.org/10.1007/s11270-008-9814-9

DEMATTÊ, J. A. M.; DOTTO, A. C.; BEDIN, L. G.; SAYÃO, V. M.; SOUZA, A. B. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. **Geoderma**, v. 337, p. 111–121, 2019a. https://doi.org/10.1016/j.geoderma.2018.09.010

DEMATTÊ, J. A. M.; DOTTO, A. C.; PAIVA, A. F. S.; SATO, M. V.; DALMOLIN, R. S. D.; ARAÚJO, M. S. B. de; SILVA, E. B. da; NANNI, M. R.; CATEN, A. ten; NORONHA, N. C.; LACERDA, M. P. C.; ARAÚJO FILHO, J. C. de; RIZZO, R.; BELLINASO, H.; FRANCELINO, M. R.; SCHAEFER, C. E. G. R.; VICENTE, L. E.; DOS SANTOS, U. J.; SÁ BARRETTO SAMPAIO, E. V. de; MENEZES, R. S. C.; DE SOUZA, J. J. L. L.; ABRAHÃO, W. A. P.; COELHO, R. M.; GREGO, C. R.; LANI, J. L.; FERNANDES, A. R.; GONÇALVES, D. A. M.; SILVA, S. H. G.; MENEZES, M. D. de; CURI, N.; COUTO, E. G.; DOS ANJOS, L. H. C.; CEDDIA, M. B.; PINHEIRO, É. F. M.; GRUNWALD, S.; VASQUES, G. M.; MARQUES JÚNIOR, J.; DA SILVA, A. J.; BARRETO, M. C. de V.; NÓBREGA, G. N.; DA SILVA, M. Z.; DE SOUZA, S. F.; VALLADARES, G. S.; VIANA, J. H. M.; TERRA, F. da S.; HORÁK-TERRA, I.; FIORIO, P. R.; DA SILVA, R. C.; FRADE JÚNIOR, E. F.; LIMA, R. H. C.; FILIPPINI ALBA, J. M.; DE SOUZA JUNIOR, V. S.; SANTOS BREFIN, M. D. L. M.; RUIVO, M. D. L. P.; FERREIRA, T. O.; BRAIT, M. A.; CAETANO, N. R.; BRINGHENTI, I.; MENDES, W. de S.; SAFANELLI, J. L.; GUIMARÃES, C. C. B.; POPPIEL, R. R.; BARROS E SOUZA, A.; QUESADA, C. A.; COUTO, H. T. Z. The Brazilian soil spectral library, BSSL: A general view, application and challenges. **Geoderma**, v. 354, p. 113793, 2019b. https://doi.org/10.1016/j.geoderma.2019.05.043

DUDKA, S.; MILLER, W. P. Permissible Concentrations of Arsenic and Lead in Soils Based on Risk Assessment. **Water, Air, & Soil Pollution**, v. 113, p. 127–132, 1999. https://doi.org/10.1023/A:1005028905396

ELANGO, D.; DEVI, K. D.; JEYABALAKRISHNAN, H. K.; RAJENDRAN, K.; HARIDASS, V. K. T.; DHARMARAJ, D.; CHARUCHANDRAN, C. V.; WANG, W.; FAKUDE, M.; MISHRA, R.; VEMBU, K.; WANG, X. Agronomic, breeding, and biotechnological interventions to mitigate heavy metal toxicity problems in agriculture. **Journal of Agriculture and Food Research**, v. 10, p. 100374, 2022. https://doi.org/10.1016/j.jafr.2022.100374

FALEH, R.; KACHOURI, A. A hybrid deep convolutional neural network-based electronic nose for pollution detection purposes. **Chemometrics and Intelligent Laboratory Systems**, v. 237, p. 104825, 2023. https://doi.org/10.1016/j.chemolab.2023.104825

FELEAFEL, M. N.; MIRDAD, Z. M. Hazard and Effects of Pollution by Lead on Vegetable Crops. **Journal of Agricultural and Environmental Ethics**, v. 26, p. 547–567, 2013. https://doi.org/10.1007/s10806-012-9403-1

GARG, A.; TAI, K. Comparison of statistical and machine learning methods in modelling of data with multicollinearity. **International Journal of Modelling, Identification and Control**, v. 18, n. 4, p. 295–312, 2013. https://doi.org/10.1504/IJMIC.2013.053535

GARTHWAITE, P. H. An interpretation of partial least squares. **Journal of the American Statistical Association**, v. 89, n. 425, p. 122–127, 1994.

HAQUE, M. A.; SUBRAMANIAN, V.; GIBBS, R. J. Copper, lead, and zinc pollution of soil environment. **Critical Reviews in Environmental Control**, v. 12, n. 1, p. 13–68, 1982. https://doi.org/10.1080/10643388209381693

HORTA, A.; MALONE, B.; STOCKMANN, U.; MINASNY, B.; BISHOP, T. F. A.; MCBRATNEY, A. B.; PALLASSER, R.; POZZA, L. Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. **Geoderma**, v. 241, p. 180–209, 2015. https://doi.org/10.1016/j.geoderma.2014.11.024

HONG, Y. K.; YOON, D. H.; KIM, J. W.; CHAE, M. J.; KO, B. K.; KIM, S. C. Ecological risk assessment of heavy metal-contaminated soil using the triad approach. **Journal of Soils and Sediments**, v. 21, p. 2732–2743, 2021. https://doi.org/10.1007/s11368-020-02750-9

JANIK, L. J.; SKJEMSTAD, J. O.; SHEPHERD, K. D.; SPOUNCER, L. R. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. **Soil Research**, v. 45, n. 2, p. 73–81, 2007. https://doi.org/10.1071/SR06083

JANIK, L. J.; SORIANO-DISLA, J. M.; FORRESTER, S. T. Feasibility of handheld mid-infrared spectroscopy to predict particle size distribution: influence of soil field condition and utilisation of existing spectral libraries. **Soil Research**, v. 58, n. 6, p. 528–539, 2020. https://doi.org/10.1071/SR20097

KÄSTNER, F.; SUT-LOHMANN, M.; RAMEZANY, S.; RAAB, T.; FEILHAUER, H.; CHABRILLAT, S. Estimating heavy metal concentrations in Technosols with reflectance spectroscopy. **Geoderma**, v. 406, p. 115512, 2022. https://doi.org/10.1016/j.geoderma.2021.115512

KINGMA, D. P.; BA, J. Adam: A Method for Stochastic Optimization. In: Proceedings of the International Conference on Learning Representations, May 7–9, 2015, San Diego, CA, USA, 2015.

KHAN, S.; EL-LATIF HESHAM, A.; QIAO, M.; REHMAN, S.; HE, J. Effects of Cd and Pb on soil microbial community structure and activities. **Environmental Science and Pollution Research**, v. 17, p. 288–296, 2010. https://doi.org/10.1007/s11356-009-0134-4

KHANAM, R.; KUMAR, A.; NAYAK, A. K.; SHAHID, M.; TRIPATHI, R.; VIJAYAKUMAR, S.; BHADURI, D.; KUMAR, U.; MOHANTY, S.; PANNEERSELVAM, R.; CHATTERJEE, D.; SATAPATHY, B. S.; PATHAK, H. Metal(loid)s, As, Hg, Se, Pb and Cd in paddy soil: Bioavailability and potential risk to human health. **Science of the Total Environment**, v. 699, p. 134330, 2020. https://doi.org/10.1016/j.scitotenv.2019.134330

KUMAR, V.; SHARMA, A.; KAUR, P.; SIDHU, G. P. S.; BALI, A. S.; BHARDWAJ, R.; THUKRAL, A. K.; CERDA, A. Pollution assessment of heavy metals in soils of India and ecological risk assessment: A state-of-the-art. **Chemosphere**, v. 216, p. 449–462, 2019. https://doi.org/10.1016/j.chemosphere.2018.10.066

LECUN, Y.; BENGIO, Y. Convolutional networks for images, speech, and time series. In: **The Handbook of Brain Theory and Neural Networks**, v. 3361, 1995.

LI, Z.; MA, Z.; VAN DER KUIJP, T. J.; YUAN, Z.; HUANG, L. A review of soil heavy metal pollution from mines in China: pollution and health risk assessment. **Science of the Total Environment**, v. 468, p. 843–853, 2014. https://doi.org/10.1016/j.scitotenv.2013.08.090

LI, M.; FENG, Y.; YU, Y.; ZHANG, T.; YAN, C.; TANG, H.; SHENG, Q.; LI, H. Quantitative analysis of polycyclic aromatic hydrocarbons in soil by infrared spectroscopy combined with hybrid variable selection strategy and partial least squares. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 257, p. 119771, 2021. https://doi.org/10.1016/j.saa.2021.119771

LIU, J.; ZHANG, Y.; WANG, H.; DU, Y. Study on the prediction of soil heavy metal elements content based on visible near-infrared spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 199, p. 43–49, 2018. https://doi.org/10.1016/j.saa.2018.03.040

LIU, J.; HAN, J.; XIE, J.; WANG, H.; TONG, W.; BA, Y. Assessing heavy metal concentrations in earth-cumulic-orthic-anthrosols soils using Vis-NIR spectroscopy transform coupled with chemometrics. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 226, p. 117639, 2020. https://doi.org/10.1016/j.saa.2019.117639

LU, Q.; XU, X.; LIANG, L.; XU, Z.; SHANG, L.; GUO, J.; XIAO, D.; QIU, G. Barium concentration, phytoavailability, and risk assessment in soil-rice systems from an active barium mining region. **Applied Geochemistry**, v. 106, p. 142–148, 2019. https://doi.org/10.1016/j.apgeochem.2019.05.010

LUCE, M. S.; ZIADI, N.; ROSSEL, R. A. V. GLOBAL-LOCAL: A new approach for local predictions of soil organic carbon content using large soil spectral libraries. **Geoderma**, v. 425, p. 116048, 2022. https://doi.org/10.1016/j.geoderma.2022.116048

MA, Y.; ROUDIER, P.; KUMAR, K.; PALMADA, T.; GREALISH, G.; CARRICK, S.; LILBURNE, L.; TRIANTAFILIS, J. A soil spectral library of New Zealand. **Geoderma Regional**, v. 35, e00726, 2023. https://doi.org/10.1016/j.geodrs.2023.e00726

MAGRISSO, S.; BELKIN, S.; EREL, Y. Lead bioavailability in soil and soil components. **Water, Air, and Soil Pollution**, v. 202, p. 315–323, 2009. https://doi.org/10.1007/s11270-009-9978-y

MAIA, A. J.; NASCIMENTO, R. C.; SILVA, Y. J. A. B.; DO NASCIMENTO, C. W. A.; MENDES, W. S.; NETO, J. G. V.; FILHO, J. C. A.; TIECHER, T.; SILVA, Y. J. A. B. Near-infrared spectroscopy for prediction of potentially toxic elements in soil and sediments from a semiarid and coastal humid tropical transitional river basin. **Microchemical Journal**, v. 179, 107544, 2022. https://doi.org/10.1016/j.microc.2022.107544

MALÝ, S.; ZBÍRAL, J.; ČIŽMÁROVÁ, E. Is Mehlich 3 soil extraction a suitable screening method for determination of some risk elements?. **Plant, Soil and Environment**, v. 67, n. 9, p. 499–506, 2021. https://doi.org/10.17221/228/2021-PSE

MINCA, K. K.; BASTA, N. T.; SCHECKEL, K. G. Using the Mehlich-3 soil test as an inexpensive screening tool to estimate total and bioaccessible lead in urban soils. **Journal of Environmental Quality**, v. 42, n. 5, p. 1518–1526, 2013. https://doi.org/10.2134/jeq2012.0450

MOTUZOVA, G. V.; MINKINA, T. M.; KARPOVA, E. A.; BARSOVA, N. U.; MANDZHIEVA, S. S. Soil contamination with heavy metals as a potential and real risk to the environment. **Journal of Geochemical Exploration**, v. 144, p. 241–246, 2014. https://doi.org/10.1016/j.gexplo.2014.01.026

MOURA-BUENO, J. M.; DALMOLIN, R. S. D.; TEN CATEN, A.; DOTTO, A. C.; DEMATTÊ, J. A. Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. **Geoderma**, v. 337, p. 565–581, 2019. https://doi.org/10.1016/j.geoderma.2018.10.015

NG, W.; MINASNY, B.; MONTAZEROLGHAEM, M.; PADARIAN, J.; FERGUSON, R.; BAILEY, S.; MCBRATNEY, A. B. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. **Geoderma**, v. 352, p. 251–267, 2019. https://doi.org/10.1016/j.geoderma.2019.06.016

NOGUEIROL, R. C.; DE MELO, W. J.; BERTONCINI, E. I.; ALLEONI, L. R. F. Concentrations of Cu, Fe, Mn, and Zn in tropical soils amended with sewage sludge and composted sewage sludge. **Environmental Monitoring and Assessment**, v. 185, p. 2929–2938, 2013. https://doi.org/10.1007/s10661-012-2761-3

NYARKO, F.; TACK, F. M.; MOUAZEN, A. M. Potential of visible and near infrared spectroscopy coupled with machine learning for predicting soil metal concentrations at the regional scale. **Science of The Total Environment**, v. 841, 156582, 2022. https://doi.org/10.1016/j.scitotenv.2022.156582

ORGIAZZI, A.; BALLABIO, C.; PANAGOS, P.; JONES, A.; FERNÁNDEZ-UGALDE, O. LUCAS Soil, the largest expandable soil dataset for Europe: a review. **European Journal of Soil Science**, v. 69, n. 1, p. 140–153, 2018. https://doi.org/10.1111/ejss.12499

PANDEY, A.; JAIN, K. A robust deep attention dense convolutional neural network for plant leaf disease identification and classification from smart phone captured real world images. **Ecological Informatics**, v. 70, 101725, 2022. https://doi.org/10.1016/j.ecoinf.2022.101725

PANDIT, C. M.; FILIPPELLI, G. M.; LI, L. Estimation of heavy-metal contamination in soil using reflectance spectroscopy and partial least-squares regression. **International Journal of Remote Sensing**, v. 31, n. 15, p. 4111–4123, 2010. https://doi.org/10.1080/01431160903229200

PADARIAN, J.; MINASNY, B.; MCBRATNEY, A. B. Using deep learning to predict soil properties from regional spectral data. **Geoderma Regional**, v. 16, e00198, 2019. https://doi.org/10.1016/j.geodrs.2018.e00198

PYO, J.; HONG, S. M.; KWON, Y. S.; KIM, M. S.; CHO, K. H. Estimation of heavy metals using deep neural network with visible and infrared spectroscopy of soil. **Science of the Total Environment**, v. 741, 140162, 2020. https://doi.org/10.1016/j.scitotenv.2020.140162

RAVIV, L.; LUPYAN, G.; GREEN, S. C. How variability shapes learning and generalization. **Trends in Cognitive Sciences**, v. 26, n. 6, p. 462–483, 2022. https://doi.org/10.1016/j.tics.2022.03.007

REDDI, S. J.; KALE, S.; KUMAR, S. On the convergence of Adam and beyond. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, Apr 30–May 3, Vancouver, CA, 2018.

REN, S.; SONG, C.; YE, S.; CHENG, C.; GAO, P. The spatiotemporal variation in heavy metals in China's farmland soil over the past 20 years: a meta-analysis. **Science of The Total Environment**, v. 806, 150322, 2022. https://doi.org/10.1016/j.scitotenv.2021.150322

SAFANELLI, J. L.; HENGL, T.; SANDERMAN, J.; PARENTE, L. Open Soil Spectral Library, training data and calibration models, v1.2 [Data set]. **Zenodo**, 2021. https://doi.org/10.5281/zenodo.7599269

SANDERMAN, J.; SAVAGE, K.; DANGAL, S. R. Mid-infrared spectroscopy for prediction of soil health indicators in the United States. **Soil Science Society of America Journal**, v. 84, n. 1, p. 251–261, 2020. https://doi.org/10.1002/saj2.20009

SANTOS, S. N.; ALLEONI, L. R. F. Methods for extracting heavy metals in soils from the southwestern Amazon, Brazil. **Water, Air, Soil Pollution**, v. 224, p. 1–16, 2013. https://doi.org/10.1007/s11270-012-1430-z

SANTANA, F. B.; OTANI, S. K.; DE SOUZA, A. M.; POPPI, R. J. Comparison of PLS and SVM models for soil organic matter and particle size using vis-NIR spectral libraries. **Geoderma Regional**, v. 27, e00436, 2021. https://doi.org/10.1016/j.geodrs.2021.e00436

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, 1964. https://doi.org/10.1021/ac60214a047

SHANG, H.; SHANG, L.; WU, J.; XU, Z.; ZHOU, S.; WANG, Z.; WANG, H.; YIN, J. NIR spectroscopy combined with 1D-convolutional neural network for breast cancerization analysis and diagnosis. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 287, 121990, 2023. https://doi.org/10.1016/j.saa.2022.121990

SHEPHERD, K. D.; FERGUSON, R.; HOOVER, D.; VAN EGMOND, F.; SANDERMAN, J.; GE, Y. A global soil spectral calibration library and estimation service. **Soil Security**, v. 7, 100061, 2022. https://doi.org/10.1016/j.soisec.2022.100061

SHI, T.; CHEN, Y.; LIU, Y.; WU, G. Visible and near-infrared reflectance spectroscopy—An alternative for monitoring soil contamination by heavy metals. **Journal of Hazardous Materials**, v. 265, p. 166–176, 2014. https://doi.org/10.1016/j.jhazmat.2013.11.059

SINGH, A.; PRASAD, S. M. Remediation of heavy metal contaminated ecosystem: an overview on technology advancement. **International Journal of Environmental Science and Technology**, v. 12, p. 353–366, 2015. https://doi.org/10.1007/s13762-014-0542-y

SOIL SURVEY STAFF. Kellogg Soil Survey Laboratory Methods Manual. **Soil Survey Investigations Report No. 42**, Version 6.0. National Soil Survey Center, Natural Resources Conservation Service, USDA, Lincoln, Nebraska, 2022.

SONG, Y.; LI, F.; YANG, Z.; AYOKO, G. A.; FROST, R. L.; JI, J. Diffuse reflectance spectroscopy for monitoring potentially toxic elements in the agricultural soils of Changjiang River Delta, China. **Applied Clay Science**, v. 64, p. 75–83, 2012. https://doi.org/10.1016/j.clay.2011.09.010

THAKUR, P. S.; CHATURVEDI, S.; KHANNA, P.; SHEOREY, T.; OJHA, A. Vision transformer meets convolutional neural network for plant disease classification. **Ecological Informatics**, v. 77, 102245, 2023. https://doi.org/10.1016/j.ecoinf.2023.102245

TODOROVA, M.; MOUAZEN, A. M.; LANGE, H.; ATANASSOVA, S. Potential of near-infrared spectroscopy for measurement of heavy metals in soil as affected by calibration set size. **Water, Air, & Soil Pollution**, v. 225, p. 1–19, 2014. https://doi.org/10.1007/s11270-014-2036-4

UDELHOVEN, T.; EMMERLING, C.; JARMER, T. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. **Plant and Soil**, v. 251, p. 319–329, 2003. https://doi.org/10.1023/A:1023008322682

ULLAH, A.; TAHIR, A.; RASHID, H. U.; UR REHMAN, T.; DANISH, S.; HUSSAIN, B.; AKCA, H. Strategies for reducing Cd concentration in paddy soil for rice safety. **Journal of Cleaner Production**, v. 316, 128116, 2021. https://doi.org/10.1016/j.jclepro.2021.128116

USEPA. **40 CFR 401.15. U.S. Environmental Protection Agency**. Retrieved December 18, 2024, from https://www.ecfr.gov/current/title-40/part-401/section-401.15, 2022.

VISCARRA ROSSEL, R.; WALVOORT, D. J. J.; MCBRATNEY, A. B.; JANIK, L. J.; SKJEMSTAD, J. O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. **Geoderma**, v. 131, p. 59–75, 2006. https://doi.org/10.1016/j.geoderma.2005.03.007

WANG, J.; CUI, L.; GAO, W.; SHI, T.; CHEN, Y.; GAO, Y. Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy. **Geoderma**, v. 216, p. 1–9, 2014. https://doi.org/10.1016/j.geoderma.2013.10.024

WANG, C.; ZHANG, T.; PAN, X. Potential of visible and near-infrared reflectance spectroscopy for the determination of rare earth elements in soil. **Geoderma**, v. 306, p. 120–126, 2017. https://doi.org/10.1016/j.geoderma.2017.07.016

WANG, S.; GUAN, K.; ZHANG, C.; LEE, D.; MARGENOT, A. J.; GE, Y.; PENG, J.; ZHOU, W.; ZHOU, Q.; HUANG, Y. Using soil library hyperspectral reflectance and machine learning to predict soil organic carbon: Assessing potential of airborne and spaceborne optical soil sensing. **Remote Sensing of Environment**, v. 271, 112914, 2022a. https://doi.org/10.1016/j.rse.2022.112914

WANG, Z.; LUO, P.; ZHA, X.; XU, C.; KANG, S.; ZHOU, M.; NOVER, D.; WANG, Y. Overview assessment of risk evaluation and treatment technologies for heavy metal pollution of water and soil. **Journal of Cleaner Production**, v. 134043, 2022b. https://doi.org/10.1016/j.jclepro.2022.134043

WIJEWARDANE, N. K.; GE, Y.; WILLS, S.; LOECKE, T. Prediction of soil carbon in the conterminous United States: visible and near infrared reflectance spectroscopy analysis of the rapid carbon assessment project. **Soil Science Society of America Journal**, v. 80, n. 4, p. 973–982, 2016.

XIANG, M.; LI, Y.; YANG, J.; LEI, K.; LI, Y.; LI, F.; ZHENG, D.; FANG, X.; CAO, Y. Heavy metal contamination risk assessment and correlation analysis of heavy metal contents in soil and crops. **Environmental Pollution**, v. 278, 116911, 2021. https://doi.org/10.1016/j.envpol.2021.116911

XIONG, X.; YANXIA, L.; WEI, L.; CHUNYE, L.; WEI, H.; MING, Y. Copper content in animal manures and potential risk of soil copper pollution with animal manure use in agriculture. **Resources, Conservation and Recycling**, v. 54, n. 11, p. 985–990, 2010. https://doi.org/10.1016/j.resconrec.2010.02.00

YANG, J.; WANG, X.; WANG, R.; WANG, H. Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using Vis–NIR spectroscopy. **Geoderma**, v. 380, 114616, 2020. https://doi.org/10.1016/j.geoderma.2020.114616

ZHANG, D. A coefficient of determination for generalized linear models. **The American Statistician**, v. 71, n. 4, p. 310–316, 2017. https://doi.org/10.1080/00031305.2016.1256839

ZHANG, X.; SUN, W.; CEN, Y.; ZHANG, L.; WANG, N. Predicting cadmium concentration in soils using laboratory and field reflectance spectroscopy. **Science of the Total Environment**, v. 650, p. 321–334, 2019. https://doi.org/10.1016/j.scitotenv.2018.08.442

ZHAO, D.; ZHAO, X.; KHONGNAWANG, T.; ARSHAD, M.; TRIANTAFILIS, J. A Vis-NIR Spectral Library to Predict Clay in Australian Cotton Growing Soil. **Soil Science Society of America Journal**, v. 82, n. 6, p. 1347–1357, 2018. https://doi.org/10.2136/sssaj2018.03.0100

ZHAO, D.; WANG, J.; JIANG, X.; ZHEN, J.; MIAO, J.; WANG, J.; WU, G. Reflectance spectroscopy for assessing heavy metal pollution indices in mangrove sediments using XGBoost method and physicochemical properties. **Catena**, v. 211, 105967, 2022. https://doi.org/10.1016/j.catena.2021.105967

6 FINAL CONSIDERATIONS

This study aimed to evaluate the potential of infrared spectroscopy (IRS) as an effective tool for monitoring soil degradation across diverse conditions, with a particular focus on concentrations of potentially toxic elements (PTEs), including results for soil erodibility and salinity-related properties. The findings strongly support the initial hypotheses and position IRS, especially when enhanced by modern data modeling techniques, as a powerful ally for environmental monitoring and sustainable land management.

Across the literature review integrated with case studies, IRS was found to demonstrate its ability to estimate soil degradation-related attributes with moderate to high accuracy. It was found that the integration of machine learning modeling approaches significantly enhances the reliability of the results, as exemplified by the predictions of soil erodibility and PTEs. Partial Least Squares Regression (PLSR) models, while effective in some cases, shows limitations when dealing with more complex or less spectrally distinctive attributes, underscoring the need for more advanced modeling strategies.

The approach of combining Convolutional Neural Networks (CNNs) with spectral data fusion techniques, integrating both near-infrared (NIR) and mid-infrared (MIR) spectra, was assessed in different scenarios. The results confirm this premise: CNNs outperformed conventional chemometric approaches in predicting salinity-related variables and PTEs concentrations in most cases, especially with MIR, and for some cases using fused spectral inputs. Both CNN-MIR and CNN-fusion combinations yielded higher R^2 and RPD values compared to PLSR across multiple case studies and environmental conditions, demonstrating robustness and scalability. However, in some cases, PLSR still outperformed CNN, demonstrating the inherent value of this classical approach.

Beyond confirming the hypotheses, the research highlights the broader evolution of IRS from an experimental laboratory-based technique to a flexible, field-deployable solution. Supported by advances in hardware, spectral libraries, and open-access datasets, IRS is now positioned to offer real-time, high-resolution soil assessments. Although challenges remain, such as the need for regional calibration models and handling low-concentration analytes, these are being addressed by ongoing developments in spectral libraries, machine learning, data fusion, and standardization efforts. In conclusion, this work strongly reinforces IRS as a pragmatic tool for soil monitoring, contributing to global efforts in combating soil degradation, enhancing food security, and promoting environmental resilience.

APPENDIX A – Convolutional Neural Network architecture

Table 1 - Architecture of the Convolutional Neural Network with depthwise separable convolutions.

Layer	Number of Filters	Kernel size	Stride	Activation
SeparableConv1D	16	5	2	ReLU
BatchNormalization	–	–	–	–
SeparableConv1D	16	5	2	ReLU
BatchNormalization	–	–	–	–
MaxPooling1D	–	–	–	–
SeparableConv1D	32	3	1	ReLU
BatchNormalization	–	–	–	–
MaxPooling1D	–	–	–	–
SeparableConv1D	32	3	1	ReLU
BatchNormalization	–	–	–	–
MaxPooling1D	–	–	–	–
SeparableConv1D	32	3	1	ReLU
BatchNormalization	–	–	–	–
MaxPooling1D	–	–	–	–
SeparableConv1D	64	3	1	ReLU
BatchNormalization	–	–	–	–
MaxPooling1D	–	–	–	–
SeparableConv1D	64	3	1	ReLU
BatchNormalization	–	–	–	–
MaxPooling1D	–	–	–	–
SeparableConv1D	128	1	1	ReLU
BatchNormalization	–	–	–	–
Flatten	–	–	–	–
Dense	16	–	–	ReLU
Dense	1	–	–	–

ReLU Rectified linear unit

APPENDIX B – Predicted *versus* measured plots for pollution and salinity variables

Figure 1 - Predicted *versus* measured scatter plot for Ce prediction using Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).

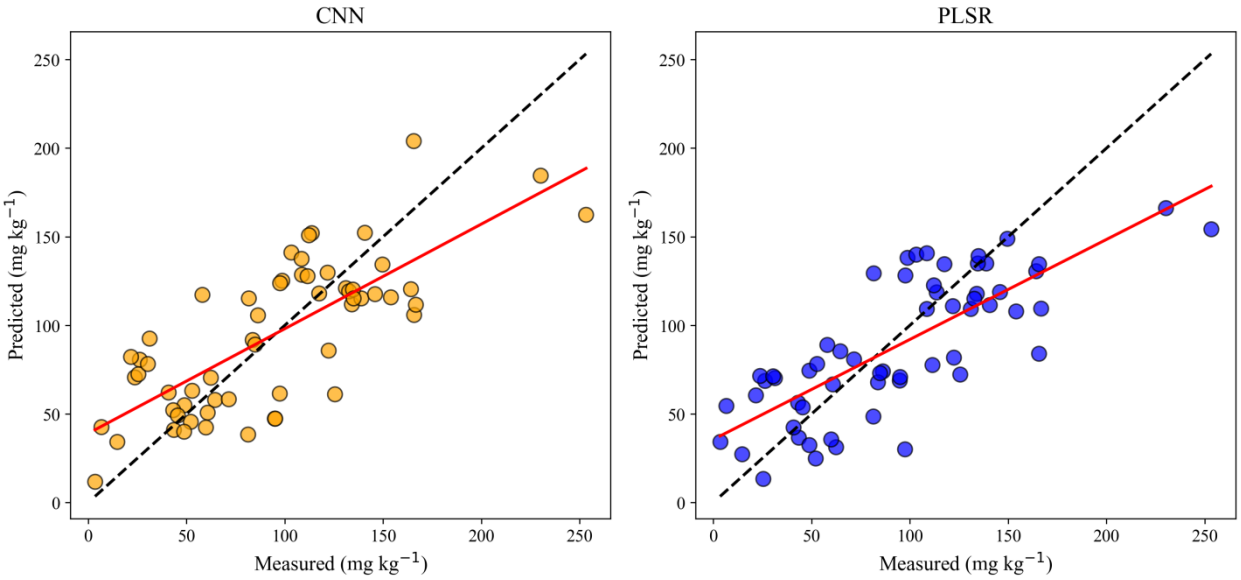


Figure 2 - Predicted *versus* measured scatter plot for La prediction using Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).

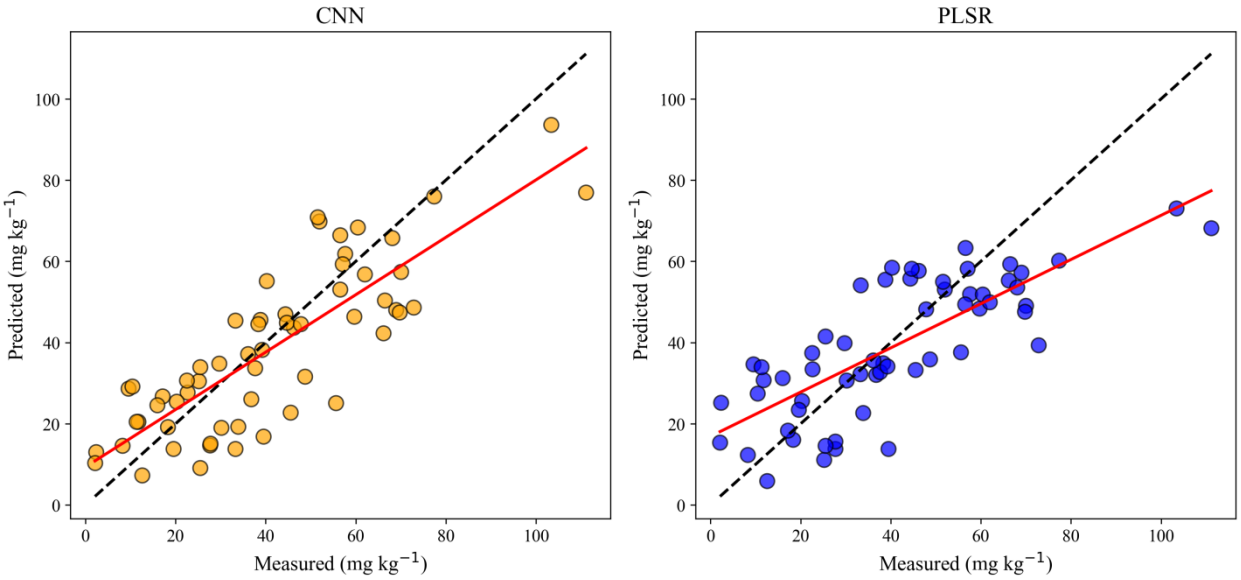


Figure 3 - Predicted *versus* measured scatter plot for Nd prediction using Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).

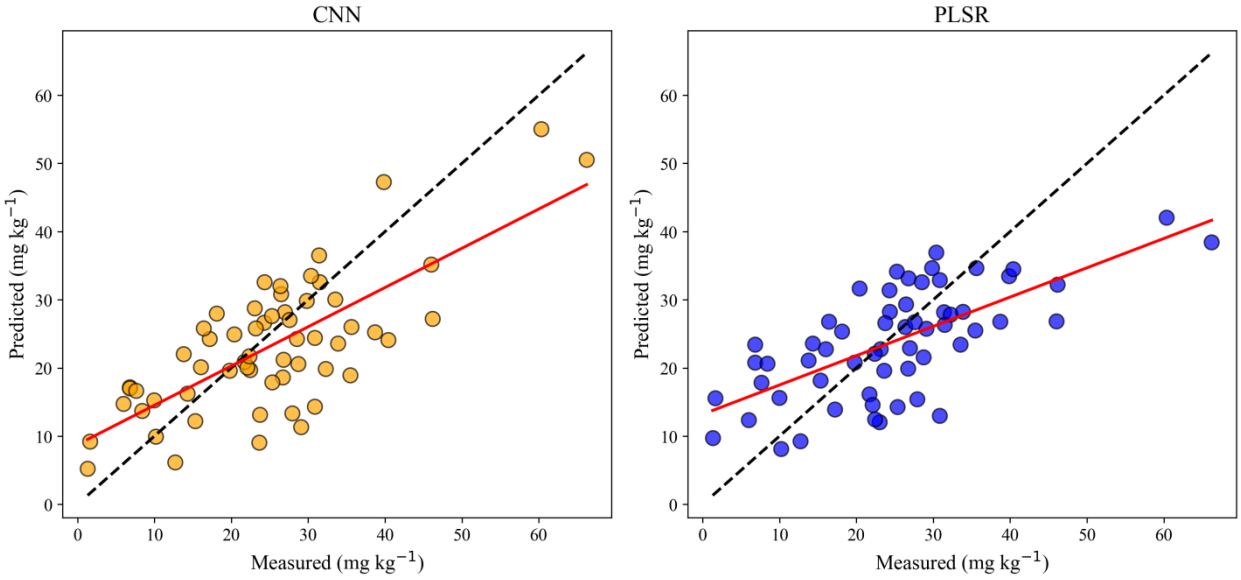


Figure 4 - Predicted *versus* measured scatter plot for Pr prediction using Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).

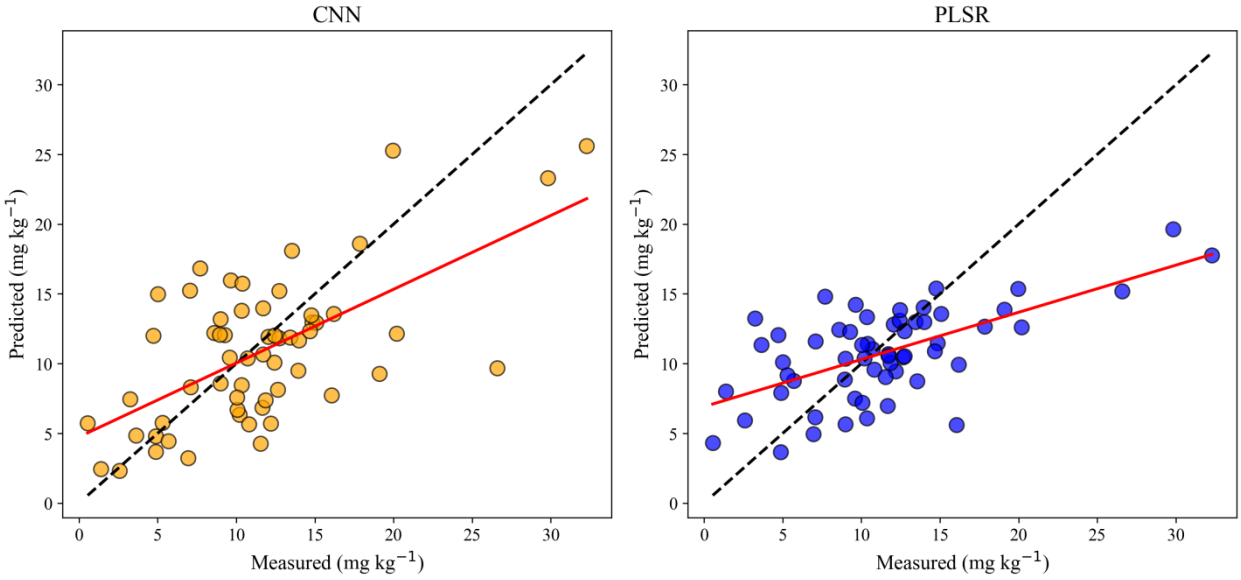


Figure 5 - Predicted *versus* measured scatter plot for Sm prediction using Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).

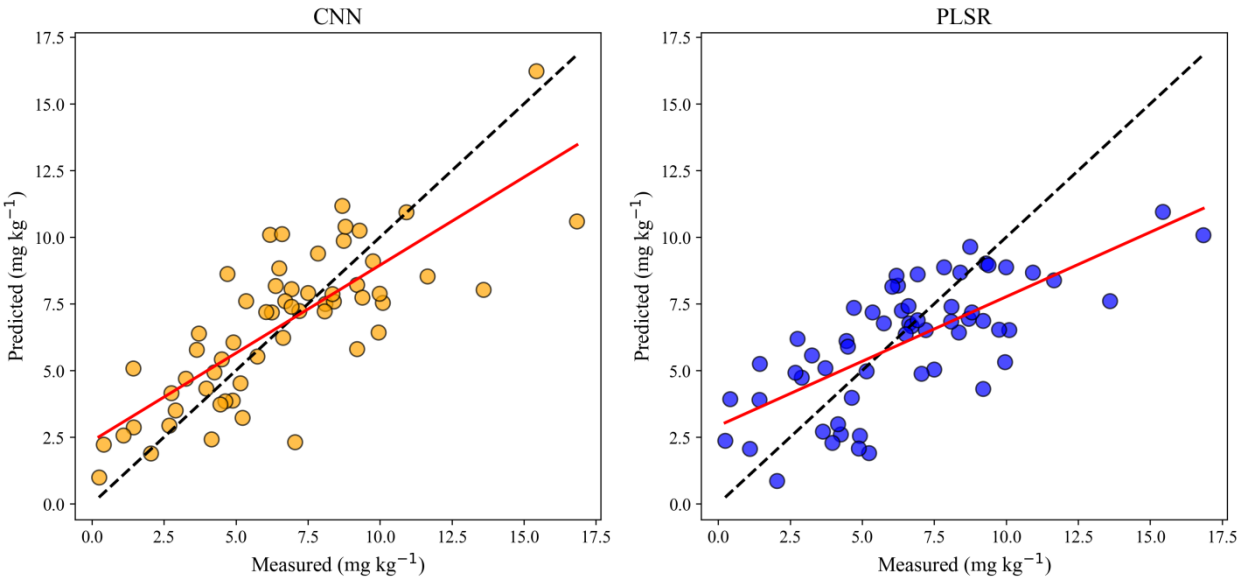


Figure 6 - Predicted *versus* measured scatter plot for Co prediction using Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).

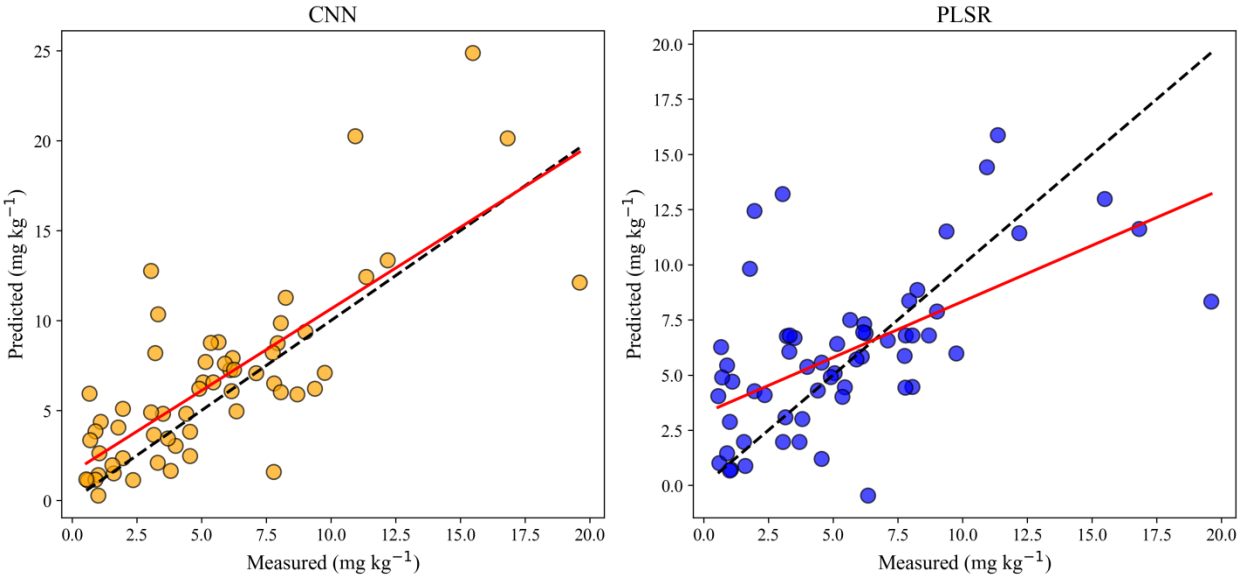


Figure 7 - Predicted *versus* measured scatter plot for Cr prediction using Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).

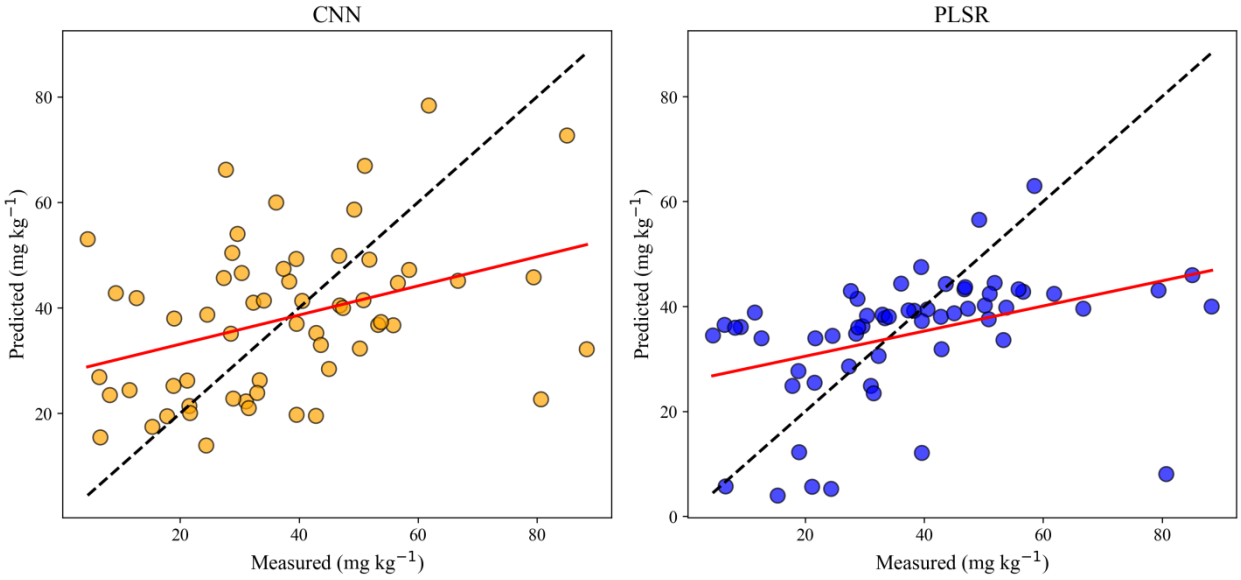


Figure 8 - Predicted *versus* measured scatter plot for Ni prediction using Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).

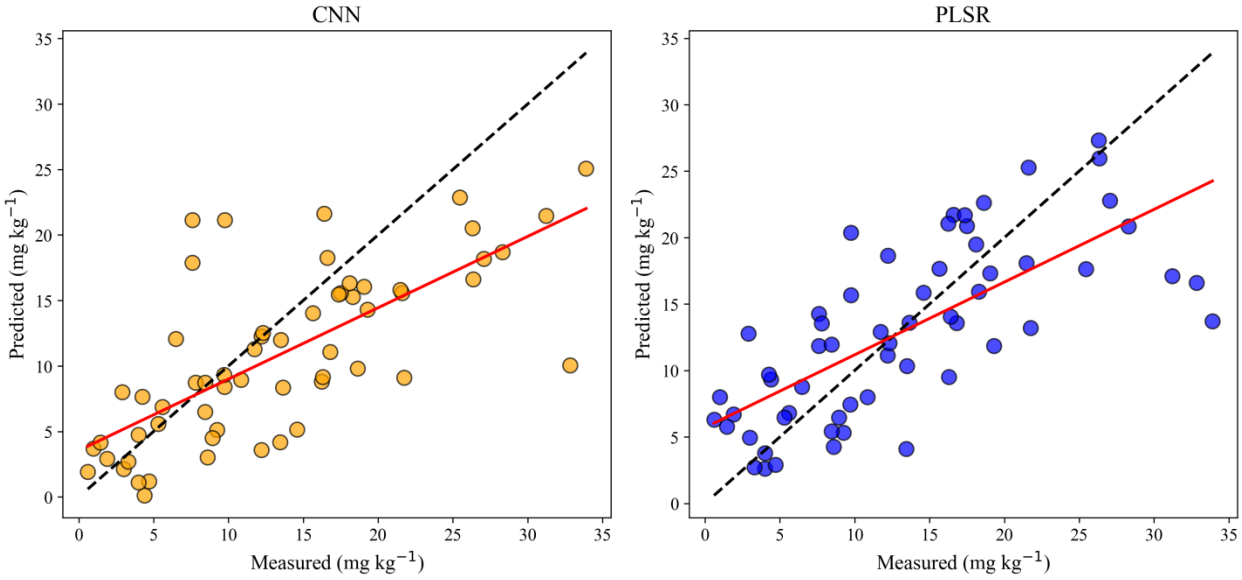


Figure 9 - Predicted *versus* measured scatter plot for Pb prediction using Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).

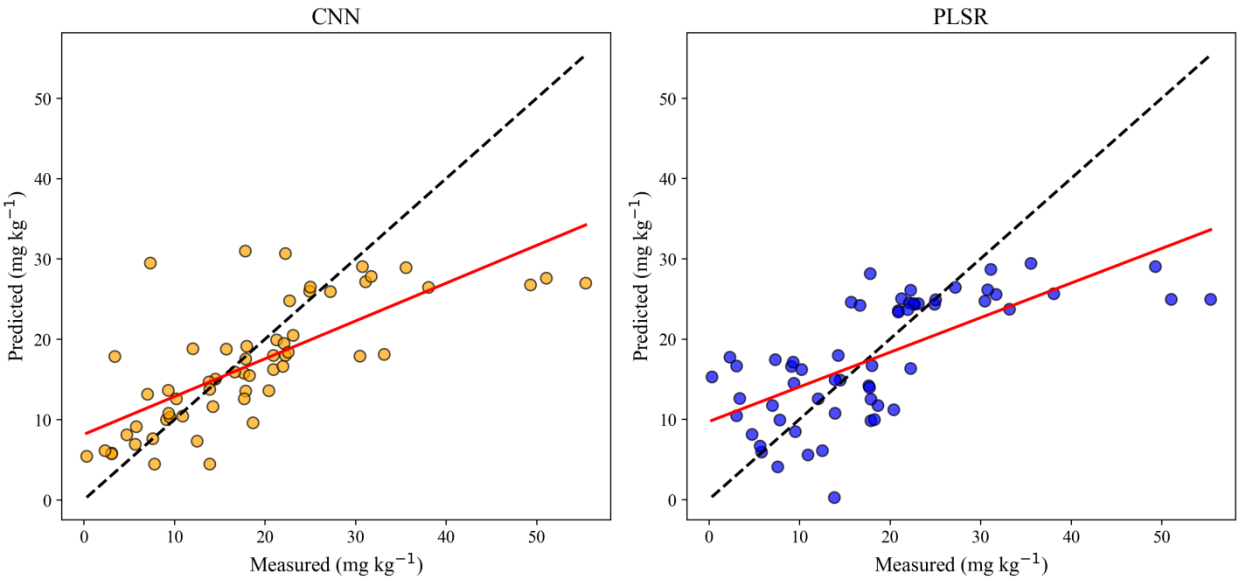


Figure 10 - Predicted *versus* measured scatter plot for Zn prediction using Convolutional Neural Network (CNN) and Partial Least Squares regression (PLSR).

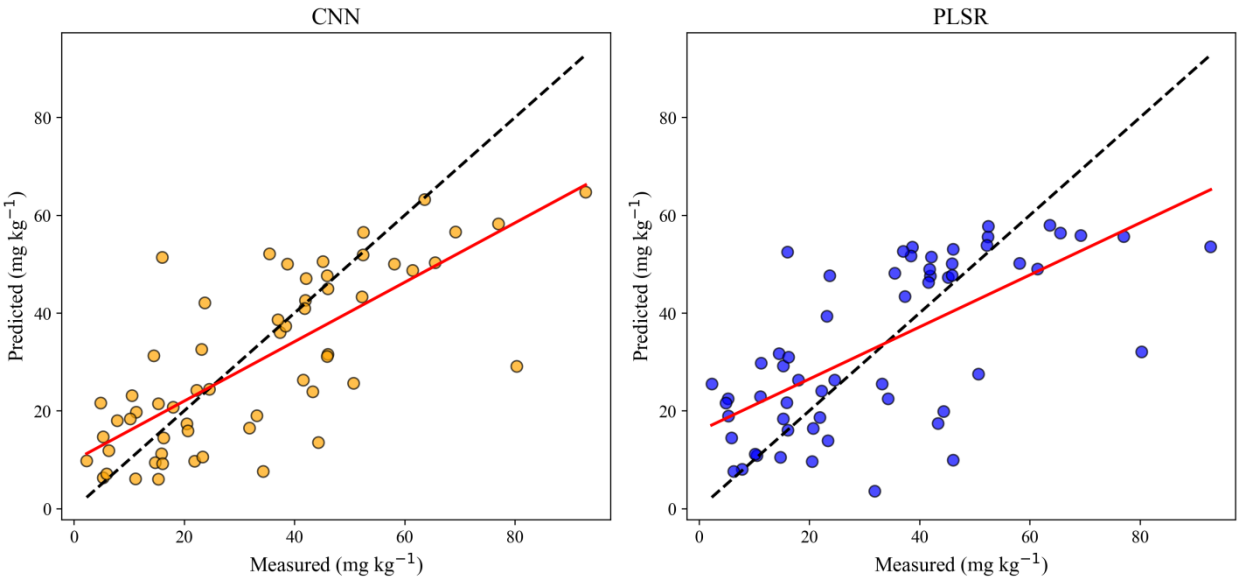


Figure 11 - Cation exchange capacity (CEC) predicted *versus* measured values for both convolutional neural network (CNN) and partial least squares regression (PLSR) using the three spectral datasets: near-infrared (NIR), mid-infrared (MIR), and the fused spectra of NIR and MIR (Fusion).

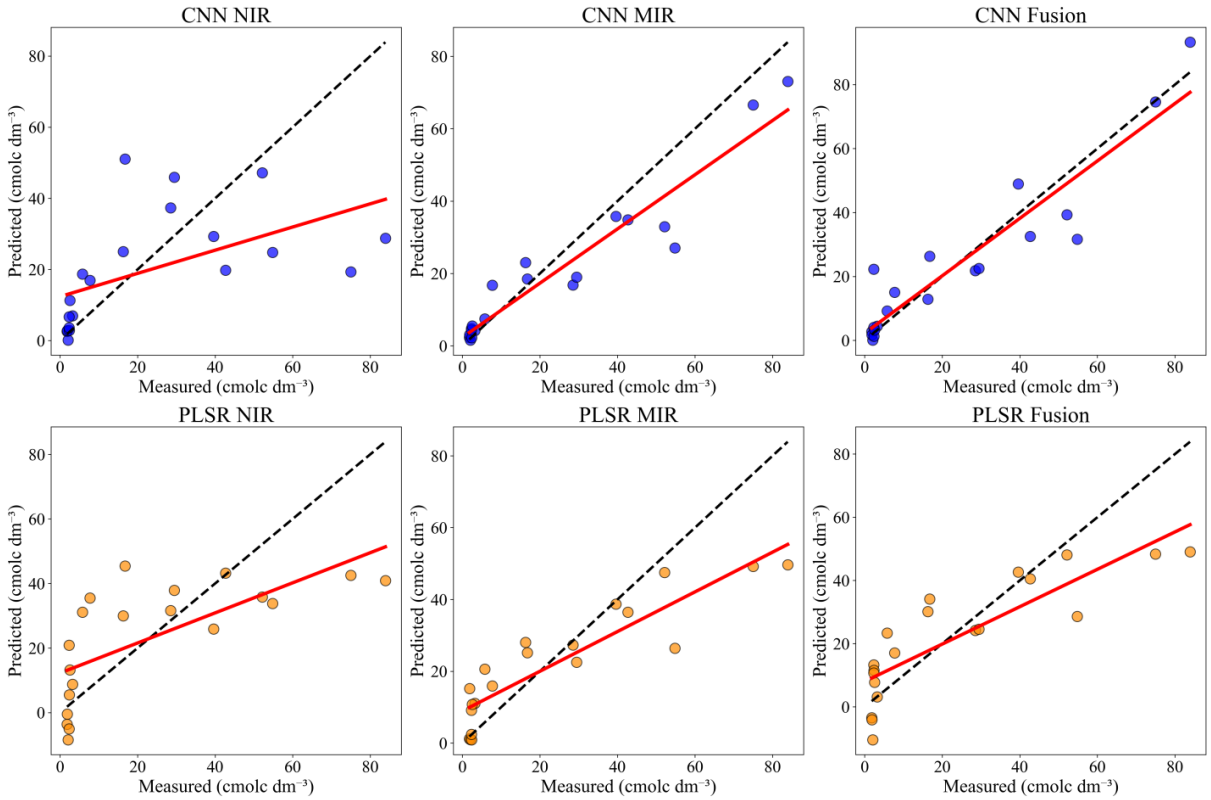


Figure 12 - Calcium (Ca) predicted *versus* measured values for both convolutional neural network (CNN) and partial least squares regression (PLSR) using the three spectral datasets: near-infrared (NIR), mid-infrared (MIR), and the fused spectra of NIR and MIR (Fusion).

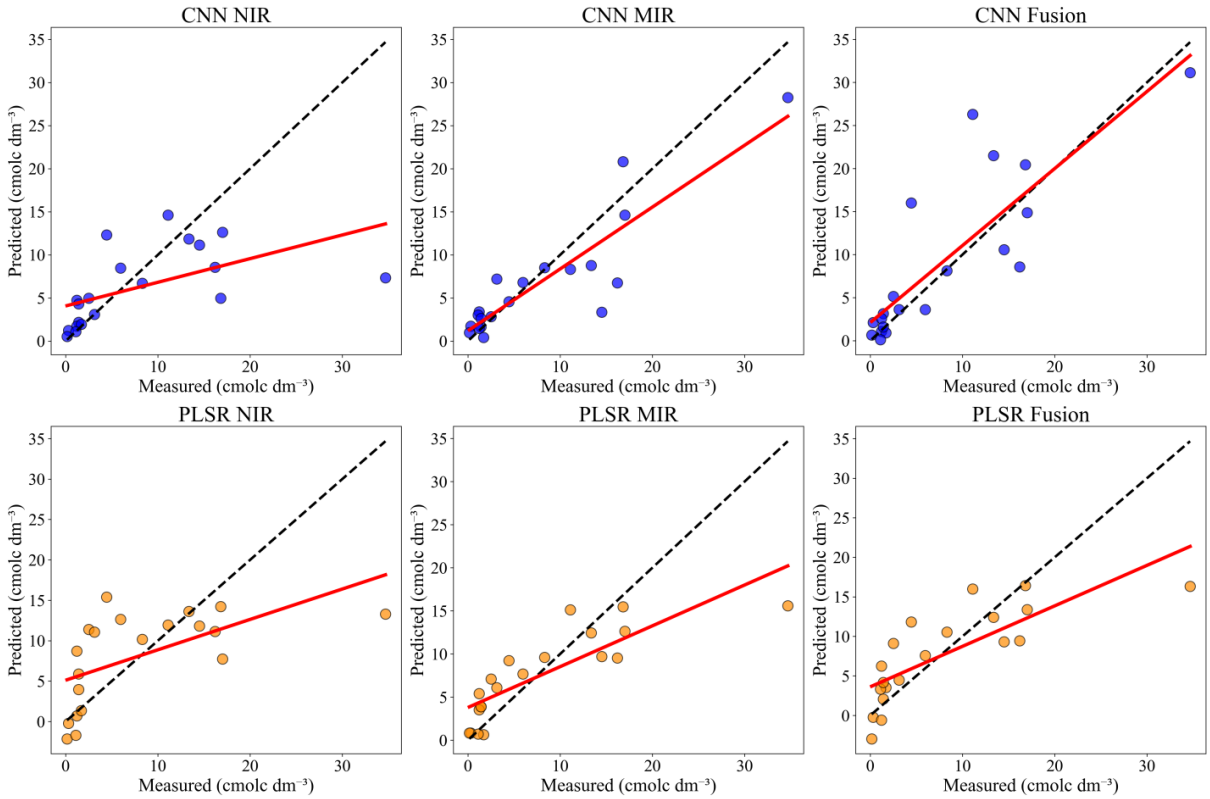


Figure 13 - Potassium (K) predicted *versus* measured values for both convolutional neural network (CNN) and partial least squares regression (PLSR) using the three spectral datasets: near-infrared (NIR), mid-infrared (MIR), and the fused spectra of NIR and MIR (Fusion).

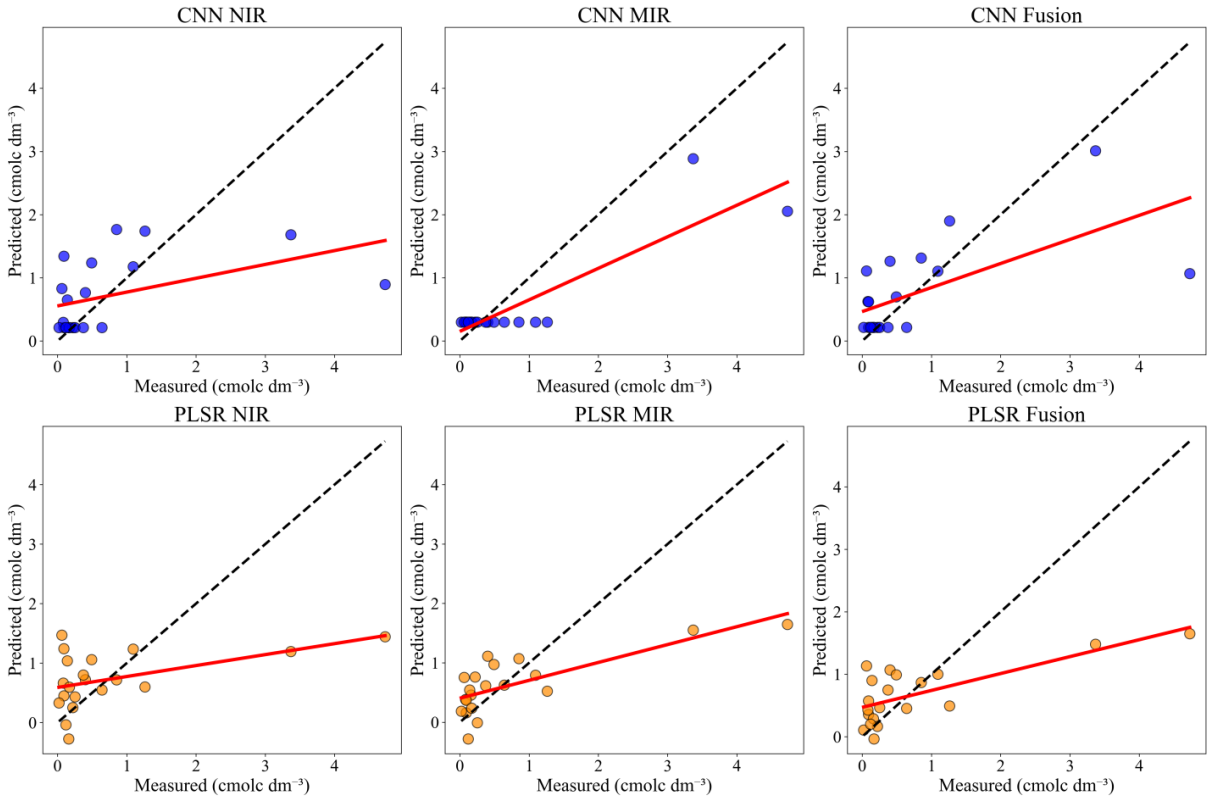


Figure 14 - Magnesium (Mg) predicted *versus* measured values for both convolutional neural network (CNN) and partial least squares regression (PLSR) using the three spectral datasets: near-infrared (NIR), mid-infrared (MIR), and the fused spectra of NIR and MIR (Fusion).

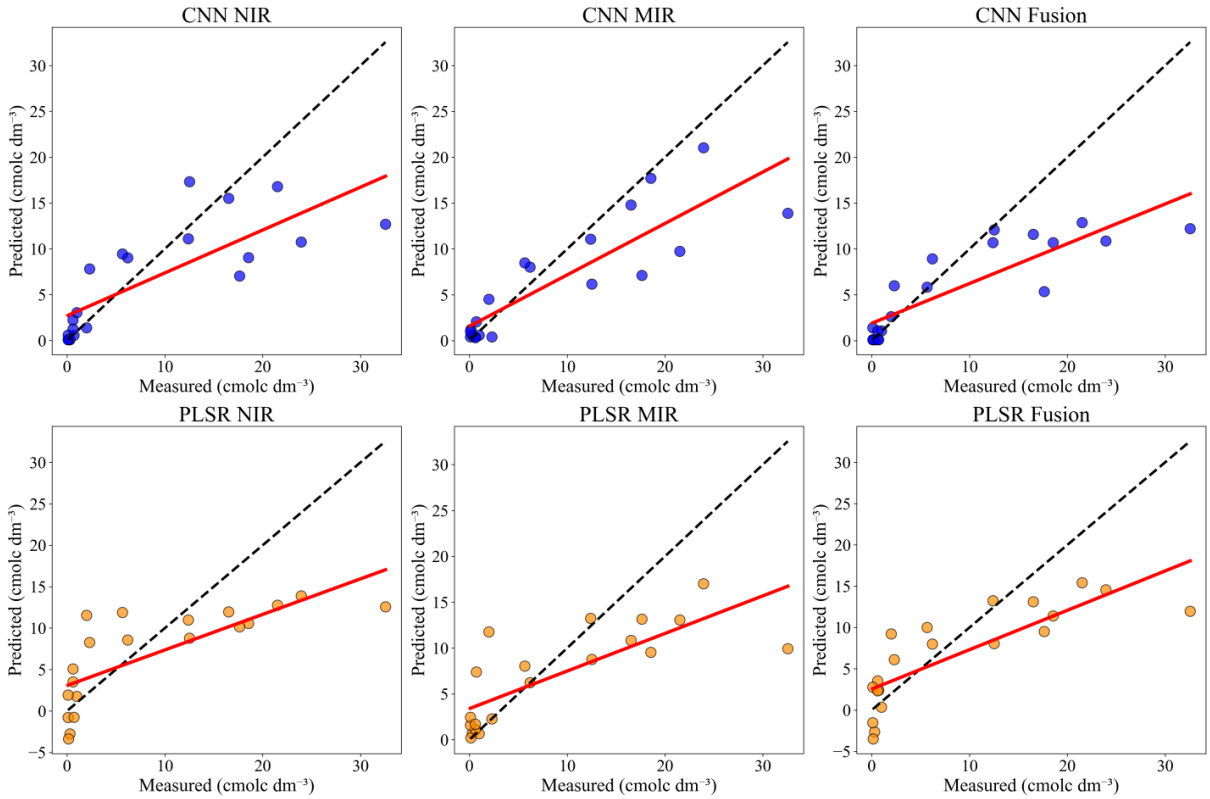


Figure 15 - Sodium (Na) predicted *versus* measured values for both convolutional neural network (CNN) and partial least squares regression (PLSR) using the three spectral datasets: near-infrared (NIR), mid-infrared (MIR), and the fused spectra of NIR and MIR (Fusion).

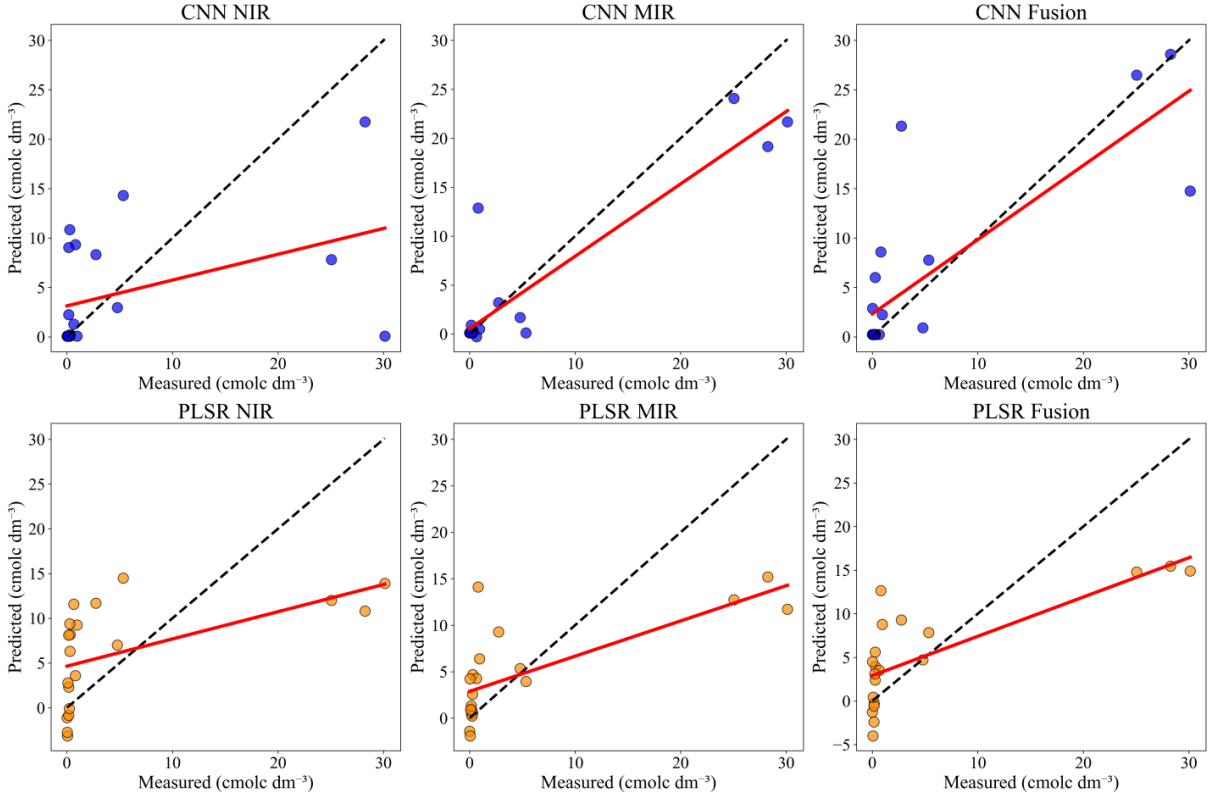


Figure 16 - pH with H_2O (pH H_2O) predicted *versus* measured values for both convolutional neural network (CNN) and partial least squares regression (PLSR) using the three spectral datasets: near-infrared (NIR), mid-infrared (MIR), and the fused spectra of NIR and MIR (Fusion).

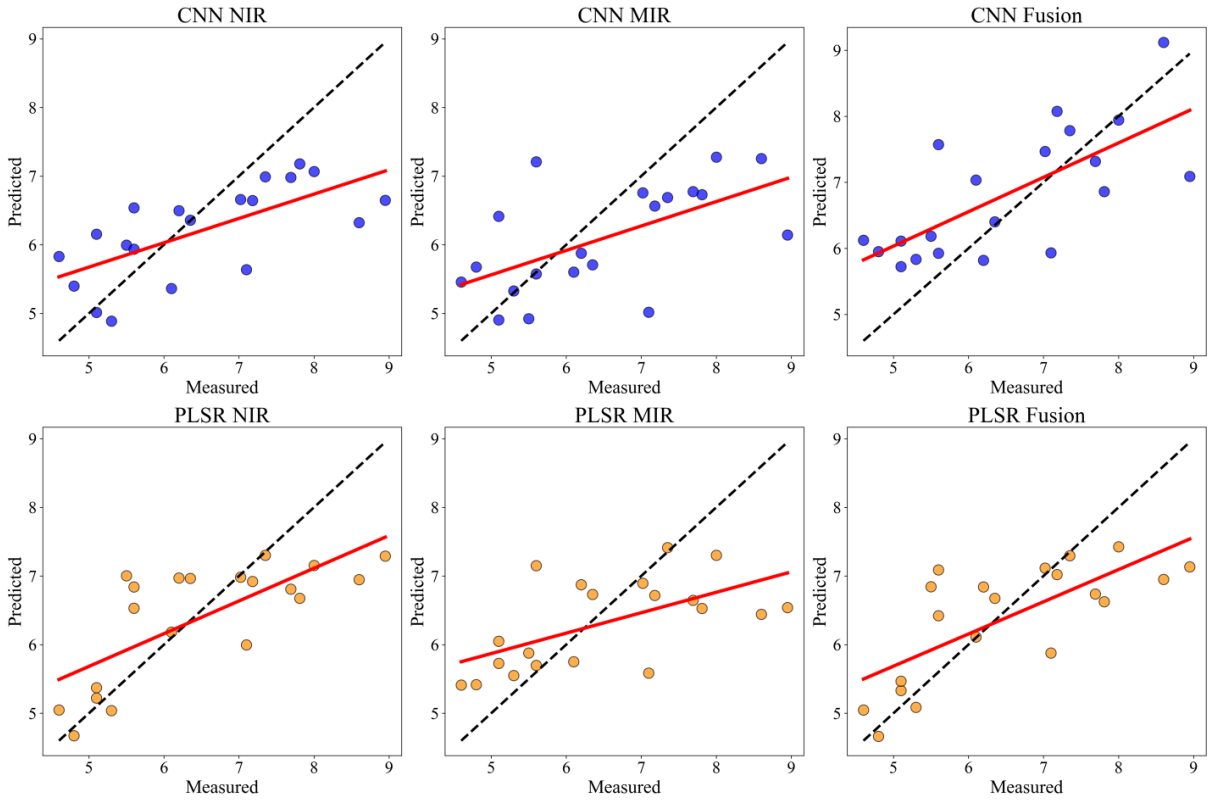


Figure 17 - pH with KCl (pH KCl) predicted *versus* measured values for both convolutional neural network (CNN) and partial least squares regression (PLSR) using the three spectral datasets: near-infrared (NIR), mid-infrared (MIR), and the fused spectra of NIR and MIR (Fusion).