CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

FACULTY OF AGROBIOLOGY, FOOD AND NATURAL RESOURCES
DEPARTMENT OF SOIL SCIENCE AND SOIL PROTECTION

Spatial Prediction of Potentially Toxic Elements Contents in Soil Using Digital Soil Mapping Approaches

…………………………………………………………………..

Doctoral dissertation

Ph.D. Student: Ing. Samuel Kudjo Ahado

Supervisor: Prof. Dr. Ing. Luboš Borůvka

Prague 2024

DECLARATION

I, **Ing. Samuel Kudjo Ahado MSc.,** hereby declare that the dissertation thesis titled " Spatial Prediction of Potentially Toxic Elements Contents in Soil Using Digital Soil Mapping Approaches" is the original work and, duly investigated using scientific protocols, has not been submitted anywhere for any degree or professional qualifications.

Prague, March 2024

Ing. Samuel Kudjo Ahado MSc.

ACKNOWLEDGEMENTS

I would like to warmly thank everyone who contributed to the success of this doctoral adventure. This thesis is a testament to the collective support and encouragement I have received over the years.

Above all, I will be forever grateful to my supervisor, Prof. Dr. Ing. Luboš Borůvka, whose leadership, wisdom and unwavering support have been invaluable throughout this process. Your mentorship has shaped my academic and personal growth, and I am grateful for the patience and dedication you have shown in navigating the complexities of this research.

My gratitude goes to the faculty and staff of the Department of Soil Science and Soil Conservation at the Czech University of Life Sciences Prague, whose dedication to academic excellence has created an enriching environment for intellectual exploration. I am grateful for the resources, facilities, and opportunities provided by the department, which have greatly facilitated my research efforts.

To my colleagues and fellow researchers, your camaraderie and shared passion for knowledge have made this academic journey more worthwhile. The countless discussions, debates, and collaborative efforts have shaped my understanding and broadened my perspective. I am indebted to my family for their continued support, encouragement and understanding through the ups and downs of this doctoral journey. Your belief in my abilities has been a constant source of motivation, and I am deeply grateful for the sacrifices you have made to allow me to focus on my studies.

Finally, I would like to express my gratitude to all the participants and people who contributed to my research. Your willingness to share your insights and experiences was crucial in shaping the empirical basis of this thesis.

In conclusion, this achievement is not mine alone, but is the result of the collective efforts of a supportive community. Thank you all for being part of this transformative academic journey.

With deep gratitude,

Ing. Samuel Kudjo Ahado MSc.

PhD candidate, Department of Soil Science and Soil Protection

Czech University of Life Sciences Prague, 2024.

PREFACE

This doctoral dissertation is completed using datasets from the Frýdek-Místek district and additional data from the Jizera Mts. To achieve the objectives of the study by understanding the spatial and vertical distributions of potentially toxic elements (PTEs) in soil, different algorithms were adopted to model, map and predict the PTEs in the soil. The thesis adopted supplementary auxiliary datasets to enhance the work and achieve the objectives. Among the additional datasets were raster and non-raster data from different sources including Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), satellite imageries (such as Landsat 7 and 8, Sentinel 2), land use-cover (from EU-Corine data), terrain attributes and soil properties (from field surveys and samplings). Further, legacy database of forest soil and topographic properties which comprises thousands of institutional samples collected throughout the Czech Republic between 1998 and 2019 were also incorporated to enhance the prediction of PTEs in the agricultural and forest soils investigated. Though the study focused on the Frýdek-Místek district, an extension was made to the Jizera Mountains following the past historical information of intensive anthropogenic activities domiciled in the area.

The work focused on selected PTEs which were known to be more impactful such as Pb, Cu, Cr, Zn, Cd, Ni, As and Mn. Several geospatial, geostatistical, and machine learning algorithms, models and tools including GIS spatial analytics, various kriging approaches, and the positive matrix factorization (PMF) were applied to effectively model, map, and predict the level and trends of PTEs in the studied soils in the study areas. In addition to the prediction modelling of the PTEs in the soils, the source distribution was evaluated. The accuracy of the applied models for effective prediction of the PTEs in the soils were affirmed by introducing various validation algorithms namely, the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), RPIQ (ratio of performance to interquartile range), and median absolute error (MdAE). The Department of Soil Science and Soil Protection under the Faculty of Agrobiology, Food and Natural Resources of the Czech University of Life Sciences Prague (CZU) served as indispensable surveillance during the entire thesis. The various grants units, technical supporters, and co-authors were given adequate compensatory acknowledgement in the appropriate sections of the publications. However, the study was not performed at a national scale yet the results from the study have promising information that will promote the integration of field sampling data in known PTEs

elevated areas to serve as adjunct to legacy data in enhancing the prediction at either regional or national scales. It could be imperatively concluded that the bibliometric outcomes gathered from the different scientific literature adopted in the thesis point that there is no definite modeling method that is spectacularly overwhelming in all studies. The following paragraph outlines the papers that form the thesis.

*Agyeman, Prince Chapman, **Samuel Kudjo Ahado**, Luboš Borůvka, James Kobina Mensah Biney, Vincent Yaw Oppong Sarkodie, Ndiye M. Kebonye, Kingsley John (2021). Trend analysis of global usage of digital soil mapping models in the prediction of potentially toxic elements in soil/sediments: a bibliometric review. **Environmental Geochemistry and Health,** 43: 1715-1739.*

***Samuel Kudjo Ahado**, Agyeman, Prince Chapman, Luboš Borůvka, Radoslava Kanianska, Chukwudi Nwaogu (2023). Using geostatistics and machine learning models to analyze the influence of soil nutrients and terrain attributes on lead prediction in forest soils. **Modeling Earth Systems and Environment,** 10:2099-2112.*

***Samuel Kudjo Ahado**, Chukwudi Nwaogu (2023). Spatial modelling and quantification of soil potentially toxic elements based on variability in sample size and land use along a toposequence at a district scale. **International Journal of Environmental Science and Technology**, 20: 1-20.*

*Agyeman, Prince Chapman, **Samuel Kudjo Ahado**, Kingsley John, Ndiye Michael Kebonye, James Kobina Mensah Biney, Luboš Borůvka, Radim Vašát, Martin Kočárek (2021). Source apportionment, contamination levels, and spatial prediction of potentially toxic elements in selected soils of the Czech Republic. **Environmental Geochemistry and Health,** 43: 601-620.*

***Samuel Kudjo Ahado**, Chukwudi Nwaogu, Vincent Yaw Oppong Sarkodie, Luboš Borůvka (2021). Modeling and assessing the spatial and vertical distributions of potentially toxic elements in soil and how the concentrations differ. **Toxics**, 9: 181-187.*

*Agyeman, Prince Chapman, **Samuel Kudjo Ahado**, Kingsley John, Ndiye Michael Kebonye, Radim Vašát, Luboš Borůvka, Martin Kočárek, Karel Němeček (2021). Health risk assessment and the application of CF-PMF: A pollution assessment–based receptor model in an urban soil. **Journal of Soils and Sediments,** 21 (9): 3117-3136*

TABLE OF CONTENTS

1.0 LITERATURE REVIEW

1.1 Soil, and soil pollution

Soil as a concept could be defined for the understanding of everyone weather experts or layman. This work attempted to provide the definitions of soil for the full comprehension of the two classes of persons (expert and layman) according to Dazzi and Papa (2022):

For the experts: Soils are living, four-dimensional natural phenomena possessing solids, water (or ice) and air that could store and transform energy and matter, as well as having a significant role of interacting with the atmosphere, lithosphere, biosphere, and hydrosphere. Though soils are components that are outside and are open systems, they also exist in shallow lakes and beneath pavement. They could be of any color, any age, be very shallow or deep, and consist mostly of a structured mixture of sand, silt, and clay (inorganics), rocks and organic material (dead and alive).

For the layman: Soils are the thin layer covering our planet earth. They are complex integration of living and mineral elements. Soils are essential Earth resources that purify and conserve water as indispensable portion of the planetary system that provide food and habitat for man, plants, and animals.

Soil pollution could be defined as a degradation process caused by presence in soils of substances detrimental to the health of humans, animals, and plants (FAO, 2018). Soil pollution can hence be seen as the alteration of soil health, defined as the continued capacity of soil to function as a vital living ecosystem that sustains plants, animals, and humans (USDA-NRCS, 2018). Soil pollution is primarily due to the existence of an anomalous chemical or substance in a higher than-normal content that has adverse effects on any non-targeted organism (Binner et al. 2023; FAO & ITPS, 2015). One group of pollutants is commonly known as the (potentially) toxic elements (PTEs). The PTEs have human sources, can exist naturally in soils as mineral materials, and could be

harmful in large amounts. The challenge with soil pollution is that it is in most cases not accurately studied or accessible. To exacerbate the situation is the fact that contaminants are rapidly increasing due to anthropogenic advancements in agrochemical and industrial processes. To complicate the issue is the complexity, diversity, uncertainty, time-sapping, and financial exploits involved in soil studies, thus making it difficult to identify pollutants. In addition, the impacts of soil pollution are more compounded by the characterization of soil which limits the free mobility and cycling of contaminant, bioavailability, as well as swaying the soil residence time (An et al. 2023; Elkhlifi, et al. 2023; FAO & ITPS, 2015).

Globally, many anthropogenic drivers have been identified as key causes of soil pollution, namely industrialization, wars, mining, and agricultural intensification (Yang & Yang 2023; Nwaogu and Cherubin 2023; Luo et al. 2009). In some cases, soil is used as a sink for dumping strong and liquid pollutants since urban expansion (Usoh et al. 2023). This act is well structured so immediately the wastes were buried, they would present no danger to the planetary health (either man or the environment) (Usoh et al. 2023; Swartjes & Siciliano, 2012). The principal sources of soil pollution are humans, causing accumulation of toxic wastes in soils that could be aggravated (Martin et al. 2023; Duarte et al. 2018). The features of soil permit it to naturally absorb and promote the cycling of inorganic chemicals or PTEs including lead (Pb), chromium (Cr), zinc (Zn), cadmium (Cd), manganese (Mn), nickel (Ni), copper (Cu), mercury (Hg), arsenic (As), antimony (Sb), and cobalt (Co) (Adomako et al. 2023; Sun & Chen, 2016). In Europe, North Africa, Asia, Northwest Pacific, North America, and Sub-Saharan Africa and Latin America, PTEs have been identified respectively as the third, fourth, sixth, seventh, eighth, and the ninth most important degraders to soil quality and functions (FAO & ITPS, 2015). In Europe, 6.24% (1,37,000 km^2) of agricultural land is polluted with PTEs (Naidu et al. 2021). In the Czech Republic for instance,

there is a coal reserve of about 705 million tons with approximately 50 million tons of being produced annually, making the country the 14th biggest producer in the world. Thus, soil pollution is an inevitable impact of mining in the Czech Republic and elsewhere (Rouhani et al. 2023).

1.2 Potentially toxic elements (PTEs)

Potentially toxic elements (PTEs) are a class of metals, semimetals, and non-metals that are of concern due to their persistence, toxicity, bioaccumulation, and biomagnification in high concentrations, posing threats to the planetary health including human, animal, plants, and environment (de Almeida Ribeiro Carvalho et al. 2022). These elements are naturally parts of the earth's crust soils that are mostly harmful when in large concentrations (Ahado et al. 2021; Eriksson et al. 2017). Both natural processes and anthropogenic interventions constitute the primary sources of the PTEs concentrations (Shar et al. 2021; Mondal et al. 2021; Kalkhajeh et al. 2021). The natural processes are defined as the materials produced from parent material, while the human sources basically emanate from intensive human activities. Many studies have demonstrated that natural sources of some PTEs (namely Pb, Cd, and Hg) have been exceeded by human discharges into the environment due to pedogenesis (Kalkhajeh et al. 2021; Lu et al. 2012). Industrial emissions, burning of fossil fuels, municipal wastewaters, and sewage sludge are components of the human sources of metals (Peli et al. 2021; Liu et al. 2020; Nwaogu et al. 2017). Further, agricultural intensifications and diversifications have been documented to elevate PTEs in the environment including soil and water (Kalkhajeh et al. 2021; Sungur et al. 2021; Cai et al. 2012).

Moreso, automobile and vehicle discharges, road dusts, and military operations are responsible for high PTEs (Barker et al. 2021; Mondal and Singh 2021). According to Luo et al (2009), farming activities are responsible for 79.6%, 56%, and 63% of the yearly total record of Cu, Zn, and Cd,

respectively, in soils. They also concluded that the total yearly input of Pb (85%), Ni (67.5%), and Cr (43%) presence in the soil comes from industrial atmospheric emissions. Among the several harmful impacts of anthropogenic activities on the aquatic and terrestrial ecosystems is the huge mobilization, cycling, and dispersal of contaminants from their natural pools into the environment (atmosphere, soil, and water) (Hou et al. 2017; Zhao et al. 2014). Soil pollution is difficult to be frequently studied or outwardly noticed, making it a hidden threat. The numerous features of pollutants are constantly increasing because of agrochemical and industrial advancements. However, the effects of soil pollution also significantly rest on the prevailing soil properties since this regulates the cycling, mobility, bioavailability, and residence time of PTEs (FAO & ITPS, 2015). The irony or challenging dichotomy about soil PTE pollution is that interventions to combat them sometimes could introduce additional issues, This is because (a) PTEs are non-destructible and frequently accumulate rather than degrade in soils (Maas et al. 2010); (b) they have a wide range of health effects, and the health vulnerability is complicated by their oxidation state and associated bioavailability disparities (Walker et al. 2003); and (c) there are numerous diffusional sources of PTE contamination (Qu et al. 2020). Elevated quantities of PTEs in soils do not only affect soil health but owing to their steadiness in the environment and lasting biological half-life, they possess the potential of accruing in the food processes and consequently impact human health (Jannetto & Cowl 2023; Assey & Mogusu 2023; Velayatzadeh 2023; Dhuldhaj et al. 2023).

Though the adverse impacts of PTEs have been continuously reported, yet the consequence of their presence continues to increase globally in many agricultural lands, forests, industrial and waste pits globally including Nigeria (Nwaogu et al. 2017), China (Wen et al. 2023), Brazil

(Cardoso et al. 2023), Australia (Tao et al. 2023), Czech Republic (Rouhani et al. 2023), and USA (Sepúlveda et al. 2023).

1.2.1 Sources of PTEs in soils

Soil is a dynamic natural resource that is composed of various gases, minerals, salts, organic-inorganic components and living organisms. It has biological, chemical, and physical properties that are susceptible toward any abrupt changes, which can be caused by natural activities (volcanic eruptions, weathering of ores, forest fires, etc.) or most commonly by various anthropogenic activities (dumping of household and industrial waste, use of chemical fertilizers and pesticides to enhance crop (Gautam et al. 2023). Although soil has natural potential to either resist or regenerate or suppress the PTEs, this is mostly when the soil is of high quality or when the PTEs are not in high concentrations. PTEs encroach into the soil from various routes, including natural sources and anthropogenically-induced sources. It is pertinent to state here that the elementary sources of soil pollution are humanly caused which aggregates over time (Duarte et al. 2018). Source of pollutant occurs from different paths, including natural enrichment, agricultural activities (artificial fertilizers, animal manures, composts, pesticides), industrial activities, transportation system, atmospheric deposition, waste management and treatment, and mining (Wang et al. 2023; Xue et al. 2023; Demir et al. 2023; Ma et al. 2023; Yuan et al. 2023; Nwaogu et al. 2017). Several studies have revealed that the PTEs of human generated sources are generally more mobile and bioavailable in soil than PTEs of natural origins (such as volcanic eruptions, geological processes/rock weathering reactions, wildfires) (Pacifico et al. 2023; Kharuk et al. 2021). Natural processes including wildfires, volcanic eruption, weathering of rocks, erosion, and rock formations contribute substantially to the release of large amounts of PTEs in the environment (Pacifico et al. 2023; Hoshyari et al. 2023; Kharuk et al. 2021). Increase in food production chain due to increase

in human population has been reported as one of the primary causes of soil contaminations (Awad, 2023). As more food items are in high demand by man, more quantities of farmlands are cultivated with applications of more quantities of agrochemicals to increase food productions (Falconnier et al. 2023; Tudor et al. 2023).

1.2.2 Spatial distribution of PTEs

Spatial distribution of PTEs is principally defined by the source of pollution, which is basically soil pollution in this case, and is predominantly widespread (Hoshyari et al. 2023; Borůvka et al. 2005) and considering their level of toxicity and doggedness could be highly harmful to the planetary health (human and environment). Meanwhile, the spatial distribution of identified contamination has an important structure, with possible pollution proximity to the point source decreasing with distance away from the primary source (Belanović et al. 2023; Borůvka et al. 2005). Soil physicochemical features generally have substantial influence on the spatial distribution processes of PTEs (Mahmoudabadi et al. 2015). Studies have shown that the enrichment index of PTEs could be linked to diverse soil attributes and their spatial distribution of index enrichment for certain PTEs including Zn, Cd, and Ni based on the spatial arrangement of pH, soil organic matter (SOM), sand, and clay (Zhao et al. 2010). The metal deposit can be enhanced by increasing soil pH, organic matter, cation exchange capacity, and the amount of iron and manganese oxides (Zaib et al. 2023; Lake et al. 1984). For example, an increase in soil pH has been observed to have affected the solubility and mobility of certain pollutants, making them less available for plant uptake or leaching into deeper soil layers or groundwater (Zaib et al. 2023).The PTE moieties have been affirmed to have exhibited serious influence on the chemical mobility and bioavailability in soil (Zhang et al. 2018). Similarly, the bioavailable components of a PTE in most soils are crucial in its conglomeration by organisms. PTEs emitted by anthropogenic activities are

thought to have a high bioavailability (Bolan et al. 2014), this could be because many human operations such as Arbuscular Mycorrhizal Fungi (AMFs) application can also change PTE bioavailability in the soil-crop system (Chen et al. 2023). Therefore, effective identification of the chemical compositions of PTEs in soils could be a good support for examining their potential environmental issues (Sun et al. 2019).

1.2.3 The PTEs for study

1.2.3.1 Cadmium (Cd)

Cadmium (Cd) is a PTE that was first identified in 1817 as a by-product of the zinc processing industry. In the Earth's crust, Cd is about 700 times more concentrated than zinc and about $0.1-0.2$ mg kg^{-1} in the lithosphere (El Rasafi et al. 2022) and coexists with zinc ores, as and is a by-product from the refinement in zinc sulfide mining locations (Rao and Kashifuddin, 2016). Further, most Cd is derived from steel remains, which is the second largest source of Cd globally (Coutinho et al. 2023). In addition, decomposition of surface, weathering of parent rock mineral might promote the discharge of different elements including Cd, which elevates Cd contents in the environment (Liu et al. 2022). In farming, the applications of artificial fertilizers containing phosphorus elevates Cd in soil, whereas landfill, sewage sludge and other contamination sources also increase Cd concentration in the soils (Rao and Kashifuddin, 2016). Other known primary sources of Cd are mine tailings, compost inputs and irrigation of mine effluents (Bolan et al. 2013). Cd has been recognized as a hazardous element to the soil and humans (Olarewaju et al. 2023).

1.2.3.2 Chromium (Cr)

Chromium is a ubiquitous PTE deposited into the environment by various anthropogenic activities like plating, corrosion control, tanning, nuclear weapon production (Mohanty et al. 2023; Gosh et al. 2016; Ray et al. 2016). Other anthropogenic sources of Cr are mining and metallurgy, metal

plating, rubber, photography, industrial dust and fumes, tanning, leather industry, chemical industry, fertilizers, textile industry, paints and pigments, and the diffuse ones that can be wastewater and sludge from dyeing and tanning industries (Smiljanić et al. 2019). Chromium occurs in six various forms including Chromium (III) and Chromium (VI) which are the two main valence states of the metal that take part in inducing environmental toxicity (Mohanty et al. 2014). The trivalent form of chromium shows relatively mild toxicity relative to the hexavalent form of chromium. The former also has glucose, lipid and protein metabolism affirming that Cr is beneficial when exists in acceptable quantities. For example, Cr is important for glucose and fat breakdown, and lipoprotein intake in living things. Its biological function is intertwined with that of insulin, and most Cr-enhanced processes are insulin dependent. An excessive amount, on the other hand, could be harmful (Majhi and Samantaray 2020). In soils treated with wastewater, toxic levels are very common. More than 7.5 million tons of Cr are estimated to be generated yearly on a global scale. In nature, Cr is among the few PTEs that do not occur in elemental but exist in compound forms such as chromite (Wuana & Okieimen, 2011). Mobility and cycling of Cr is influenced by soil sorption properties including clay content, iron oxide content, and the amount of available organic matter. Further, surface runoff can transport Cr in its soluble or precipitated form to surface waters; soluble and desorbed Cr complexes can leach from soil into groundwater; the leachability of Cr(VI) increases as soil pH increases; however, the majority of Cr released into natural waters is particle associated and is finally discharged in the sediment or soil (Smith, 1995).

1.2.3.3 Copper (Cu)

Copper is known as the 25th most abundant component of the Earth's crust (Tu et al. 2024), it is also an essential micronutrient element for the growth and development of plants, animals, and humans. However, excess Cu in the soil is actively risky and might be very toxic to microflora,

flora, fauna, and humans. Though a PTE, Cu is malleable, ductile, and a good conductor of heat and electricity. A feature distinguished by a crystalline structure that absorbs frequencies in the visible range. Copper rapidly integrates to organics in the soil, suggesting that possibly a small fraction of Cu might be found in solution as ionic copper, Cu(II). Copper solubility is dramatically decreased at pH 5.5, close to the optimal farmland pH of 6.0 - 6.5 (Martínez & Motto, 2000). According to Smiljanić et al. (2019), the sources of Cu in the soil and the environment are point sources such as mining and metallurgy, plating, rayon, electrical and electronic waste, pesticides, paints, and pigments. Others area textile industry, explosive, and diffuse sources like manures, fertilizers, pesticides, sewage sludge and atmospheric fall out resulting from the combustion of fossil fuels and industrial processes.

1.2.3.4 Iron (Fe)

Iron takes the second place after aluminum among metals on the total concentrations in earth crust and the fourth most abundant element in the Earth's crust and occupies the 26th elemental position in the periodic table (Sánchez et al. 2017; Wedepohl, 1995). Iron is an indispensable element for growth and survival of almost all living organisms (Valko et al. 2005). It is an attractive transition metal for various biological redox processes because of its inter-conversion between ferrous (Fe^{2+}) and ferric (Fe^{3+}) ions (Phippen et al. 2008). The source of iron in surface soil and/or water is anthropogenic and is related to mining activities. The production of sulphuric acid and the discharge of ferrous (Fe^{2+}) takes place due to oxidation of iron pyrites (FeS_2) that are common in coal seams (Valko et al. 2005).

Iron mobilization in soils is regulated by redox processes, pH, and DOC availability for complexation. Iron as a metal is further considered to affect the mobilization and bioavailability of phosphorus (Bakker et al. 2016) and is reported to show a strong influence on carbon (C) stocks

and sequestration by modifying flocculation, sedimentation, and conservation of organic C in mineral sediments (Lalonde et al. 2012; von Wachenfeldt et al. 2008). The main sources of Fe to the soil are human activities including mining, industry, and agriculture. Naturally, weathering of mineral components of the soil is the first step of Fe mobilization (Giesler et al. 2000). Chemical weathering is promoted by acidifying compounds of different origins, namely organic acids from vegetation and microbes (Landeweert et al. 2001; Chen et al. 2000). Iron concentrations are commonly observed in catchments with coniferous forest, and constant land-use changes which are known to have relocated Fe from soils to waterbodies (Björnerås et al. 2017). The mobilization of Fe from soils depends extremely on either its link to organic matter (OM) (Jansen et al. 2004), redox reactions prompted by oxygen deficiency (Grybos et al. 2009), or low soil pH (Kuesel et al. 2001).

1.2.3.5 Lead (Pb)

Pb is among the primary elements in nature globally, and its content in the lithosphere ranges from 10 to 30 mg kg^{-1} (USDHHS, 2007). It exists naturally as a bluish-grey metal that is primarily seen as a mineral in combination with other elements such as Sulphur (PbS, $PbSO_4$) and/or oxygen ($PbCO_3$). Based on its elemental physical and chemical properties, Pb has many industrial applications. It stands as the fifth most widely exploited metal globally due to its high industrial needs. More than 900 different industries including oil refineries, drug manufacturing, and quarrying depend largely on Pb for most of their operations. There is a rise in the metal ion concentration in industrial effluent due to an increase in outflow from firms operating close to rivers. Worldwide, the average surface soil content of Pb content is about 32 mg kg^{-1} and ranges from 10 to 67 mg kg^{-1} (Kabata-Pendias, 2010a). Pb is introduced into the soil through various pathways including point and diffuse sources which are primarily anthropogenic. Some of the

sources are mining and metallurgy, industrial dust and fumes, application of lead in gasoline, combustion fossil fuel, solid waste, solid waste combustion and incineration, industrial waste, paints and pigments, explosives, ceramics and dishware, and some types of PVC. Others are agrochemicals (pesticides, fertilizers, herbicides), manufacturing of lead-acid batteries, urban runoff, exhaust gases of petrol engines, and atmospheric fallout from the combustion of fossil fuels.

1.2.3.6 Manganese (Mn)

Manganese is one of the most prevailing metals in soils, existing as oxides and hydroxides and cycling through its three oxidation states which are known primarily as pyrolusite (MnO_2) and, to a little extent, rhodochrosite ($MnCO_3$) (LENNTECH, 2008). The anthropogenic source of Mn that is exported into the soil and the environment in a point and the diffuse source is through the production of ferromanganese steels, electrolytic manganese dioxide applied in batteries, alloys, catalysts, fungicides, antiknock agents, pigments, dryers, wood preservatives, coating welding rods (Bradl, 2005b). Manganese is a mineral nutrient that is required by all plant species (PlantProbs.net, 2019). Further, Mn is accumulated by species such as diatoms, mollusks, and sponges. Manganese dioxide is adopted as a catalyst, and when chemically combined to potassium to form potassium permanganate, it becomes a powerful oxidant and disinfectant. Manganese oxide (MnO) and manganese carbonate ($MnCO_3$) are two other distinct compounds that have been exploited: the former is used in fertilizers and ceramics, whereas the latter is the starting material to produce other manganese compounds (LENNTECH, 2008). Manganese is one of the most widespread microelements with the average clark of total Mn in earth's crust which is about 850–900 mg kg^{-1} (Hettler et al. 1997).

1.2.3.7 Nickel (Ni)

Nickel is one of the most popular metals on earth and exists in various metal alloys used in the steel industry, also in colorants, taps, as well as in dry cells. Nickel is not only found in soils but could erode into water bodies due to its availability in contaminated water (Ogunlalu et al. 2021). Other human source for Ni are fertilizers, manures, metal refining, smelting, burning of coal and industrial sewage sludge, emissions from mining and smelting operations, an atmospheric fallout from the combustion of fossil fuels, mining and metallurgy, electroplating, production of iron and steel, industrial dust, industrial aerosols, incineration of waste, fertilizers, combustion of coal, battery, chemical industries, food processing industries (Alloway, 2013). In the environment, Ni is mostly available in little amounts, while food plants cultivated in polluted soils can absorb large amounts. Numerous cytotoxic activities carried out by Ni include the production of free radicals, genetic regulation, and the control of transcription factors. It has been discovered that Ni contributes to the control of the expression of a few lengthy non-coding chromosomes. Nickel has also been shown to produce reactive species, which contribute to neurotoxic development (Engwa, et al. 2019). Most of the Ni on earth is unavailable because it is caged in the planet's iron-nickel molten core, which has 10% nickel. It has been postulated that large quantities of nickel dissolved in the sea ca. 8 billion tons. Nickel concentrations in soil can vary from 0.2 - 450 ppm in clay and loamy soils.

1.2.3.8 Zinc (Zn)

Zinc (Zn) plays a crucial role in various biological processes and is subsequently an essential micronutrient for living beings (Broadley et al. 2012; Wessells et al. 2012; Brown et al. 1993). Zinc can also be toxic when present in excess (Giller et al. 1998; Tyler et al. 1989). As a result, too low or too high soil Zn concentrations negatively affect soil functions and may harm animal and

human health. Primary and secondary Pb smelters have been verified as the main point sources of Pb, Zn, and Cd pollution (Luo et al. 2023).

Beginning from the late 19th century till date, the production of Zn has rapidly increased (Han et al. 2002). Zinc is a widely used metal, mainly for galvanizing steel products, brass products, or die casting (Graedel et al. 2005). The exploitation and production of Zn contributes to the huge quantities of Zn released into the air and its subsequent discharge to soils, followed by waste incineration and fossil fuel combustion (Stemweis et al. 1988). The concentrations of Zn in soils globally ranged from 30-100 mg kg^{-1} on average, whereas higher quantities can be found in calcareous and organic soils (Kabata-Pendias and Szteke, 2015). Meanwhile, anthropogenic Zn sourcing from different agricultural and mining activities might elevate Zn levels in some soils (Araújo et al. 2017). Furthermore, Zn emanating from industrial processes and products has drastically increased the local contents of the metal. For instance, corrosion of galvanized facilities might lead to exacerbated contents of Zn, especially in the soils (Bertling et al. 2006). Zinc as part of the particulate matter discharged by brake linings or tires was reported to be associated with heightened Zn concentrations in the soils closer to the major roads with heavy traffic (Nwaogu et al. 2017; Hjortenkrans et al. 2007). Moreover, agricultural inputs such as chemical fertilizers, and manure can introduce unplanned quantity of Zn. For instance, chemical fertilizers having trace elements release Zn in soils through agricultural applications (Nziguheba and Smolders, 2008). On the other hand, organic fertilizers such as sewage sludge or manure could possess substantial quantities of Zn because of feed additives (Mantovi et al. 2003). Therefore, these Zn inputs as emphasized above have led to soil pollution globally as reported by several studies in Europe (Kabala et al. 2020; Douay et al. 2008; Steinnes et al. 1997), Africa (Famuyiwa et al. 2022; Nwaogu et al. 2017; Kazapoe et al. 2021), Asia (Wen et al. 2023; Yang et al. 2023; Yang and Yang

2023; Xue et al. 2023; Wang et al. 2023; Hoshyari et al. 2023), Southern America (Sepúlveda et al. 2023; de Almeida Ribeiro Carvalho et al. 2022; Reyes et al. 2021), Northern America (Klapstein et al. 2020), and Oceania (Tao et al. 2023).

1.3 Knowledge gap

Judging from the above literature, it is obvious that there have been problems associated with spatial prediction of PTEs and impacts in forest and agricultural soils (Agyeman et al. 2021). The role of forest and agricultural soils in sequestering CO_2 and providing the growing population with food has been very crucial lately. This has therefore put insurmountable pressure on the forest and agricultural soils, thereby making the soil lose their natural ecosystem services because many contaminants (e.g. PTEs) have saturated them. Since human population continuously increases, more food, fibre, carbon stocks are needed from the forest and agricultural soils, thus many research works have recently focused on new innovations to redeem the soil by ameliorating PTEs concentrations (Agyeman et al. 2022).

Many conventional approaches adopted have failed to provide sustainable solutions in the reduction of increasing contents of PTEs in the soils because these approaches cannot produce accurate spatial predictions for precise quantifications of the PTEs over time. Digital soil mapping approaches have brought the solution by helping to close the gap by enhancing the availability of updates, quantitative, dependable, and valid soil data. In contrast to the traditional soil mapping approaches, DSM provides reliable information to support the policy-makers in developing timely, effective, and achievable decisions regarding sustainable soil health and management (Ahado et al. 2023). Since the advent of DSM, prediction of PTEs integrated large arrays of environmental factors, namely remote sensing, terrain attributes, digital elevation model, and other related datasets that support the prediction of PTEs in the soil. This is because it encompasses valid facts

and reliable data about the surface and subsurface terrain features including highlands, lowlands, fluvial erosions, and sediments contaminations, as well as landslides. These support the decision makers and end users with timely and comprehensive information required to achieve excellent predictions (Ahado et al. 2023).

Potentially toxic elements concentrations in soil significantly influence soil quality, processes, functions, and biota which consequently have negative impacts on agriculture and forest plants and productions (Ke et al. 2023; Haj Heidary et al. 2023; Zhao et al. 2017). An exacerbated content of PTEs in agricultural and forest soils diminishes the productivity of agriculture, forestry, hindering the microbial activity in the soil, renders soil poor in fertility, and it enters the food chain. (Cabral et al. 2023; Ke et al. 2023; Haj Heidary et al. 2023; Vácha, 2021; Toth et al. 2016). More particularly, increased PTEs in the soil have great risks to planetary health including human health, flora, and fauna. Various crop species, cultivation and production systems have different responses to PTE uptake from soil to plant, resulting in diverse health risks to people through the food chain (Mwelwa et al. 2023; Antoniadis et al. 2017). For example, Mwelwa et al. (2023) recently investigated biotransfer of PTEs along the soil-plant-edible insect-human food chain in Africa and concluded that high dietary intakes of PTEs is a major safety concern for consumers in the region. The prevailing idea of health risk assessment in all areas of interest is mainly on the average values of the study area, vis-a-vis the maximum and minimum values. However, these kinds of health risk assessments rarely give a total scenario of the health conditions of the study area, but a reasonable knowledge about it. In contrary, applying a sample location technique to investigate health in each area of interest is a novel approach in health risk assessment that gives a detailed understanding of the area of interest as well as comprehensive data and information on health risk status for every sampled point (Agyeman et al. 2021).

To fully comprehend, assess, analyze, monitor, and regulate PTEs pollution needs good knowledge of the sources of PTEs in the area(s) of interest. Different DSM tools and approaches and their combinations have been proved highly effective and efficient. These include Principal component analysis (PCA), MLR, MLA, geostatistics, RS, GIS and others (Fei et al. 2019). Further, the positive matrix factorization (PMF), developed by the United States Environmental Protection Agency (USEPA, 2014), has also been identified as a highly effective receptor model that can quantitatively compute the contributions of potential sources to soil PTE pollutions at every data location under nonnegative constraints and data uncertainty (Yang et al. 2023; Kim et al. 2023; Anastasopolos et al. 2023; van der Westhuizen et al. 2023; Vieira et al. 2023; Huang et al. 2018; USA- EPA 2018; Sowlat et al. 2016). However, the application of the PMF receptor model has its own drawbacks which results in disparities between measured and predicted PTE content, and consequently creates impacts on PTEs factor contributions (Brown et al. 2015; Paatero et al. 2014). To increase prediction efficiency and enhance model's performances in DSM, there is the necessity to integrate the modeling method with a different model. In respect of this, ecological risk was joined with PMF to produce ER-PMF, which supplies less differences and has a higher likelihood of minimizing errors between measured and predicted PTE content. This is a novel approach that gives sustainable and effective room to decrease to the minimal rate errors emanating from source distribution assessment.

Though, studies have affirmed that some metals including Zn, Mn and Fe are essential for plants survival and productivity, yet when these metals become essence in the soil they mar and harm the growth and development of the plants (Kalkhajeh et al. 2021). The irony of it is that the concentration of these metals is at times difficult to identify if alternative auxiliary datasets are not employed to predict their presence in soil. However, the introduction of covariates data including

the satellite imageries containing the soil properties, terrain attributes, as well as PTEs concentrations and distributions, with the application of PTEs prediction algorithms such as kriging and machine learning algorithms, the challenges associated with lack of data becomes significantly resolved. Auxiliary dataset from Landsat images can easily be combined through the fusion processes to enhance the PTEs prediction efficiency which is a new development in PTEs assessment and prediction (Agyeman et al. 2022).

1.4 Digital soil mapping (DSM)

Digital soil mapping, which can also be referred to as predictive soil mapping, is at the moment the best approach to predict the spatial distribution, variation and mobility of soil or sediment at given locations (McBratney et al. 2003). The potential and need for DSM and/or predictive soil mapping has become a promising subdiscipline in soil science and agronomy (Minasny and McBratney 2016). Though, studies have revealed that the spatial variability of soil attributes within or between soil types is, in most cases, inherent because of geological and pedological soil formation processes (Iqbal et al. 2005), but some of the variability could be related to human management systems. At temporal and spatial scales, the drivers and processes influencing soil variability interwove to produce a unified content which became later well refined and adjusted by the spatial heterogeneity deposition of soil properties. It has been demonstrated by Zhu et al. (2016) that environmental covariates and soil interactions in spatial predictions are shaped with a model and the comprehended association and are readopted to spaces or areas where soil is unspecified. The essence of DSM is the formation of a quantitative soil domain nexus hinged on the modelling points or sample observation points to define the interactions between soil and other environmental covariates including vegetation indices, climate parameters, geological elements, slope, and topographic wetness index (Penizek & Boruvka, 2008). As a perfect tool, DSM adopts models to

determine soil property values at uncertain locations (McBratney et al. 2003a; Heung et al. 2016; Minasny and McBratney, 2016; Zhu et al. 2001). According to Arrouays et al. (2014), the soil science world has embraced DSM for mapping and modeling soil attributes, classes, and types and, to a large extent, to predict the content of PTEs in the soil. Many stake holders in soil studies including FAO, UNEP, and soil societies have adopted DSM because of its high accuracy relative to the conventional soil mapping. As a unique method with reliable and high precision in the spatial prediction of PTEs, and consistency for sustainable land management, DSM has become a globally accepted approach (Padarian et al. 2019). The potential of DSM has been enhanced by the progression of geographical information technology and computational technology (Zhang et al. 2017). According to Lagacherie and McBratney (2006), DSM is the identification, building and collation of spatial soil information systems by numerical models referencing the spatio-temporal indicators of soil features from soil observation and knowledge from related environmental variables. The accruing of PTEs in the soils has been a global issue of paramount concern (González-Macías et al. 2006; Liu et al. 2003), because it presents visible threat to planetary health (human and environment) (Chen et al. 2015). Chen et al. (2009) affirmed that DSM is key possible goal of studies is the inhibition of PTEs in the soil. On the other hand, spatial prediction of PTEs gives an opportunity to identify and delineate the distribution of PTEs, their contents, occurrence and defining their source(s) of pollution.

The DSM approaches are related to many methods such as conventional statistical approaches, machine learning algorithms (MLA), geostatistical methods as well as the hybrid methods (Chen et al. 2019). The routine approaches are the adoption of regularly used non-spatial statistical methods namely multiple linear regression (Jiang et al. 2019), partial least square regression (Lago et al. 2021), generalized linear models and linear mixed models (Doetterl et al. 2013). The

statistical approaches became famous in the modelling of soil organic carbon (SOC) (Gomes et al. 2019), PTEs (Ballabio et al. 2018) and soil attributes (Shi et al. 2011). In as much as such statistical techniques are not too complex to execute, their needs for reasonable and equal distribution with large datasets are seldomly challenging (Chen et al. 2019). These methods are in most cases influenced by a dearth of spatial data, which renders them less reliable and unsuitable for classifying and studying regional differences and trends (Lian et al. 2009). Other authors have reported that DSM applicability is primarily based on scientific research and are regional-based oriented (Kempen et al. 2012). Anthropogenic interventions might contaminate restricted places with well-defined boundaries from point sources or pollute broader land surfaces circumlocutorily. Meanwhile, discovering the source of PTEs is often strenuous if the point source cannot be found at the location where high contents of the element are seen (Tóth et al. 2016). In this scenario, the potential of DSM of PTEs concentrations of surface soil within given regions is crucial as it helps in analyzing spatial trends and hotspots (Tóth et al. 2016). However, the spatial variability of PTEs is naturally related to point sources of contamination, other variables including climate (wind speed and direction), and topography/terrain surface are also vital and ought to be considered (Taghizadeh-Mehrjardi et al. 2021; Costa et al. 2018). For example, Molla et al. (2023) applied DSM approaches on the assessment of soil properties and topographic features as significant covariates to improve the prediction accuracy of PTEs in soil. Similarly, Behrens et al. (2018) used DSM in multi-scale terrain feature development and their respective effectiveness for deep learning techniques. Several other studies have demonstrated extensive study of digital elevation models (DEM) and PTEs using DSM (Suleymanov et al. 2023; Agyeman et al. 2023a, 2023b; Ahado and Nwaogu 2023; Costa et al. 2018).

1.4.1 Inverse distance weighting (IDW)

Inverse distance weighting is a geospatial based digital soil mapping method where interpolation explicitly works on the assumption that things which are closer to each other are more alike than those which are farther apart (Agyeman et al. 2021). Greater weight will be assigned to the points which are closest to the target location, and hence the allocated weights change as an inverse function of 'pth power of distance', where power function (p) is a positive real number. Greater values of p grant greater influence on values which are closest to the point to be interpolated. The parameter prediction for the target location is a summation of the product of 'allotted weights' and 'measured values' for all sites. After reviewing numerous literatures, p is taken to be 2 for the current study. However, the major disadvantages of IDW are that it is sensitive to outliers and there is no indication of error or uncertainty.

1.5 Geostatistical methods

Geostatistical methods include simple/ordinary kriging and regression kriging. The techniques are mostly applied to interpolate geographical features with considerable spatial autocorrelation of environmental parameters including climate variables, soil/sediments properties, and geological components. The application of geostatistical modelling techniques cut across wider spectra namely application in soil SOC (Bangroo et al. 2020; Wang et al. 2015; Peng et al. 2013), soil properties (Vašát et al. 2013; López-Granados et al. 2005), PTEs (Linnik et al. 2020; Łyszczarz et al. 2020) as well as investigating remote/proximal sensing images (van der Meer, 2012; Zawadzki et al. 2005). Two main benefits of geostatistical interpolation are (a) its capability to handle directional factors, such as soil pollution (PTEs), soil erosion, siltation flow, lava flow, and wind movement, and (b) the ability to surpass the lowest and highest point values. On the contrary, its

major constraints are issues of the smoothing effects of kriging, and the reality that spatial interpolation analyzes physical data in a continuous environment.

1.5.1 Kriging

Ordinary kriging is a globally used digital soil mapping technique which presents many advantages over other interpolation techniques. It interpolates using weights independent of the data, hence practically, the weights after the first estimation can be used for all data sets. Also, Kriging can be thought of as a two-step process: first, the spatial covariance structure of the sampled points is determined by fitting a variogram, and then weights derived from this covariance structure are used to interpolate values for unsampled blocks or points across the spatial field (Zimmerman & Homer, 1991). Kriging has a smoothening effect on the result where it underestimates the higher values and overestimates the lower values. Since the data has a smooth trend with lesser fluctuations, this was not a cause for concern in the study. The necessary steps to predict with kriging are uncovering the rule for getting the dependence and making estimations. The most significant contribution of kriging-based dependence estimation is the effect of 'statistical stationarity' in few cases of predicting values. Some data sets are observed to be random to the extent of contributing to a situation, in which the prediction methodology fails to establish any correlation to carry out dependable estimation.

1.5.2 Regression-kriging (RK)

The RK approach is a workhorse in the toolbox of a digital soil mapper (Keskin and Grunwald, 2018). It is a hybrid algorithm which works by combining a regression of the target soil property on environmental covariates with the kriging of the prediction residuals for predicting soil properties (Hengl et al. 2004). In this hybrid model, a regression model is used to assess the deterministic trend between the target attribute and environmental predictors, whereas stochastic

and spatially auto-correlated residuals are interpolated using a kriging function (Yigini et al. 2018). The RK is an efficient tool for predicting the spatial distributions of soil features and attributes (Ma et al. 2017).

1.6 The positive matrix factorization (PMF) model

The PMF model which is commonly referred to as the US EPA's Positive Matrix Factorization (PMF) Model is a mathematical receptor model created by the US EPA scientists that gives scientific support for the assessment, development and implementation of environmental quality standards, and environmental forensics of environmental variables on land, water, and air (US EPA 2014; 2010). The PMF Model can analyze a wide range of environmental sample data: sediments/soil, wet deposition, surface water, and ambient air. The US EPA's PMF Model decreases the huge number of variables in complex analytical datasets to integrate species called source types and source contributions. In the model, the source types are classified by relating them to tested profiles. The source contributions on the other hand are applied to quantify the quantity each source contributed to a sample. Further, the PMF has great potential for supporting robust uncertainty estimates and diagnostics.

The functionality of PMF is simple as the users provide files of sample species designated as concentrations and uncertainties together with the number of sources. The model calibrates and computes the source profiles (also called fingerprints), source contributions, as well as the source profile uncertainties. The PMF Model outputs/results are constrained to provide positive source contributions, and the uncertainty weighted difference between the observed and predicted species concentration is effectively reduced. The Model outputs are shown graphically in the users' interfaces that ease data input, visualization of model diagnostics, and exporting of the results. One great advantage of the model is that it is freely available online and does not require a license or

other software to use, as well as it can be used offline. Meanwhile, the algorithms used in the PMF Model have been peer reviewed by leading air, water, and soil quality management scientists. This model has been used globally in many studies on soil pollution and quality assessment including in the Czech Republic (Ahado and Nwaogu 2023; Agyeman et al. 2023), India (Ambade et al. 2023), China (Yang et al. 2023), South Korea (Kim et al. 2023), Canada (Anastasopolos et al. 2023), USA (Sowlat et al. 2016), South Africa (van der Westhuizen et al. 2023), Brazil (Vieira et al. 2023). The flexibility and rapid accessibility of the PMF model in assessing soil pollution is commendable. PMF has high functionality for the evaluation of the PTEs in soil. It provides significant benefit in detecting and monitoring PTEs in organic and mineral soils, thus, it is one of the best and latest models (Chen et al. 2010).

In addition to applying singly, the PMF model has been integrated with other models to investigate soil pollution. For example, Ling et al. (2022) combined PMF models with geostatistical analysis for source identification while, Wang et al. (2021) investigated the accuracy of PMF source contribution analysis with correlation analysis and principal component analysis. In sum, PMF model is good at distinguishing different types of pollution sources and has unique advantages of processing and optimizing data.

1.7 Machine learning algorithms (MLA)

The MLA is a robust approach that can tolerate nonlinearity and multicollinearity and counteract overfitting especially with low soil data and auxiliary environmental data (Gautam et al. 2011). The MLA models consist of series of methods including Cubist, Stochastic gradient boosting, Random forest, Quantile regression forest, and K-nearest-neighbor.

1.7.1 Artificial neural networks (ANN)

Neural networks are known as a combination of algorithms with a purpose of simulating human learning. The ANN techniques were initially applied in computer science and later became popular in different fields of study. Substantial volumes of simple and advanced algorithms are available in various publications. The multi-layer perceptron (MLP) neural network which has a back propagation algorithm tends to be the common algorithm (Schmidhuber 2015). In the same vein, the less complex learning algorithms with data on dependent parameters and covariates could be easily recollected and remembered. Typically, the ANN is made up of three different layers namely, input layer, hidden layer, and output layer. Each layer is linked to the neurons operating simultaneously and could change the input to the output layer. These algorithms are not new as they have been widely applied in soil analysis and mapping several years ago (Song et al. 2016). Nowadays, there has been significant improvement of ANN with the introduction of the deep learning (DL) that could reach apex level extractions from input. The development of DL was initiated by Hinton et al. (2006). A detailed overview of DL could be found in Schmidhuber (2015). The DL is ideally conducted via the supervised learning scheme. Meanwhile, only a few soil mapping studies have so far applied this novel approach. Integration of the DL and macroscopic cellular automata were performed by Song et al. (2016) to model the spatiotemporal distribution of soil moisture content.

1.7.2 Boosted regression tree (BRT)

The BRT algorithms integrate decision tree techniques and the boosting approaches. In line with random forest (RF) models, BRT attempts to fit several decision trees repeatedly to enhance the accuracy of the method. Whereas RF models apply retaining techniques permitting every input datum to possess uniform likelihood for the successive drawing of subsets, BRT employs a

24

boosting method by examining the input data in the successive building of trees. The BRTs are powerful to outliers and missing values (Elith et al. 2008) and can eventually provide a satisfactory prediction of the SOC stock in spite of the size of the study area (Ottoy et al. 2017).

1.7.3 Cubist (CUB)

Cubist implements a related approach to boosting but is called "committees", which make iterative decisions in sequence. This model employs instance and model-based coupling techniques to create a multivariate regression from training data. Quinlan (1992) and Kuhn and Johnson (2013) reported that the cubist primary value is to enhance the multiple trainings committees and augment the weight to ensure it is well balanced. Similarly, the cubist model training committees (above one mostly) boosting method shares similarities with sequential series tree development with weight-adjusted. Kuhn et al. (2013) recounted that the cubist model is typically used to apply amended depending on the number of neighbours, based on prediction rules. However, Kuhn et al. (2014) stated that the cubist regression model utilized in classification and regression is prevalent and extensively used in R as a package. The cubist model follows the same method as in random forests.

1.7.4 K-nearest-neighbor (KNN)

K-nearest neighbors is a popular data mining strategy for solving classification problems (Moreno et al. 2003). It is an excellent choice due to its flexibility and relatively fast computational efficiency. However, the main disadvantage of the K-nearest neigbour classifier is that it requires a lot of memory to store the entire sample. If the sample size is large, the response time of a sequential computer is also high (Alpaydin, 1997). Despite the memory requirement, it performs well in classification tasks on various datasets (Liang and Li 2009). In this study, the K-nearest-neighbor was performed with regression kriging in R using the caret package KNN.

1.7.5 Random forest (RF)

Random forest is a data mining technique that is commonly the ensemble learning method for classification and regression that functions by creating a multitude of decision trees at the training stage, which are later gathered to produce a single prediction (Breiman, 2001). In recent years, RF is becoming popular among digital soil mappers for predicting soil elements as evidenced by the number of articles using it. The RF is also a form of machine learning method which has an up-to-date ensemble technique using the recursive partitioning for the exploring of the association between predictors and the variable of focus (Breiman et al. 1996). The RF has the same model building process as CART, sharing similar model benefits such as flexibility with different types of datasets, and insensitive to missing data. Several trees can be developed in RF, in which variables are randomly selected and only one fraction of the parameters will be installed to recognize splits in each tree. Thus, compared with CART, the RF can give a concise error measurement and is less prone to over-fitting (Breiman 2001), which has been validated in the mapping of regional soil type (Barthold et al. 2013), soil organic carbon stocks (Were et al. 2015), and soil texture (Song et al. 2016). An additional benefit of RF could be its potential to measure the importance of variables and generate acceptable results for the noisy dataset. Various authors argue that the performance of RF is still not robust enough when comparing with other machine learning techniques (Were et al. 2015).

1.7.6 Quantile regression forest (QRF)

Quantile regression forest is a variation on the RF method (Meinshausen, 2006). It keeps track of all observation samples in each decision tree node, their average values, and their variation. It also evaluates the provisional distribution of prediction results predicted on this information (Dharumarajan et al. 2019).

1.7.7 Stochastic gradient boosting (SGB)

Stochastic gradient boosting is associated with mixture boosting and bagging. Many microscale categorization or regression trees are progressively generated from pseudo-residuals (the gradient of the loss function of the previous tree) (Friedman, 2001; 2002). At each iteration, a tree is built from a random subsample of the data (selected without substitution), which causes the model to progress incrementally. Using only part of the training data increases computational efficiency and prediction accuracy while trying to avoid data overfitting. An advantage of stochastic gradient enhancement is that no predictor variable needs to be pre-selected or transformed. It is also resistant to outliers, as the steepest gradient method highlights points related to their correct categorization. In this study, the stochastic gradient boosting was performed in R using the caret package gbm.

1.8 Environmental covariates and GIS

Environmental covariates are a set of variables consisting of an empirical quantitative function of seven environmental covariates expressed as SCORPAN model (McBratney et al. 2003). In description, S stands for soil properties, or prior knowledge of the soil at a point; C represents climate, climatic properties of the environment at a point; O indicates organisms, vegetation or fauna or human activity; R signifies topography, landscape attributes; P denotes parent material, lithology; A is for age, the time factor; and N represents space, relative spatial position. The nature and characteristics of soil at any point has been identified as the influence of the interactions of SCORPAN factors. Therefore, for effective assessment and prediction of PTEs, these environmental covariates are essential. The introduction of multiple approaches embedded in the DSM has made it easier to incorporate all the environmental parameters associated with soil thereby promoting the sustainable prediction of PTEs. The combination of remote sensing, GIS-

based models with the MLA helps to generate meaningful mathematical formula used to produce

good results in soil PTEs predictions.

2.0 OBJECTIVES

The thesis aims at examining the trends, relationships, and the concentrations of PTEs in forest and agricultural soils by using various DSM methods including geostatistical and machine learning algorithms. It consists of six papers, each with a specific objective that is in line with this primary aim.

Paper 1: Trend analysis of global usage of digital soil mapping models in the prediction of potentially toxic elements in soil/ sediments: a bibliometric review.

Objectives: The objective of this paper was to conduct a review of articles, summarize and analyze the spatial prediction of potentially toxic elements, determine, and compare the models' usage as well as their performance over time.

Paper 2: Using geostatistics and machine learning models to analyze the influence of soil nutrients and terrain attributes on lead prediction in forest soils.

Objectives: The primary goal of this paper was to investigate the possibility of predicting lead (Pb) in forest soils by combining terrain attributes and soil nutrients using geostatistics and machine learning algorithms (MLAs).

Paper 3: Spatial modelling and quantification of soil potentially toxic elements based on variability in sample size and land use along a toposequence at a district scale.

Objectives: The key aim of this paper was to apply ordinary kriging (OK), geographically weighted regression (GWR), and positive matrix factorization (PMF) to model soil Cu and Mn in the Frýdek-Místek district based on different sample sizes, topography, and land use.

Paper 4: Source apportionment, contamination levels and spatial prediction of potential toxic elements in selected soils of the Czech Republic.

Objectives: The objective of this paper was to investigate the source apportionment, concentration levels and spatial distribution of PTEs in selected soils in Frýdek-Místek District of the Czech Republic.

Paper 5: Assessment of the spatial and vertical distributions of potentially toxic elements in soil and how the concentrations differ.

Objectives: This paper aimed at assessing the concentrations of PTEs in soils by modelling their spatial and vertical distributions using a GIS-based ordinary interpolation technique, and to identify the possible sources of the PTEs using PMF model.

Paper 6: Health risk assessment and the application of CF-PMF a pollution assessment-based receptor model in the Frýdek - Místek district).

Objectives: to assess human health risk exposure, to apply a novel pollution assessment-based receptor model CF-PMF (contamination factor-positive matrix factorization) in the Frýdek - Místek district.

3.0 METHODOLOGY: GENERAL OVERVIEW FOR THE STUDY

3.1. Brief description of the study area(s)

3.1.1 Study area I

The area under study is situated in the district of Frýdek-Místek that is within the Moravian-Silesian Region in the Czech Republic (**Fig 1**). The area under study is situated within the geographical coordinates for the North and East ranges between 49.39° and 49.76° (N) and between 18.18° and 18.86° (E), at an altitude between 225 and 1323 m above sea level, characterized by a cold temperate climate and a high amount of rainfall even in dry months. In Frýdek-Místek, the summers are hot and partially cloudy, and the winters are cold, dry, windy, and mainly cloudy. Over the course of the year, temperatures usually range from -4.4 °C to 23.9 °C and are rarely below -13.3 °C or above 30°C while the average annual precipitation ranges from 685 to 752 mm (Weather Spark 2016). The survey area of the district is measured at 1208 km^2 with 39.38% of the land size designated for agricultural activities and 49.36% for forest lands (Agyeman et al. 2021). The study area comprises of the following cities: Havířov, Těrlicko, Třinec, Bystřice, Jablunkov, Mosty u Jablunkova and Hrčava, which are affected by intensive farming and active industries such as the steel industry. Třinec and Vítkovice, a part of Ostrava city where the steel industry is located, become an essential area for the assessment of PTEs distribution and health risk within and around neighboring communities (Agyeman et al. 2020).

The soils differ in their color, structure, and carbonate content. The soil shows a medium and fine texture material that is derived from parent materials. It is mostly colluvial, alluvial, or aeolian deposits. Some parts of the soil show mottles in the top and subsoil that is primarily accompanied by concretions and bleaching. The potential toxic elements pollution in the area is anticipated to occur from atmospheric deposition emitted from the steel industry nearby, vehicular emission,

abrasion from tires, and agricultural activities (e.g. pesticide and insecticide applications) (Agyeman et al. 2020).The dominant soil type is Cambisols, with minor element of Stagnosols and other soil types (Kozák 2010).

Landuse-landcover

•	Sampling points		Pastures
	Discontinuous urban fabric		Complex cultivation patterns
	Industrial or commercial units		Agric. and Natural vegetation areas
	Road and rail networks area		Broad-leaved forest
	Dump sites		Coniferous forest
	Construction sites		Mixed forest
	Green urban areas		Natural grasslands
	Sport and leisure facilities		Transitional woodland-shrub
	Non-irrigated arable land		Water bodies

Fig. 1: Study area map with sampling locations in forests in Frýdek-Místek district in the Czech Republic with CORINE landuse-landcover information.
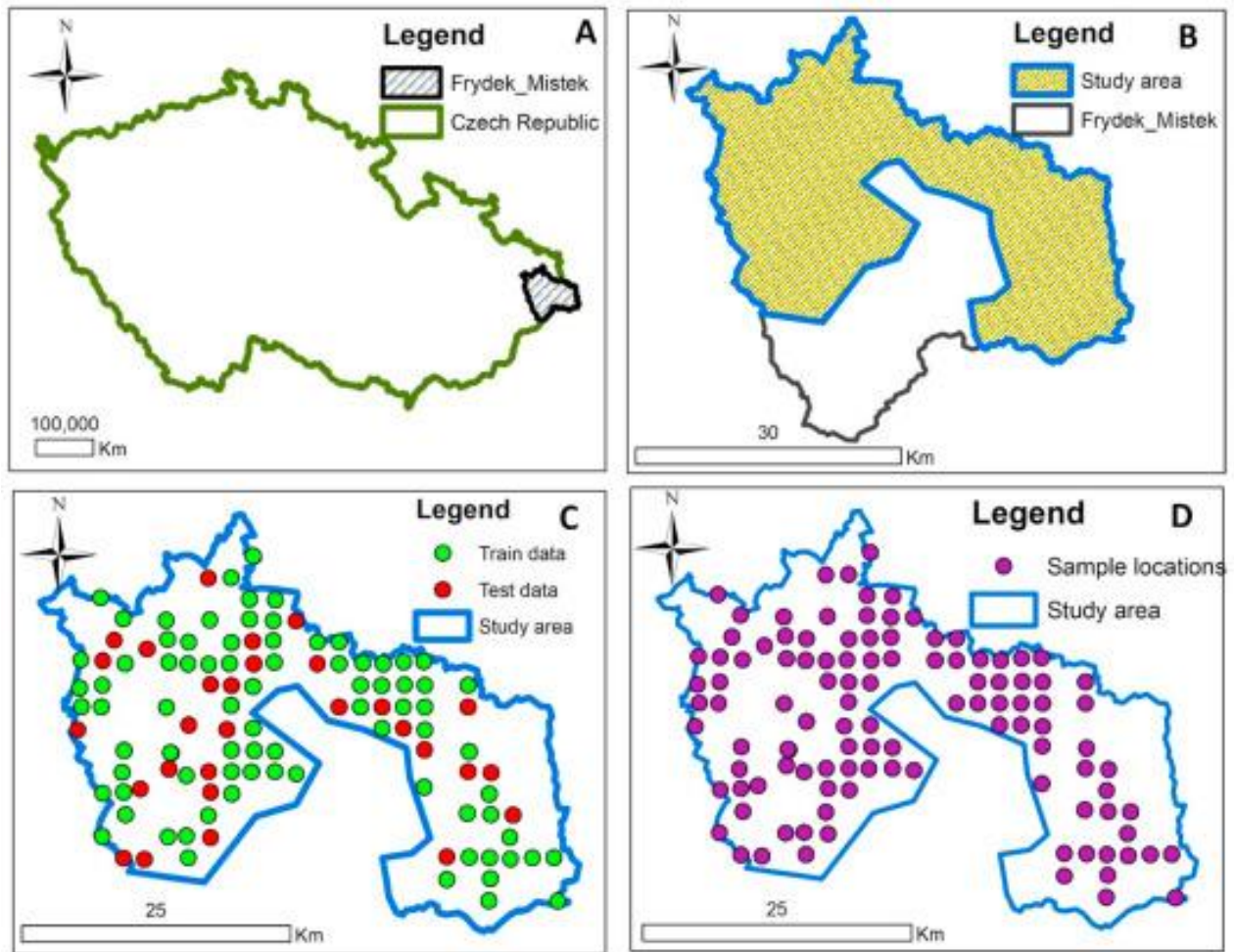
Fig. 2: (A) Study area map with sampling locations in forest soil, (B) Frýdek-Místek district, (C) Research location coupled with the partitioned dataset employed (i.e. training and test), (D) Sampling locations of the Frýdek-Místek district.

3.1.2 Study area II

This study area, which is about 110 km², covers a greater part of the Jizera Mountains region (**Fig. 3**). The altitude ranges from 600 to 1100 m above sea level. The mean annual temperatures are in the range of 3 - 6 °C contingent upon the altitude. The annual precipitation amount reaches about 1500 mm at the top of the mountains. Most areas are covered by forests (**Fig. 3**), though in some areas it recovers only slowly after forest decline in 1980s and 1990s. Norway spruce (*Picea abies*)

is the most abundant tree species. The European beech (*Fagus sylvatica*) is the main broadleaved species as the admixture in mixed forests. Purely broadleaved forests are rather rare in the area under study except the UNESCO National Nature Reserve Beech Forests of the Jizera Mountains. There are also areas having pockets of peatbogs. The PTE pollution in the area is assumed to originate from atmospheric deposits emitted from the coal, textile and steel industries, vehicular emission, abrasion from tires. Geologically, the area is characterized by granite (granodiorite) and gneiss which form the principal acidic bedrocks. Haplic/Entic Podzols, Stagnosols and Cambisols are predominant soils (Vacek et al. 2020; IUSS Working Group WRB, 2015). At most of the area especially in the higher altitudes, the mor form of humus dominates, while the moder humus type is observable only at lower altitudes.
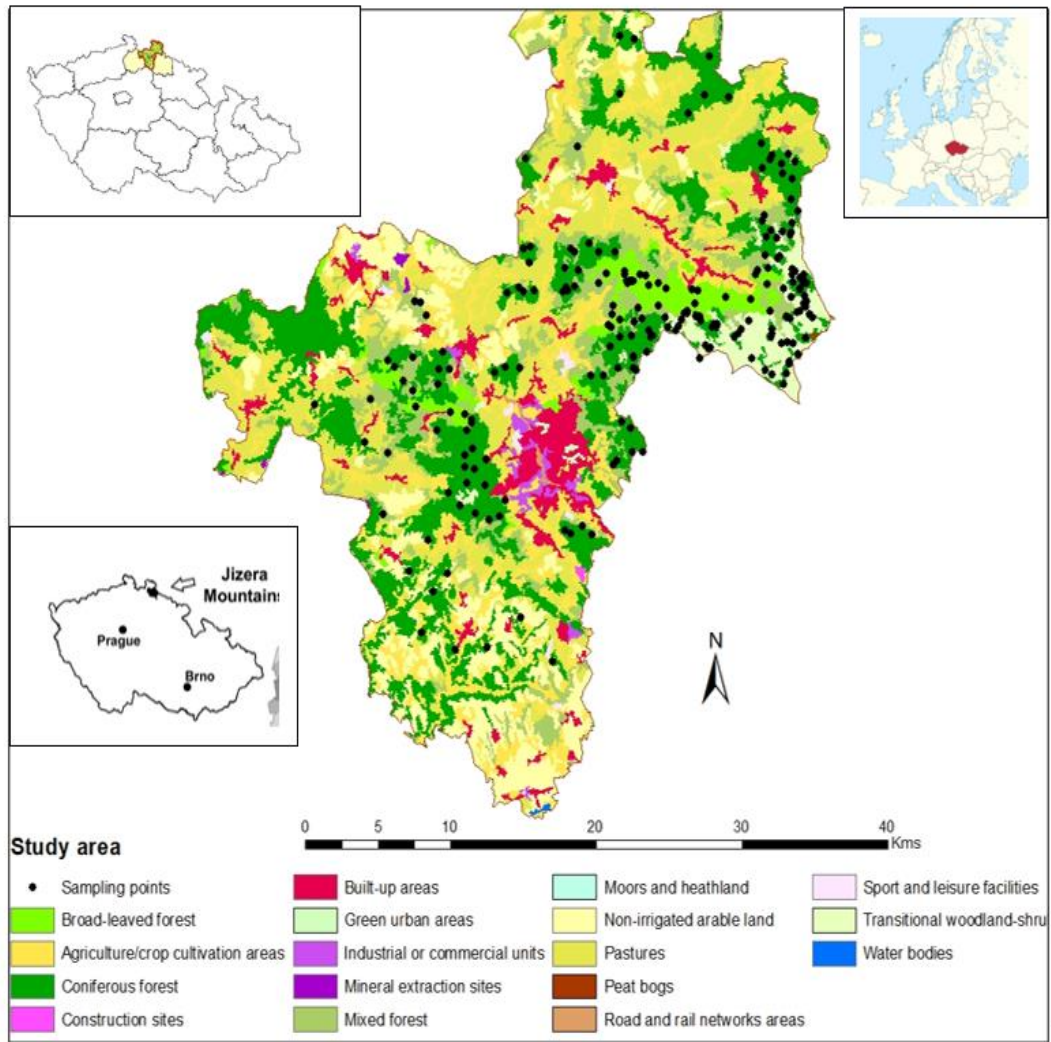
Fig. 3: Sampling points and land use–land cover of the Jizera Mts (in the Liberec district). and vicinity derived from the CORINE database (*Central Map*), location of the Jizera Mts. in Liberec region in the Northern part of the Czech Republic (*Top and down left maps*), and the location of the Czech Republic in Europe (*Top right map*).

3.2 Soil sampling and laboratory analyses

The principal source of soil data used for this study are existing soil data on Frýdek-Místek and Liberec districts and a new set of soil samples (agricultural fields – cropland and grassland) in Frýdek-Místek district collected in 2019. The legacy data will include a database of forest soil properties which consists of several thousands of sampling locations from institutional samples

collected between the years 1998 and 2019 across the whole Czech Republic, with the environmental covariates of the sampled stands (Borůvka et al. 2022). The soil sampling depth from the institutional data went through a transformation (weighted averages) into three depths: surface organic horizons (F+H), mineral topsoil (0 – 30 cm) and subsoil (30 – 80 cm; on the part of sampling sites only). The transformation is necessary because the soil data is from various depths of the soil profile and from the different surveys. Environmental covariates and pedological data of the sampled stands within the forests indicate among other covariate properties, the forest type (broadleaved, coniferous, mixed) and other land use and land cover in the area.

In Frýdek-Místek district, new sampling on agricultural land only was performed (**Fig.2**). A composite sample, composed from three mixed subsamples collected on an area of approximately 1 m^2, was collected on each place. The sampling spots were found using a handheld GPS, and samples were taken with a bucket auger, depending on the topography. There are a total number of 115 topsoil samples and 21 subsoil samples were collected across the study area. The sampling grid was 2x2 km (topsoil) and 4x4 km (subsoil).

The collected soil samples were stored in well-labeled plastic bags and taken to the laboratory. The collected samples were air-dried, grounded, and sieved with a mesh of size 2.0 mm to get a pulverized sample. In the Teflon container, 1 g of the dried, homogenized, and sieved soil sample was placed and labeled. 7 mL of 35% HCl and 3 mL of 65% HNO_3 were dispensed in each bottle of Teflon and gently closed by the cap to enable the sample to remain overnight for reactions to take place (aqua regia procedure).

The supernatant was placed on a hot metal plate for 2 h to promote digestion of the sample and left to cool when the soil sample was dissolved. The mixture was transferred into a prepared 50-mL volumetric flask and then diluted with deionized water 50 mL. The diluted supernatant was

then filtered into 50 mL PVC tubes. 1 mL of diluted concentration was further diluted with 9 mL of deionized water and filtered into a 12 mL test tube prepared to evaluate pseudototal PTE content. The ICP-OES (inductively coupled plasma optical emission spectrometry) (Thermo Fisher Scientific Corporation, USA) was utilized to ascertain the concentration of PTEs (Mn, Ni, As, Pb, Zn, Cr, Cu, and Cd) in compliance with standard procedures and protocols. Moreover, the quality control and quality assurance process were ensured by checking standard reference material (SRM NIST 2711aMontana II soil). The detection limits of the PTEs utilized in this study are 0.0002 mg kg kg^{-1}(Cd), 0.0007 mg kg^{-1}(Cr), 0.0060 mg kg^{-1}(Cu), 0.0001 mg kg^{-1}(Mn), 0.0004 mg kg^{-1}(Ni), 0.0015 mg kg^{-1}(Pb), 0.0067 mg kg^{-1}(As), and 0.0060 mg kg^{-1}(Zn). Duplicate analysis was carried out to ensure that the error was minimized. The analysis of soil samples was conducted at the Czech University of Life Sciences Prague.

In addition, the samples were also analysed using a portable X-ray fluorescence (XRF) spectrometer Olympus Delta. Each sample was measured in triplicates. The reference material was occasionally measured alongside soil samples to ensure that the analysis remained accurate until completion. The detection limits of XRF for the PTEs tested were 10 mg kg^{-1}(Ni), 10 mg kg^{-1}(Cu), 5 mg kg^{-1}(Sr), 20 mg kg^{-1}(Ba), 5 mg kg^{-1}(Ti), 10 mg kg^{-1}(Fe), 10 mg kg^{-1}(Cr), 5 mg kg^{-1}(Y), 5 mg kg^{-1}(Zr), 5 mg kg^{-1}(Th) and 5 mg kg^{-1}(Rb). The PTEs recovery percentage was 82.3 (Ni), 89.9 (Cu), 86.4 (Sr), 88.1 (Ba), 84.7 (Ti), 87.9 (Fe), 81.2 (Cr), 96.2 (Y), 92.5 (Zr), 100.9 (Th), and 98.7 (Rb) of the measured soil samples.

3.3 Spatial prediction and digital soil mapping

3.3.1 Selection of covariates

Environmental covariates were extracted from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) dataset using a digital elevation model (DEM) at a spatial

resolution of 30 m (http://earthdata.nasa.gov/search/) and processed for terrain analysis using the SAGA-GIS toolkit. In any case, a DEM with a spatial resolution of 30 m was resampled to a spatial resolution of 20 m using the bilinear resampling method in ArcGIS. The topographic parameters used in this study are slope, elevation, valley depth, LS factor, aspect and relative slope location. Terrain attributes were chosen because of their association with the response variable through a simple correlation matrix analysis.

3.3.2 Spatial modelling and digital soil mapping methods

Prior to the digital soil mapping, basic statistical parameters (mean, median, minimum, maximum, standard deviation, coefficient of variance, etc.) were first calculated for each soil property separately by horizons. The following DSM techniques were applied to analyze and model the distribution of the PTEs in the soil. These include kriging, geographical weighted regression (GWR), inverse distance weighting (IDW) and MLAs such as cubist (CUB), Quantile regression forest (QRF), Random forest (RF), stochastic gradian boosting and K-nearest-neighbor (KNN). These were applied to produce digital maps of potentially toxic elements Cr, Cu, Fe, Cd, Pb, Zn, Mn) after the log transformation of the dataset. To determine the relationship between environmental covariates and PTEs, correlation analysis was employed in Rstudio software version (4.0.5). The variogram and semivariance interpolation were used. This interpolation technique enhanced the creation of the spatial distribution maps of PTEs of the study area. ArcGIS, version 10.7.1 (ESRI 2019) was used for spatial data processing and visualization. Further, a hybrid model that coupled ordinary kriging with machine learning techniques (e.g., random forest, cubist, and extreme gradient boosting) to create a regression kriging approach was implemented. Details of the used methods are described in each paper.

3.3.3 Assessment of the accuracy and validation of the models

The study thoroughly examined the performance of the models selected for this work. The models were trained using 75% of the dataset, which was validated using the remaining 25% of the dataset. The coefficient of determination (R^2), root mean square error (RSME), mean absolute error (MAE), RPIQ (ratio of performance to interquartile range) and median absolute error (MdAE) were used to evaluate the performance, efficiency and accuracy of the various models applied in this research. R^2 describes the variability in response proportions represented by the regression model. MdAE confirms the true measured value, while RMSE and MAE determine the magnitude of the models expected accuracy. The value of R^2 should be high and the accuracy should be near to 1 to select the best model method using the validation criteria. According to Li et al. (2016), a standard R^2 value of 0.75 or higher should be considered excellent predictions, while 0.50 to 0.75 should be considered satisfactory predictions. The lower obtained values are suitable and considered the best choice for model selection for the validated standard estimation methods RMSE, MdAE and MAE. The RPIQ is calculated by dividing the interquartile range (IQ = Q3–Q1) by the RMSE and depicts the largest spread of population residuals (Bellon-Maurel et al. 2010).

3.4 Contamination level analysis for PTEs

The PTEs pollution level of the study area was examined through various contamination assessment indices, including the contamination factor (CF) and the pollution load index (PLI).

3.4.1 Contamination factor (CF)

CF is defined as the ratio of the element content of the sample to the background value of the same metal. It is given by:

$$CF = C \text{ (metal) Sample} / C \text{ (metal) background value} \qquad (1)$$

Where C (metal) is the concentration of metal analyzed from sampled soil; C (metal) background value is the geochemical background value (or concentration) of that metal. It is important to state here that the baseline values used were the World mean values (Kabata-Pendias, 2011).

3.4.2 Pollution load index (PLI)

The PLI is a measure of estimation which was first proposed by Tomlinson et al. (1980). Pollution load index has been in use for the detection of pollution. It is robust and effective in the comparison of pollution levels in space and time. The PLI was calculated based on the concentration factor of each PTE by focusing on the background value in the soil where CF is the contamination factor earlier stated (Eq. 1) and 'n' is the number of metals studied. The pollution load index (PLI) equation is given by:

$$\text{PLI} = \sqrt[n]{\text{CF}_1 \text{ x CF}_2 \text{ x CF}_{,3} \text{ x } ... \text{ x CF}_n} \tag{2}$$

Where CF is the contamination factor derived for each metal; n is the number of metals. A PLI less than 1 indicates the optimal soil quality, and a PLI that is equal to 1 signifies that only the baseline levels of contaminants are present, while a PLI greater than 1 infers the degradation of the quality of the site by pollution (according to Tomlinson et al. 1980).

3.4.3 Source apportionment using positive matrix factorization (PMF) model

The positive matrix factorization (PMF) model is effective by using the software EPA-PMF v. 5.0 (U.S. EPA 2014). It was applied to determine the contribution of PTEs sources to contamination in the study area. The mathematical method is a receptor model used in calculating the contribution of the sources to samples built on the composition or fingerprints of the sources. The PMF model is given by:

$$C_{ij} = \sum_{k=1}^{p} G_{ik} + F_{kj} + E_{ij} \tag{3}$$

In which C_{ij} represents the content of PTE j in soil sample i, p is the number of factors (i.e., pollution sources), G_{ik} denotes contribution of factor k to soil sample i, F_{kj} is the content of PTE j in factor k, and E_{ij} signifies the residual. Assume the matrix of PTEs contents in the soil samples; PMF can derive the factor contribution matrix G as well as factor profile matrix F by minimizing the objective function Q, which is presented as:

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{E_{ij}}{U_{ij}} \right)^2 \qquad (4)$$

In which m is the number of PTEs investigated, n denotes the number of soil samples, E_{ij} signifies the residual and U_{ij} signifies the uncertainty of PTEs j in soil sample i. U_{ij} is determined based on the PTEs content (C_{ij}), the relative standard deviation (σ), and the method detection limit (C_{MDL}). Therefore, it implies that the PTEs content is above C_{MDL} value; U_{ij} is computed as:

$$U_{ij} = \sqrt{\left(\sigma \times C_{ij} \right)^2 + C_{MDL}^2} \qquad (5)$$

The model recommends that the data below the detection limit would be substituted with the value of $C_{MDL}/2$, i.e., data that does not occur in this study, and the associated uncertainty is calculated as:

$$U_{ij} = \frac{5}{6} C_{MDL} \qquad (6)$$

Besides the constraint of no significant negative contribution (G_{ik}), the utmost optimal factors were derived using the multilinear engine algorithm in PMF. It is vital to note that the minimum Q can be global or local. Consequently, multiple attempts using diverse starting points were carried out to reach the global minimum Q and a reliable solution.

The uncertainty and the bias of the results obtained were further estimated using two error estimation methods encompassed in the PMF model: classical bootstrap (BS) and displacement of factor elements (DISP) (Brown et al. 2015). The uncertainty of solutions can arise generally based on three causes: (1) errors coming from the data set randomly, which are based on measurement procedures; (2) ambiguity resulting from rotation and the fact that multiple PMF solutions can have the similar or very close values of object function Q; and (3) modelling errors triggered by the simplification of the real system. Bootstrap was performed to curb random errors, whereas displacement was engaged to explore data errors and rotational ambiguity. The bootstrapping allowed us to create a new data set that randomly selects subsamples from the original observation with replacement. This set of data generated was then fitted by the PMF model to derive various factors and the concentration contribution. The fingerprint and the factor contribution of each element in the soil were assessed based on the results generated by the software from repeating the resampling and model fitting procedure. The object function Q would then increase by a predetermined maximum value change dQ_{max}. The interpretation is based on the adjustment of the upper or lower interval of the displaced element.

3.5 Health risk assessment

The presence of industries, productive agriculture, and other anthropogenic factors exposes individuals within the study area enclave to risk posed by PTEs. Based on the risk of people being exposed, inhalation, ingestion and dermal intake are three known pathways that the residents can be exposed to. According to Chen et al. (2015), an individual is exposed to soil PTEs by three main pathways: ingestion, inhalation, and dermal contact of soil. This is specified in the following equations below.

$$CDI_{ing} = \frac{C \times IR_{ing} \times EF \times ED}{BW \times AT} 10^{-6} \qquad (7)$$

$$CDI_{inh} = \frac{C \times IR_{inh} \times EF \times ED}{PEF \times BW \times AT} \qquad (8)$$

$$CDI_{derm} = \frac{C \times SA \times AF \times ABS \times EF \times ED}{BW \times AT} \times 10^{-6} \qquad (9)$$

$$CDI_{total} = CDI_{ing} + CDI_{inh} + CDI_{derm} \qquad (10)$$

The definition of the parameters for *(CDI,* mg kg^{-1} d^{-1}), CDI_{ing} CDI_{inh} and CDI_{derm} as well as reference values of the parameters in the equations (especially, equations 7-10) are listed in the **Table 1**.

3.5.1 Non–carcinogenic risk assessment

The potential non-cancinogenic risk of one PTE was determined as the hazard quotient (*H*Q), in which the equation is given by:

$$HQ = \frac{CDI_{total}}{RfD} = \frac{CDI_{ing} + CDI_{inh} + CDI_{derm}}{RfD} \qquad (11)$$

Where RfD implies the reference dosage (mg kg^{-1} d^{-1}) and is the projected daily human exposure.

The computational HQ values were used to assess the detailed risk to health of all the PTEs studied.

The values were summed up and expressed by equation as a hazard index (HI), which is given by

equation 12:

$$HI = \sum HQ = HQ_{ing} + HQ_{inh} + HQ_{derm} \qquad (12)$$

In which HQ$_{ing}$, HQ$_{inh}$ and HQ$_{derm}$ correspond to the hazard quotient for ingestion, inhaling and dermal. USEPA report (US EPA 1989) specifically asserted that if HI < 1, there is no possibility to have a negative effect upon a person's health who is exposed to PTEs, and if the HI is >1, adverse health effects are possible.

3.5.2. Carcinogenic risk assessment

The US EPA report (EPA 2002) stressed that a human exposed to carcinogenic risk (CR) could be attributable to the probability of cancer development of any type. Equations 13 and 14 were used to calculate for PTEs such as As, Ni and Cr carcinogenic risk.

$$CR = DCI \times SF \tag{13}$$

$$TCR = \sum CR = CR_{ing} + CR_{inh} + CR_{derm} \tag{14}$$

Where CR, TCR and SF values symbolize carcinogenic risk (no unit), total carcinogenic risk (no unit)and slope factor for carcinogenic risk, DCI denotes daily chronic intake of PTEs (mg kg^{-1}d^{-1}) respectively. The acceptable threshold value of CR is $1.0.10^{-4}$ and the range of tolerable values of CR are from $1.0 \; 10^{-6}$ to $1.0 \; 10^{-4}$ (US EPA, 2011). This is the acceptable criterion which shows that human health is not significantly endangered.

Table 1: Exposure factors used in CDI estimation for non-carcinogenic and carcinogenic risk.

Variables	Description	Units	Values	Values	
			Adults	Children	
C	Concentration of PTE of present study	mg/kg			
IR_{ing}	Ingestion rate	mg/d	100	200	US EPA, 2011
IR_{inh}	Inhaling rate	m^3/d	20	7.65	US EPA, 1991
EF	Exposure frequency	days/year	350	350	US EPA, 2011
ED	Exposure duration	year	30	6	US DOE, 2011
SA	Skin surface area	cm^2	5700	2800	US EPA, 2002
AF	Soil adherence factor	mg/cm^2/d	0.07	0.2	US EPA, 2011
ABS	Dermal absorption factor	mg/kg/day	0.001	0.001	US EPA, 2011
PEF	Particle emission factor	m^3/kg	1.36×10^9	1.36×10^9	US EPA, 2011
BW	Average body weight	kg	70	15	US EPA, 2013
AT N-C	Average time for non-carcinogenic risk	day	ED $\times$ 365	ED $\times$ 365	Wang et al., 2017;
AT CR	Average time for carcinogenic risk	day	70 $\times$ 365	70 $\times$ 365	Eziz et al., 2018; Wu et al., 2019
FC	Units' conversion factor	$kg.mg^{-1}$	1×10^{-6}	1×10^{-6}	US EPA, 2002
Specific reference dose for ingestion	RfD_o	mg/kg/day	Cr (3×10^{-3}), Cu (4.0×10^{-2}), Ni (2×10^{-2}), Pb (3.50×10^{-3}), Zn (3×10^{-1}), As (3×10^{-4}) and Mn (0.14)	Li et al. 2015; USDOE 2011; Qing et al. 2015; De Miguel et al. 2007; Teng et al., 2015	
Specific reference dose for dermal contact	RfD_{ABS}	mg/kg/day	Cr (6×10^{-5}), Cu (1.2×10^{-2}), Ni (5.4×10^{-3}), Pb (5.3×10^{-4}), Zn (6×10^{-2}) and As (01.23×10^{-04}) Mn (0.05)		
Specific reference dose for inhalation	RfD_i	mg/m^3	Cr (2.86×10^{-5}), Cu (4.02×10^{-2}), Ni (2.06×10^{-2}), Pb (3.52×10^{-3}), Zn (3×10^{-1}) As (0.3) and Mn (0.8)		
Oral slope factor	SF_o	$(mg/kg/day)^{-1}$	As (1.5)		
Absorbed dose slope factor	SF_{ad}	$(mg/kg/day)^{-1}$	As (3.66)		
Inhalation slope factor	SF_i	$(mg/m^3)^{-1}$	Cr (0.5), Ni (0.84), and As (15.1)		

3.6 Materials and methods summarized for each paper.

3.6.1 Methodology 1

Trend analysis of global usage of digital soil mapping models in the prediction of potentially toxic elements in soil/ sediments: a bibliometric review

This study adopted the bibliographic review method to enhance the understanding and close the knowledge gap on the trends of scientific publications on DSM in the prediction of soil PTEs. This was conceived in this work to get the overview of existing studies and publications on the main topic of this thesis. In this paper, various articles were collected from different scientific databases such as Scopus, Web of Science, and Google Scholar. Some keywords were identified which were geared towards the topic under review, and the keywords were tuned consistently to ascertain articles that are relevant to the study. The combined keywords used for the search engine included Soils*/sediments* AND ("digital soil mapping" OR "digital mapping" OR "spatial prediction" OR "predictive mapping") AND ("toxic elements" OR "risk elements" OR "heavy metals" OR pollution OR Cadmium OR Copper OR Chromium OR Manganese OR Nickel OR Lead OR Iron OR Zinc). This paper made use of only completed articles that were duly published within the ambit of prominent journals of which all the required information needed was collected for the analysis. No unpublished papers or articles were included in this review. The period for this review spans from 2001 to the first quarter of 2019. Over 523 articles were downloaded, and thoroughly read through to ascertain 319 articles that were relevant to the study based on these following criteria:

i. Spatial prediction or interpolation by predictive models

ii. Potentially toxic elements (PTEs)

iii. Soil or sediments-based analysis

iv. Geostatistics model or

v.   Remote sensing or

vi.   Machine learning model.

Data filtering and mining: In the data mining and filtering stages, several variables were considered, such as spatial prediction using either machine learning algorithm or geostatistics. Further, under each variable, additional subtopics were included to help in collecting the vital information and to have numerous as well as copious data to analyze. 319 articles were carefully reviewed while some articles were removed based on the quality and relevance to the topic. Finally, a total of 238 papers that are highly relevant to the research, comprising 208 soil-based analyzes and 30 sediments-based research were selected. Below are the headings and information used in the data filtering and mining process (**Table 2**).

Table 2. List of information/options on which information was collected from papers on PTEs modeling and mapping.

Headings	Information collected, options
Paper identification	authors, journal, and year
Auxiliary data-remote sensing	remote sensing (RS)–satellites, RS–airborne, RS – drones, laboratory spectroscopy (VisNIR), laboratory spectroscopy (MIR), field spectroscopy (VisNIR), field spectroscopy (MIR), other)
Region	Europe w/o Russia, Northern America, Southern America, Africa, Australia, Russia, China, Asia (without China)
Objective	mapping, temporal changes/development, spatial distribution/variation assessment, comparison of methods, other
Studied PTEs	Arsenic, Beryllium, Cobalt, Cadmium, Chromium, Mercury, Molybdenum, Nickel, Lead, Zinc, Antimony, Other

PTEs source	geology, atmospheric deposition, mining, metallurgy, industry, transportation, urban, agriculture (pesticides, fertilizers…), flooding, alluvial areas, other
Modeling methods	inverse distance weighting (IDW), geostatistics, kriging (ordinary, simple, universal), regression-kriging, fuzzy methods, multiple linear regression (MLR), cubist, polynomial models, Bayesian methods, (boosted) regression trees (RT), classification/decision trees, random forests (RF), partial least squares regression (PLSR), principal component regression (PCR), artificial neural networks (ANN), multiple additive regression splines (MARS), support vector machines (SVM), other machine learning techniques
Uncertainty assessment	leave-one/group-out, validation subsample, independent validation samples, number of validations, maximum R^2 obtained (validation), minimum root mean square error (RMSE)–validation (wt), minimum root mean square error (RMSE)–validation ((t ha^{-1})

3.6.2 Methodology 2

Using geostatistics and machine learning models to analyze the influence of soil nutrients and terrain attributes on lead prediction in forest soils.

In this study, from the legacy database of forest soil properties which comprised of thousands of institutional samples collected throughout the Czech Republic between 1998 and 2019, as well as topographic attributes of the sampled forests, data for the Frýdek-Místek district were extracted, in total 336 sampling points. The depth of topsoil samples used for the legacy data was 0-30 cm. Environmental covariates and pedological data derived from forest sampled stands within the forest indicate the forest type (broadleaf, coniferous, mixed forest) and other local land uses and land covers. The contents of Pb, Cd, Cr, Cu, Mn, Zn, Fe and other elements in the soil were obtained by the standard aqua regia method (ISO 11466:1995, 1995). Geostatistics technique was

applied to examine and predict the values related to spatial heterogeneity of physical processes. This included the spatial and temporal coordinates of the datasets. The geostatistical approaches helped as a good approach to distinguish spatial features and linear interpolation values for areas in which samples have not been collected. For example, the study used regression kriging (RK) as a Geostatistics technique to hybridize with machine learning approaches for mapping and predicting lead (Pb) contents in soils. This study also employed machine learning algorithms (MLA) including Cubist, Stochastic gradient boosting, Quantile regression forest, and K-nearest-neighbor for efficient modelling, mapping, and prediction of the soil PTEs in the study. Further, since the PTEs (e.g., Pb) are associated with topographic features, environmental covariates were extracted from the Aster dataset using a digital elevation model (DEM) at a spatial resolution of 30 m (http://earthdata.nasa.gov/search/) and processed for terrain analysis using the SAGA-GIS toolkit. In any case, a 30 m DEM with a spatial resolution of 30 m was resampled to a spatial resolution of 20 m using the bilinear resampling method in ArcGIS. The topographic parameters applied in the study are slope, elevation, valley depth, LS factor, aspect, and relative slope location. Terrain attributes were chosen because of their association with the response variable(s). To determine the prediction accuracy of the used models, the dataset was split into calibration (75%) and validation (25%) datasets. The calibration dataset was used to establish the prediction models, and validation dataset was adopted for independently evaluating the prediction accuracy. The coefficient of determination (R^2), root mean square error (RSME), mean error (MAE), RPIQ (ratio of performance to interquartile range) and median absolute error (MdAE) were Used to evaluate the models' efficiency and accuracy of the methods adopted in this study.

3.6.3 Methodology 3

Spatial modelling and quantification of soil potentially toxic elements based on variability in sample size and land use along a toposequence at a district scale.

Legacy soil data and current soil maps for the Czech Republic were the primary sources of soil data for this study. The same set of data mineral topsoil (0 -30 cm) from 336 sampling locations in the Frýdek-Místek district as in the previous study was used. Environmental covariates and pedological data of the samples collected within the forests covered broadleaves, coniferous, mixed forests, and pollution risk forest areas. The pseudototal content of PTEs (Cu and Mn) in the soil was determined using the aqua regia standard method (ISO 11466:1995, 1995). The Rstudio software version (4.0.5) was used to further resample the soil samples into five different subsets by reducing the total sample (n = 336) by 60 from the original data set. R software code was used to model the five (5) different sample sizes, n = 336, n = 276, n = 216, n = 156, and n = 96, respectively. Given the total sample size of n = 336, each of the subsequent sub samples was randomly selected by assigning a unique seed [e.g., seed (276)] to ensure reproducibility of results. The various sample sets were randomly generated together with the x, y coordinates, respective terrain height equivalent, and associated land use-land cover (LULC) as derived from the field. The sequential reduction of the number of models by 60 was based on the author's discretion as it could have been any other number. Basic statistical parameters such as mean, median, minimum, maximum, standard deviation, coefficient of variance, kurtosis, and skewness were determined. The OK interpolation technique was used to enhance the creation of the spatial distribution maps of the PTEs using the non-transformed values of Cu and Mn. Furthermore, the positive matrix factorization (PMF, EPA version 5.0, Washington, DC, USA) was used for the estimation of source apportionment and contamination level of the PTEs, whereas the ordination model of CANOCO 5.0 was used to show the association among LULC, relief, sample size, and PTEs.

3.6.4 Methodology 4

Source apportionment, contamination levels and spatial prediction of potentially toxic elements in selected soils of the Czech Republic (Frýdek-Místek)

In this study, a grid sampling design was adopted for the soil assessment. The topsoil (0 to 20 cm) samples were collected at intervals of 2 km using a handheld GPS system (Leica Zeno 5 GPS) and subsoil (21 to 40 cm) samples at every 4 km. The total number (n) of samples obtained for topsoil (ts) and subsoil (ss) was n=49 and n=21, respectively. They were taken into plastic bags and appropriately labelled and transported to the laboratory. The samples were air-dried, crushed by a mechanical device, and then sieved (< 2 mm) to obtain a pulverized sample. These samples were analyzed using a portable X-ray fluorescence spectrometer (Delta Premium 2019). Each sample was measured in triplicates. The quality assurance and control process, the standard reference material for a portable device (i.e. XRF 2711a and NIST 2711a, the National Institute of Standards and Technology) was used in the analysis to ensure quality compliance. The detection limits for the PTEs tested were < 10 mg kg^{-1} (Ni),< 10 mg kg^{-1} (Cu),< 10 mg kg $^{-1}$ (Fe),< 10 mg kg $^{-1}$ (Cr). The PTEs recovery percentage were 82.3 (Ni), 89.9 (Cu), 84.7 (Ti), 87.9 (Fe), 81.2 (Cr), and 92.5 (Zr). PTEs contamination analysis such as contamination factor and pollution load index were performed. The statistical analyses were performed using Excel (maximum and minimum number, average value), PMF EPA 5.0 for estimation of source apportionment, and RStudio for mapping as well as estimation of the Pearson correlation matrix. Inverse distance weighting (IDW) interpolation was used in allocating weights change as an inverse function of 'pth power of distance and the similarities between neighbours with a proportional distance between them. This interpolation technique supported the creation of the spatial distribution maps of PTEs of the study area under investigation.

3.6.5 Methodology 5

Assessment of the spatial and vertical distributions of potentially toxic elements in soil and how the concentrations differ

In this study at 3 km intervals, the soil samples were collected were collected for both organic soil organic soil (F+H horizons) and mineral soil (30 cm) in the Jizera Mts. Region, Liberec district. The sampling points were located using a handheld GPS system (Leica Zeno 5 GPS) and samples collected using either a push probe or bucket auger depending on the terrain. A total number (N) of 221 samples each were derived for organic soil (F+H horizons) and mineral soil (A) across the study area. The collected soil samples were stored in well labelled plastic bags and taken to the laboratory. The soil samples were air-dried, crushed by a mechanical device, and then sieved with a mesh of size 2.0 mm to get a pulverized sample. The presence of elements such as Cr, Cu, Pb, Mn, and Fe) in the soil were extracted using aqua regia standard method(ISO 11466 :1995, 1995) (Melo et al. 2016) to determine their pseudo-total content. For the quality control (QC) of the method, the standard addition technique was adopted. For example, the QC of the concentration determination was guaranteed using the SRM 2711 (Montana II soil) reference material (National Institute of Standards and Technology, Gaithersburg, Maryland, USA). The values achieved were consistent with the reference data. The recovery differences were generally < 10% ($n = 3$). The detection limits for the elements based on the applied method were as follows: Cr (0.03 mg L^{-1}), Cu (0.015 mg L^{-1}), Pb (0.05 mg L^{-1}), Mn (0.05 mg L^{-1}), and Fe (0.15 mg L^{-1}). The source apportionment of the soil PTEs were determined by using the PMF model. To determine the relationship between environmental covariates and PTEs, correlation analysis was employed. ArcGIS, version 10.7.1 was used for spatial data processing and visualization. Maps of the spatial distribution of these soil properties were created using ordinary kriging.

3.6.6. Methodology 6

Health risk assessment and the application of CF-PMF: a pollution assessment-based receptor model in urban soil.

In this study data from 49 samples collected from the Frýdek-Místek district based on agricultural and industrial soil areas were used, the same as in paper 4. Contamination assessment indices such as contamination degree (C_{deg}), modified degree of contamination (mCd), Nemerow pollution index, PMF receptor model, and contamination factor receptor model (CF-PMF) was applied to determine and predict the soil PTEs. The presence of industries, productive agriculture, and other anthropogenic factors exposes individuals within the study area to PTEs. To assess the health impacts of the PTEs contamination, health risk assessment (HRA) was performed. The HRA was conducted by determining the non-carcinogenic risk assessment, and carcinogenic risk assessment.

4.0 SUMMARY AND CONCLUDING REMARKS

4.1 Summary of key findings and discussion

Overall, the goal of the study is to assess the spatial prediction of soil PTEs by applying DSM approaches including geostatistical and MLA in agriculture and forest soils, as well as determining the PTEs source(s). Data for the study were diverse and from different sources, namely legacy datasets, field survey data, and satellite imageries covering different variables such as the soil properties, as well as terrain parameters and land use to achieve effective PTEs predictions. The introduction of this diverse dataset promoted the expedition of distinct pathways for realizing a more accurate soil PTEs prediction for better decision making in the study area. At some given scenario, the integration of terrain attributes was coupled and/or decoupled with comparative analyses to determine the most suitable auxiliary dataset together with the modeling techniques in the PTEs prediction. The study demonstrated that the integration of diverse datasets produced more acceptable and desirable results than using a single dataset. This data harmonization approach intensified and strengthened the prediction efficiency and coherence vis-à-vis reducing negligible errors.

4.1.1 Paper 1: Trend analysis of global usage of digital soil mapping models in the prediction of potentially toxic elements in soil/sediments: a bibliometric review

The findings from this study revealed that soil pollution originates from diverse sources and clarifies the motives for authors to focus on a particular area of their interests. The results further observed the uncertainties in mapping and identified the number of publications in the subject area in the related journals. Further, the continental and global-based endeavors to study and publish on trending issues regarding DSM were also established in this study. For example, Europe accounted for the highest number of publications, followed by Asia, Africa, North America, South America, then Oceania. In terms of publishing journals, articles were extracted from related

journals including soil science, soil contamination or pollution, spatial prediction, and related journals repository. The journals with the highest publications on the topic were Geoderma (25), the Science of the Total Environment (18), Environmental Pollution (18), the Journal of Geochemical Exploration (12), Environmental Earth Science (11) and Environmental Monitoring & Assessment (10), and other recorded less than 10 articles each during the study period. Regarding the modelling methods, the study found a total of sixteen modelling approaches that are widely used in DSM as were shown during the study period. These were inverse distance weighting, kriging (simple, ordinary, universal), indicator kriging, regression, kriging, fuzzy methods, multiple (linear) regression, cubist, classification and regression trees, random forest, partial least square regression, principal component regression, artificial neural network, multivariate adaptive regression splines, support vector machine. However, the study revealed that only 9 out of the 16 models were the most used. The kriging (e.g., simple, ordinary, universal) model was the most preferred model of choice by most authors when compared with the other models. In addition, regression kriging, multiple linear regression, and artificial neural networks have been used quite a few times. Random forest, partial least square regression, multiple linear regression, inverse distance weighting was relatively adopted.

The DSM is primarily based on geostatistics and MLA techniques. The geostatistics is a set of statistical approaches that rely on spatial/temporal modelling data as well as focuses on the high-precision estimation with quantifiable uncertainty. Machine learning, on the other hand, is a relatively new set of techniques, other than geostatistics, which some schools of thought assume gives a high rate of predictive accuracy concerning the type of model used. The spread of MLA use has increased in recent years, and some of the DSM experts believe that there is a gradual paradigm shift from the use of the geostatistics model to machine learning techniques. Although

MLA is becoming more popular in DSM due to the availability of computational power, geostatistical kriging is the most commonly used in prediction (Webster and Oliver 2007). Although the geostatistics model remains the citadel of the DSM juxtaposing it with the papers reviewed, it is evident that geostatistics is still widely used in the soil sciences community. From 2011 until now, the soil science community has had a flair in the blending of the usage of MLA and geostatistics models. There is no doubt about the popularity of the kriging model and its usage in spatial prediction in DSM, as Veronesi and Schillaci (2019) have recently argued. Conversely, to Veronesi and Schillaci earlier arguments, they went on to report that practitioners relied less on interpolation (kriging models) but rather more on MLA. This assertion reinforces the fact that algorithms could be used based on the preference and diverse conditions available at the time. Numerous studies have been conducted to compare the two algorithms to determine the contrast and superiority between geostatistics and MLA. It is important to note that, for academic findings, it is possible to compare algorithms within DSM to find the best results that fit the modelling that a practitioner or author is trying to model.

In respect to the studied potentially toxic elements, this study observed that from 2001 to 2018, Cd, Cr, Cu, Pb, and Zn were the predominant in most of the soils examined by the scientists. The study showed that many papers mentioned the study of Pb (181 times), Cu (153 times), Zn (153 times), and Cd (141 times), Cr (130 times), Ni (116 times), and Mn (58 times), while other PTEs such as As, Co, Hg, V, and Mo were also mentioned many times. Among the PTEs, Pb was the most mentioned PTEs.

Considering the objectives of the authors for using the DSM algorithm, the most frequent reasons were 'for soil mapping' which accounted for 133 times in the articles (representing 31.7%), 'for

determining the extent of the pollution (i.e., for spatial distribution or variation)' which recorded 172 times in the articles, and with the largest percentage (> 40%).

In terms of uncertainty and validation for model accuracy, the study revealed that no validation was carried out on 128 articles reviewed, 46 papers used R square (R^2) validation method, 57 root mean square error (RMSE), 14 mean absolute error (MAE), 12 residual prediction deviation (RPD), 13 mean error (ME), and the rest of the validation criteria put together had 45 papers.

This paper reveals the complementary role machine learning algorithms, and the geostatistical models play in DSM. Nevertheless, geostatistics approaches remain the most preferred model compared to machine learning algorithms because MLA approaches are still new but with time more studies might consider MLA than geostatistics because of the potential.

4.1.2 Paper 2: Using geostatistics and machine learning models to analyze the influence of soil nutrients and terrain attributes on lead prediction in forest soils

This paper investigates the possibility of predicting lead (Pb) in forest soils by combining terrain attributes, soil nutrients and their combinations, and regression kriging in the Frýdek-Místek district of the Czech Republic. The results showed that the prediction of Pb in forest soil using separately terrain attributes and soil nutrients as additional data sets combined with regression kriging would produce satisfactory results. On the other hand, it performed well with the combination of soil nutrients and terrain attribute. Overall, the evaluation showed that using cubist_RK in combination with terrain attributes and soil nutrients provided the best prediction accuracy and the lowest error in Pb prediction in forest soils. Interactions between Pb and soil nutrients, as well as terrain attributes, can help to better identify sources of PTE pollution while improving predictive efficiency. It was found that applying a robust digital soil mapping (DSM) model in combination with terrain attributes and soil nutrients is efficient in predicting the spatial

distribution and estimation of uncertainty levels of lead (Pb) in forest soils based on the model's accuracy parameters. Therefore, it is established that the use of soil chemical properties as a proxy, together with appropriate environmental covariates, and DSM will improve modeling efficiency for the predictions of lead (Pb) in forest soils. This inference was drawn based on the following summarized contextual findings that: (i) when the terrain attributes were used as the only auxiliary dataset, QRF_RK proved to be the most effective method for predicting Pb in forest soil (context 1) ; (ii) when soil nutrients were used as the only auxiliary dataset, the most effective method for predicting Pb in forest soil was a combination with SGB_RK (context 2); (iii) combining cubist_RK with an ancillary dataset of soil nutrients and terrain attributes together is the most effective method for predicting Pb in forest soils (context 3). In sum, this finding is assumed to be a contributor to the existing knowledge on DSM using geostatistics and MLAs under forest soils.

4.1.3 Paper 3: Spatial modelling and quantification of soil potentially toxic elements based on variability in sample size and land use along a toposequence at a district scale

This study applied ordinary kriging (OK), geographically weighted regression (GWR), and positive matrix factorization (PMF) to model soil Cu and Mn in the Frýdek-Místek district based on different sample sizes, topography, and land use. It was found that the predicted PTEs concentrations and their spatial distribution in the soil between the different calibration sample sizes were highly heterogeneous. The PTEs in all the samples had a clear positively skewed distribution: Cu (n = 336) having the highest skewness of 4.46 and Cu (n = 96) having the lowest skewness of 1.04. The skewness value for Mn ranged from 2.37 to 1.58. The mean values obtained for both Cu and Mn were lower compared to the European mean values, worldwide mean value (Kabata-Pendias 2011), and Crati basin values (Guagliardi et al. 2012). This may suggest that the soil had low pollution.

The northern parts of the study district were found to have the highest concentration (hotspots) of all PTEs. This could be attributed to the fact that the socio-economic and industrial activities are concentrated towards the northern part of the study area, and thus toxic elements are regularly discharged (Wang et al. 2020; Guo et al. 2019). This study further revealed spatial dependency values in relation to the structure of the PTEs pollution in the study area, a trend that might be attributed to a typical soil-geochemical patterns of PTEs pollution formation by the disposal of human-made waste in the forest soils (Xu et al. 2016). Further, the study established that the prediction accuracy of PTEs was not dependent on total sample size but on other environmental variables such as elevation and land use are involved. For example, all the sample sizes showed the highest concentrations of the PTEs in the lowlands below 500 meters where the industrial, commercial and the arable activities dominated in contrast to the highlands (above 500 m) where forests were dominant. Therefore, the selection of appropriate prediction models will enhance the achievement of accurate prediction of PTEs spatial distribution in soils.

The result further established that the concentrations of Cu and Mn in the soils of the study area were within the reference range recommended by EU, and/or US-EPA which are EU (50–140 mg kg^{-1}), and US-EPA (28-80 mg kg^{-1} average) for Cu, and EU (524 mg kg^{-1} on average), and US-EPA (< 750 mg kg^{-1} average) for Mn, respectively. Although the rate of pollution according to the mean values indicated non-polluted soils, there may be an intermittent need to assess the soils for control measures to be taken to curtail excessive accumulation and escalation to safeguard the wellbeing of the inhabitants and the ecosystem. Also, the results might support policy-developers in sustainable farming and forestry for the health of the ecosystem towards food security, forest safety, as well as animal and human welfare. The research also concluded that sample size does not necessarily influence the quantification of spatial variability and predictive accuracy of PTEs,

rather other environmental factors such as covariates, land use, topography and the selection of appropriate models influence the prediction and predicted distribution of these elements in soils. Different subsets with the same number of samples can provide different results.

4.1.4 Paper 4: Source apportionment, contamination levels, and spatial prediction of potentially toxic elements in selected soils of the Czech Republic

This study investigated the spatial and vertical distribution of the PTEs in the soil by analyzing the concentration at the topsoil (0-20 cm) and subsoil (21-40 cm). The result established that both topsoil and subsoil showed homogeneous variability. Ni showed high variability in both the topsoil and subsoil, whereas Fe showed high variability in the topsoil only. The spatial distribution of the non-homogeneity of Ni and Fe predicts the existence of a local source of enrichment. Comparing the findings from this study with the PTEs from Crati Basin (Guagliardi et al. 2012), it is apparent that the Crati Basin is highly enriched in Cr, Ni,Cu, Rb, Sr and Ba compared to the current studies. However, the current study is also enriched in Y, Zr, Ti and Fe compared to the Crati Basin.

The findings revealed that the area consisted of low to high pollution sites. For the topsoil, the average PTE concentration decreased in the order of: Fe>Ti>Ba>Zr>Rb>Sr>Cr>Y>Cu>Ni>Th, while the subsoil decreased in the order of Fe>Ti>Zr>Ba>Rb>Sr>Cr>Y>Cu>Ni>Th. The result further revealed that Cr, Ni, Cu, Rb, Y, Zr, Ba, Th, and Fe were far above both the European average value (EAV) and the World average value (WAV) of baseline/level.

The PTEs average value concentrations of Cr, Cu, Y, Ba, Th, and Fe in the topsoil are higher than the PTEs in the subsoil. On the other hand, the concentration of Ni, Rb, Sr, Zr, and Ti in the subsoil were higher than in the topsoil. The enrichment of the topsoil by the relevant PTEs could be due to anthropogenic activities such as atmospheric deposition, steel industry, and vehicular emission

as well as agriculture (Jones et al. 2019). In contrast, the enrichment of the subsoil by the dominant PTEs might be attributed to geogenic origin, pedogenic processes, and leaching (Li et al. 2017).

The source apportionment showed the dominance of Cr, Ni, Rb, Ti, Th, Zr, Cu, and Fe in the topsoil while in the subsoil, all the PTEs showed similar pattern of factor loadings except for Ba. Therefore, it is essential to highlight that there should be periodic monitoring process to put structures in place to discourage the release of these PTEs into soils in order to safeguard the health of humans, flora, and fauna. It is incumbent to articulate from the finding ascertain that the soil health in the study area is under threat and therefore calling for attention and actions for the reduction and the prevention of PTEs contamination, which serve as an imminent threat to the environment. In sum, it is highly recommended that further studies be performed to assess the ecological and health risk of the PTEs in the soil of the Moravian-Silesian Region in Frýdek-Místek district.

4.1.5 Paper 5: Modeling and assessing the spatial and vertical distributions of potentially toxic elements in soil and how the concentrations differ

The findings from this study established the evidence and supported the notion that the Jizera Mountains area is within the "Black Triangle" which is affected by industrial activities connected with the extraction and exploitation of coal and other natural resources in Central Europe. The content of Pb exceeded the mean values for Europe and the world (Kabata-Pendias 2011).

The application of ordinary kriging approach for the spatial analysis of the PTEs gave a clear indication of the spatial distribution and hotspots of the study area. The basic summary statistics indicate that in the organic soil, the elements increased in the order: Fe< Pb< Cu<Cr<Mn .It was further observed that all the elements with exemption of Mn indicated significant relationship with altitude above sea level in both the organic soil and the mineral soil horizons. This result is in

tandem with a study conducted by Nozari et al. (2024) in the study area. Similarly, studies exist in other regions and countries (Gerdol and Bragazza, 2020 and Ding et al.2017). For example, in the Suxian district of China, it was revealed that PTEs concentrations decreased at low elevation but increased considerably with increasing elevation (Ding et al. 2017). Other studies have reported that fine-particle elements including Cr and Cu accumulate more at lower elevations (Liu et al. 2020).

It was also established in the study that for both organic soil and the mineral soil horizons, the northern part of the kriged maps revealed higher contents of the PTEs when compared with either the southern or central part. This could be explained by the historical distribution pattern of industrial and agricultural activities in the study area which were mainly located in the northern part (Černik et al. 2016).

Application of the positive matrix factorization (PMF) model, ArcGIS-based ordinary kriging, and contamination level analysis were effective for the source identification, hotspot location, and assessment of the contamination level of the PTEs. By introducing these techniques, it was further observed that except for samples 40, 65, and 174 in the organic soil, all the 442 sample points revealed low contamination by the PTEs, in contrast to the fact that the Jizera Mountains area is indeed within the "Black Triangle" which is affected by industrial activities (Borůvka et al. 2020; Marx et al. 2017). However, the level of pollution in the area is quite low based on the findings of pollution indices of this study. The findings from this study may serve as baseline information for the pollution assessment of Czechia and European farmland and forest soil quality. The results might support policy-developers in sustainable farming and forestry for the health of the ecosystem towards food security, forest safety, as well as animal and human welfare.

4.1.6 Paper 6: Health risk assessment and the application of CF-PMF: a pollution assessment–based receptor model in an urban soil

This study revealed that the mean concentration of the PTEs such as Cr, As, Mn, Pb and Zn in the Frýdek-Místek district exceeded World average, whereas the mean concentrations of As, Pb and Zn were higher than upper continent crust. However, the concentration of Cu, Mn, Pb and Zn also exceeded the tolerable European average values limit.

The comparison of CF-PMF receptor model to EPA-PMF receptor model applied in the study showed that the PMF receptor model is a robust receptor model, but the hybridization of PMF and CF increases the source apportionment efficiency and reduces the error. A comparative assessment between the hybrid model and PMF revealed that consistently the hybrid model performed better than the single-parent model. For example, in this study, the estimated coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) showed that out of the seven PTEs evaluated, CF-PMF showed superior performance in all the seven PTEs relative to PMF model.

In terms of the potential health risk, the estimated total carcinogenic risk (TCR) for children was 2.33 times higher than that of the adults. The TCR for adults and children were 6.9×10^{-6} and 1.61×10^{-5}, respectively. Oral ingestion of soil was the main pathway for the CR of As for children, while dermal contact of soils was the main pathway for the CR of As for adults. Therefore, it was established that the calculated TCR for both adults and children were a good proof that according to US- EPA the carcinogenic health risk within the study area falls within the acceptable limits (i.e., TCR values should range in between 1×10^{-6} to 1×10^{-4}). Conclusively, the likelihood for indigenes within the study area to be exposed to carcinogenic related health risk associated with soil PTEs is relatively low or zero. However, it is pertinent for the people to take pragmatic

strategies to protect the soil by ensuring that PTE accumulations are continuously either abated and/or reduced. To achieve this, the study suggests regular monitoring of the soil PTEs contents using more effective tools such as the DSM approaches.

4.2 Concluding remarks

This thesis concluded that:

(i) Applying a robust and hybridized digital soil mapping (DSM) model in combination with terrain attributes, soil nutrients and appropriate auxiliary datasets was efficient in achieving effective prediction of the spatial distribution and estimation of the uncertainty levels of PTEs in forest and agricultural soils.

(ii) In the past decades, more studies focusing on the spatial prediction of PTEs in the soil applied geostatistical approaches, but recently, the combination of other DSM models such as machine learning algorithms is fast gaining recognition.

(iii) The integration of data fusion, terrain attribute, and the kriging modeling techniques produced optimal results with a high R^2 value, high prediction accuracy, marginal errors, and less bias.

(iv) For better prediction outcomes, proxies or additional data sets can be combined with soil characteristics that have a strong correlation with response variables.

(v) The combination of legacy datasets, coupled with appropriate modeling method(s) and a well-correlated environmental covariate dataset, generates valuable and reliably acceptable PTEs prediction results.

(vi) Adopting a higher spatial resolution remote sensing datasets together with input data in the prediction of PTEs or soil properties is not a guarantee for achieving good results. But the preferrable best results will be realized by combining environmental covariates

with a high correlation of the response variable, coupled with appropriate modeling approaches for predicting PTEs in either agricultural or forest soil.

(vii)   The application of mean, maximum, and minimum values as criteria for health risk estimation might not ensure a comprehensive scenario of the health status in any study area; other health risk assessment procedures and standards and factors (e.g., socioeconomics) need to be put into consideration.

(viii)   The adoption of a pollution assessment-based receptor model (ER-PMF) has proved more reliable and practical in estimating distribution sources.

(ix)   Introduction of geographical weighted regression ordinary kriging is highly dependable and efficient in the mapping and modeling of PTEs in the agricultural and forest soils compared to the use of only geographical weighted regression.

(x)   No specific geostatistics and/or machine learning algorithm (MLA) is an all-round the best in the prediction of PTEs. But the efficiency and reliability of any model depends mainly on the available datasets, and partly on the users' experience. For a good result, every study needs a different modeling method that is most appropriate for the type of dataset used because there is no single modeling technique that fits all datasets, and for the study objectives.

5.0 REFERENCES

Adhikari, K., Kheir, R.B., Greve, M.B., Bøcher, P.K., Malone, B.P., Minasny, B., McBratney, A.B., & Greve, M.H. (2013). High-resolution 3-D mapping of soil texture in Denmark. Soil Science Society of America Journal, 77 (3), 860–876.

Adomako, E.E., Raab, A., Norton, G.J., & Meharg, A.A. (2023). Potentially toxic element (PTE) soil concentrations at an urban unregulated Ghanaian e-waste recycling centre: Environmental contamination, human exposure and policy implications. Exposure and Health, 15(3), 677-686.

Agyeman, P. C., John, K., Kebonye, N.M., Borůvka, L., & Vašát, R. (2023). Combination of enrichment factor and positive matrix factorization in the estimation of potentially toxic element source distribution in agricultural soil. Environmental Geochemistry and Health, 45(5), 2359-2385.

Agyeman, P.C., Ahado, S.K., Kingsley, J., Kebonye, N.M., Biney, J.K.M., Borůvka, L., Vasat, R. & Kocarek, M. (2021). Source apportionment, contamination levels, and spatial prediction of potentially toxic elements in selected soils of the Czech Republic. Environmental geochemistry and health, 43, pp.601-620.

Agyeman, P. C., Khosravi, V., Kebonye, N.M., John, K., Borůvka, L., & Vašát, R. (2022). Using spectral indices and terrain attribute datasets and their combination in the prediction of cadmium content in agricultural soil. Computers and Electronics in Agriculture, 198, 107077.

Agyeman, P. C., John, K., Kebonye, N.M., Khosravi, V., Borůvka, L., & Vašát, R. (2023). Prediction of the concentration of antimony in agricultural soil using data fusion, terrain attributes combined with regression kriging. Environmental Pollution, 316, 120697.

Ahado, S. K., & Nwaogu, C. (2023). Spatial modelling and quantification of soil potentially toxic elements based on variability in sample size and land use along a toposequence at a district scale. International Journal of Environmental Science and Technology, 21, 3567–3586.

Ahado, S.K., Nwaogu, C., Sarkodie, V.Y.O., & Borůvka, L. (2021). Modeling and assessing the spatial and vertical distributions of potentially toxic elements in soil and how the concentrations differ. Toxics, 9(8), 181.

Alpaydin, E., (1997). Voting over multiple condensed nearest neighbors. Lazy Learning, 11, 115-132.

Alloway, B. J. (2013). Sources of heavy metals and metalloids in soils. In: Heavy metals in soils: trace metals and metalloids in soils and their bioavailability. Springer, 11-50.

Ambade, B., Sethi, S.S., & Chintala Cheruvu, M.R. (2023). Distribution, risk assessment, and source apportionment of polycyclic aromatic hydrocarbons (PAHs) using positive matrix factorization (PMF) in urban soils of East India. Environmental Geochemistry and Health, 45(2), 491-505.

An, Q., Zhou, T., Wen, C. & Yan, C. (2023). The effects of microplastics on heavy metals bioavailability in soils: a meta-analysis. Journal of Hazardous Materials, 460, 132369.

Anastasopolos, A.T., Hopke, P.K., Sofowote, U.M., Mooibroek, D., Zhang, J.J., Rouleau, M., Peng, H. & Sundar, N. (2023). Evaluating the effectiveness of low-sulphur marine fuel regulations at improving urban ambient PM2. 5 air quality: Source apportionment of PM2. 5 at Canadian Atlantic and Pacific coast cities with implementation of the North American emissions control area. Science of the Total Environment, 904, 166965.

Antoniadis, V., Shaheen, S. M., Boersch, J., Frohne, T., Du Laing, G., & Rinklebe, J. (2017). Bioavailability and risk assessment of potentially toxic elements in garden edible vegetables and soils around a highly contaminated former mining area in Germany. Journal of Environmental Management, 186, 192-200.

Araújo, D.F., Boaventura, G.R., Machado, W., Viers, J., Weiss, D., Patchineelam, S. R., Ruiz, I., Rodrigues, A. P. C., Babinski, M., & Dantas, E. (2017). Tracing of anthropogenic zinc sources in coastal environments using stable isotope composition. Chemical Geology, 449, 226–235.

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., & dL Mendonca-Santos, M. (2014). GlobalSoilMap: Toward a fine-resolution global grid of soil properties. Advances in Agronomy, 125, 93-134.

Assey, G.E., & Mogusu, E. (2023). Effects of persistent organic pollutants on environment, health and mountains: a review. Journal of Environmental Engineering and Science, 19(1),1-8.

Awad, A. (2023). The determinants of food insecurity among developing countries: Are there any differences? Scientific African, 19, e01512.

Bakker E.S., Van Donk E., & Immers A.K. (2016). Lake restoration by in-lake iron addition: a synopsis of iron impact on aquatic organisms and shallow lake ecosystems. Aquatic Ecology, 50, 121–135.

Ballabio, C., Panagos, P., Lugato, E., Huang, J.H., Orgiazzi, A., Jones, A., Fernández-Ugalde, O., Borrelli, P. & Montanarella, L. (2018). Copper distribution in European topsoils: An assessment based on LUCAS soil survey. Science of The Total Environment, 636, 282-298.

Bangroo, S.A., Najar, G.R., Achin, E., & Truong, P.N. (2020). Application of predictor variables in spatial quantification of soil organic carbon and total nitrogen using regression kriging in the North Kashmir Forest Himalayas. Catena, 193, 104632.

Barker, B.J., Clausen, J.L., Douglas, T.L., Bednar, A.J., Griggs, C.S., & Martin, W.A. (2021). Environmental impact of metals resulting from military training activities: A review. Chemosphere, 265, 129110.

Barthold F K, Wiesmeier M, Breuer L, Frede H G, Wu J, & Blank F.B. (2013). Land use and climate control the spatial distribution of soil types in the grasslands of Inner Mongolia. Journal of Arid Environments, 88, 194–205.

Behrens, T., Schmidt, K., MacMillan, R. A., & Viscarra Rossel, R. A. (2018). Multi-scale digital soil mapping with deep learning. Scientific Reports, 8(1), 1–9.

Belanović Simić, S., Miljković, P., Baumgertel, A., Lukić, S., Ljubičić, J., & Čakmak, D. (2023). Environmental and health risk assessment due to potentially toxic elements in soil near former antimony mine in Western Serbia. Land, 12(2), 421.

Bertling, S., Odnevall Wallinder, I., Leygraf, C., & Berggren Kleja, D. (2006). Occurrence and fate of corrosion-induced zinc in runoff water from external structures. Science of The Total Environment, 367, 908–923.

Binner, H., Sullivan, T., Jansen, M.A.K., & McNamara, M.E. (2023). Metals in urban soils of Europe: A systematic review. Science of The Total Environment, 854, 158734.

Björnerås, C., Weyhenmeyer, G.A., Evans, C.D., Gessner, M.O., Grossart, H.P., Kangur, K., Kokorite, I., Kortelainen, P., Laudon, H., Lehtoranta, J. and Lottig, N. (2017). Widespread increases in iron concentration in European and North American freshwaters. Global Biogeochemical Cycles, 31(10), 1488-1500.

Bolan, N.S., Makino, T., Kunhikrishnan, A., Kim, P.J., Ishikawa, S., Murakami, M., Naidu, R. & Kirkham, M.B., (2013). Cadmium contamination and its risk management in rice ecosystems. Advances in Agronomy, 119,183-273.

Borůvka, L., Vacek, O., & Jehlička, J. (2005). Principal component analysis as a tool to indicate the origin of potentially toxic elements in soils. Geoderma, 128(3-4 SPEC. ISS.), 289–300.

Borůvka, L., Vašát, R., Němeček, K., Novotný, R., Šrámek, V., Vacek, O., Pavlů, L., Fadrhonsová, V., & Drábek, 0. (2020). Application of regression-kriging and sequential Gaussian simulation for the delineation of forest areas potentially suitable for liming in the Jizera Mountains region, Czech Republic. Geoderma Regional, 21, e00286.

Bradl, H.B. (2005). Sources and origins of heavy metals. In Interface science and technology Chapter 6. (Vol.6, pp. 1-27). Elsevier.

Breiman, L. (1996). Stacked regressions. Machine Learning, 24(1), 49–64.

Breiman,L.(2001). Random forests. Machine Learning, 45(1), 5–32.

Broadley, M., Brown, P., Cakmak, I., Rengel, Z., & Zhao, F. (2012). Function of nutrients: micronutrients. In Marschner's mineral nutrition of higher plants. Academic Press, 191-248.

Brown, P. H., Cakmak, I., & Zhang, Q. (1999). Form and function of zinc plants. In Zinc in Soils and Plants: Proceedings of the international symposium on 'zinc in soils and plants' held at the University of Western Australia, 27–28 September 1993. Dordrecht: Springer Netherlands, 93-106.

Brown, S.G., Eberly, S., Paatero, P., & Norris, G.A. (2015). Methods for estimating uncertainty in PMF solutions: examples with ambient air and water quality data and guidance on reporting PMF results. Science of The Total Environment, 518, 626-635.

Cabral, J.B., Gentil, W.B., Ramalho, F.L., Braga, C.C., Becegato, V.A., & Paulino, A.T. (2023). Harmful effects of potentially toxic elements in soils of cerrado biomes. Water, Air, & Soil Pollution, 234(6), 334.

Cai, L., Xu, Z., Ren, M., Guo, Q., Hu, X., Hu, G., Wan, H., & Peng, P. (2012). Source identification of eight hazardous heavy metals in agricultural soils of Huizhou, Guangdong Province, China. Ecotoxicology and Environmental Safety, 78, 2–8.

Cardoso, K.M., Boechat, C.L., do Nascimento, C.W.A., Escobar, M.E.O., da Silva, D. G., da Silva Lins, S.A., & Morais, P. G. C. (2023). Watershed-scale assessment of environmental background values of soil potential toxic elements from the Caatinga and Atlantic forest ecotone in Brazil. Chemosphere, 338, 139394.

Černik, J., Kunc, J., & Martinat, S. (2016). Territorial-technical and socio-economic aspects of successful brownfield regeneration: A case study of the Liberec region (Czech Republic). Geographia Technica, 11, 22-38.

Chen J, Blume H.P, Beyer L. (2000). Weathering of rocks induced by lichen colonization—a review. CATENA, 39:121–146

Chen, H., Teng, Y., Lu, S., Wang, Y., & Wang, J. (2015). Contamination features and health risk of soil heavy metals in China. Science of the Total Environment, 512–513, 143–153.

Chen, L., Ren, C., Li, L., Wang, Y., Zhang, B., Wang, Z., & Li, L. (2019). A comparative assessment of geostatistical, machine learning, and hybrid approaches for mapping topsoil organic carbon content. ISPRS International Journal of Geo-Information, 8(4), 174.

Chen, L., Wang, F., Zhang, Z., Chao, H., He, H., Hu, W., Zeng, Y., Duan, C., Liu, J. & Fang, L. (2023). Influences of arbuscular mycorrhizal fungi on crop growth and potentially toxic element accumulation in contaminated soils: A meta-analysis. Critical Reviews in Environmental Science and Technology, 53(20), 1795-1816.

Chen, L., Lowenthal, D.H., Watson, J.G. Koracin, D., Kumar, N., Knipping, E.M., Wheeler, N., Craig, K., & Reid, S. (2010). Toward effective source apportionment using positive matrix factorization: Experiments with simulated PM2.5 data. Journal of the Air & Waste Management Association, 60, 43–54.

Chen, T., Liu, X., Li, X., Zhao, K., Zhang, J., Xu, J., Shi, J., & Dahlgren, R. A. (2009). Heavy metal sources identification and sampling uncertainty analysis in a field-scale vegetable soil of Hangzhou, China. Environmental Pollution, 157(3), 1003–1010.

Cheng-Ping, Z., Chuan, L., & Hai-wei, G. (2011). Research on hydrology time series prediction based on grey theory and [epsilon]-support vector regression. In: 2011International Conference on Computer Distributed Control and Intelligent Environmental Monitoring, 1673–1676.

Cherkassky, V. & Mulier, F.M. (2007). Learning from data: concepts, theory, and methods. John Wiley & Sons.

Congdon, P. (2007). Bayesian statistical modelling. John Wiley & Sons.

Cortes, C., Vapnik, V. (1995). Support-vector networks. Machine Learning, 20 (3), 273–297.

Costa, E.M., Samuel-Rosa, A., & dos Anjos, L.H.C. (2018). Digital elevation model quality on digital soil mapping prediction accuracy. Ciencia e Agrotecnologia, 42(6), 608–622.

Coutinho, G.B.F., Moreira, M.D.F.R., Fischer, F.M., Dos Santos, M.C.R., Feitosa, L.F., de Azevedo, S.V., Borges, R.M., Nascimento-Sales, M., Christoffolete, M.A., Santa-Marinha, M.S. & Valente, D. (2023). Influence of environmental exposure to steel waste on endocrine dysregulation and PER3 gene polymorphisms. International Journal of Environmental Research and Public Health, 20(6), 4760.

Dazzi, C., & Papa, G.L. (2022). A new definition of soil to promote soil awareness, sustainability, security and governance. International Soil and Water Conservation Research, 10(1), 99-108.

de Almeida Ribeiro Carvalho, M., Botero, W.G. & de Oliveira, L.C. (2022). Natural and anthropogenic sources of potentially toxic elements to aquatic environment: a systematic literature review. Environmental Science Pollution Research, 29, 51318–51338.

De Miguel, E., Iribarren, I., Chacon, E., Ordonez, A., & Charlesworth, S. (2007). Risk-based evaluation of the exposure of children to trace elements in playgrounds in Madrid (Spain). Chemosphere, 66(3), 505-513.

De'ath, G. (2007). Boosted trees for ecological modeling and prediction. Ecology 88 (1), 243–251.

Delerce, S., Dorado, H., Grillon, A., Rebolledo, M.C., Prager, S.D., Patiño, V.H., Garcés Varón, G. & Jiménez, D., (2016). Assessing weather-yield relationships in rice at local scale using data mining approaches. PloS One, 11(8),

Demir, M., Tunç, E., Thiele-Bruhn, S., Çelik, Ö., Tsegai, A. T., Aslan, N., & Arslan, S. (2023). Status, sources and assessment of potentially toxic element (PTE) contamination in roadside orchard soils of Gaziantep (Türkiye). International Journal of Environmental Research and Public Health, 20(3), 2467.

Dharumarajan, S., Hegde, R., Janani, N &. Singh, S.K. (2019). The need for digital soil mapping in India. Geoderma Regional, 16, p.e00204.

Dhuldhaj, U.P., Singh, R., & Singh, V.K. (2023). Pesticide contamination in agro-ecosystems: toxicity, impacts, and bio-based management strategies. Environmental Science and Pollution Research, 30(4), 9243-9270.

Ding, Q., Cheng, G., Wang, Y., & Zhuang, D. (2017). Effects of natural factors on the spatial distribution of heavy metals in soils surrounding mining regions. Science of The Total Environment, 578, 577–585.

Doetterl, S., Stevens, A., van Oost, K., Quine, T. A., & van Wesemael, B. (2013). Spatially explicit regional-scale prediction of soil organic carbon stocks in cropland using environmental variables and mixed model approaches. Geoderma, 204–205, 31–42.

Douay, F., Pruvot, C., Roussel, H., Ciesielski, H., Fourrier, H., Proix, N., & Waterlot, C. (2008). Contamination of urban soils in an area of Northern France polluted by dust emissions of two smelters. Water, Air, and Soil Pollution, 188, 247-260.

Duarte, A.C., Cachada, A., Rocha-Santos, T. (2018). Soil and pollution: From Monitoring to Remediation. Chapter 1.

Ehsani, M.R., Upadhyaya, S.K., Slaughter, D., Shafii, S. & Pelletier, M. (1999). A NIR technique for rapid determination of soil mineral nitrogen. Precision agriculture, 1(2), 219-236.

El Rasafi, T., Oukarroum, A., Haddioui, A., Song, H., Kwon, E.E., Bolan, N., Tack, F.M., Sebastian, A., Prasad, M.N.V. & Rinklebe, J. (2022). Cadmium stress in plants: A critical review

of the effects, mechanisms, and tolerance strategies. Critical Reviews in Environmental Science and Technology, 52(5), 675-726.

Elith, J., Leathwick, J.R., & Hastie, T. (2008). A working guide to boosted regression trees. Journal of Animal Ecology 77 (4), 802–813.

Elkhlifi, Z., Iftikhar, J., Sarraf, M., Ali, B., Saleem, M.H., Ibranshahib, I., Bispo, M.D., Meili, L., Ercisli, S., Torun Kayabasi, E. & Alemzadeh Ansari, N. (2023). Potential role of biochar on capturing soil nutrients, carbon sequestration and managing environmental challenges: a review. Sustainability, 15(3), 2527.

Engwa, G.A., Ferdinand, P.U., Nwalo, F.N., & Unachukwu, M.N. (2019). Mechanism and Health Effects of Heavy Metal Toxicity in Humans. In Poisoning in the Modern World—New Tricks for an Old Dog? Karcioglu, O., Arslan, B., Eds., IntechOpen: London, UK,

Eriksson, J., Dahlin, S.A., Sohlenius, G., Söderström, M., & Öborn, I., (2017). Spatial patterns of essential trace element concentrations in Swedish soils and crops. Geoderma Regional, 10, 163–174.

ESRI, (2019). ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Eziz, M., Mohammad, A., Mamut, A., & Hini, G. (2018). A human health risk assessment of heavy metals in agricultural soils of Yanqi basin, silk road economic belt, China. Human and Ecological Risk Assessment, 24(5), 1352–1366.

Falconnier, G.N., Leroux, L., Beillouin, D., Corbeels, M., Hijmans, R.J., Bonilla-Cedrez, C., van Wijk, M., Descheemaeker, K., Zingore, S., Affholder, F. & Lopez-Ridaura, S. (2023). Increased mineral fertilizer use on maize can improve both household food security and regional food production in East Africa. Agricultural Systems, 205, 103588.

Famuyiwa, A. O., Davidson, C. M., Ande, S., & Oyeyiola, A. O. (2022). Potentially toxic elements in urban soils from public-access areas in the rapidly growing megacity of Lagos, Nigeria. Toxics, 10(4), 154.

FAO & ITPS. (2015). Intergovernmental technical panel on soils. Status of the World's soil resources. Intergovernmental Technical Panel on Soils, 100–146.

FAO. (2018). Concept Note. Global Symposium on Soil Pollution (GSOP18). 2-4 May 2018. FAO257 Rome, Italy. http://www.fao.org/3/I8411EN/i8411en.pdf [accessed 25.9.18].

Fei, X., Christakos, G., Xiao, R., Ren, Z., Liu, Y., & Lv, X. (2019). Improved heavy metal mapping and pollution source apportionment in Shanghai City soils using auxiliary information. Science of the Total Environment, 661, 168-177.

Fotheringham, A.S, Brunsdon, C., & Charlton, M. (2002). Geographically weighted regression: The analysis of spatially varying relationships. Wiley, Chichester.

Friedman, J. H. (1991). Multivariate adaptive regression splines. The annals of statistics, 19(1), 1-67.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

Friedman, J.H. (2002). Stochastic gradient boosting. Computational statistics & data analysis, 38(4), 367-378.

Gamon, J.A., Penuelas, J. & Field, C.B. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sensing of environment, 41(1), 35-44.

Gao, P., Xie, Y., Song, C., Cheng, C., & Ye, S. (2023). Exploring detailed urban-rural development under intersecting population growth and food production scenarios: Trajectories for China's most populous agricultural province to 2030. Journal of Geographical Sciences, 33(2), 222-244.

Gautam, K., Sharma, P., Dwivedi, S., Singh, A., Gaur, V.K., Varjani, S., Srivastava, J.K., Pandey, A., Chang, J.S. & Ngo, H.H. (2023). A review on control and abatement of soil pollution by heavy metals: Emphasis on artificial intelligence in recovery of contaminated soil. Environmental Research, 115592.

Gautam, R., Panigrahi, S., Franzen, D., & Sims, A. (2011). Residual soil nitrate prediction from imagery and non-imagery information using neural network technique. Biosystems Engineering, 110(1), 20–28.

Gerdol, R., Bragazza, L. (2020). Effects of altitude on element accumulation in alpine moss. Chem, 64, 810-816.

Ghosh, S., Mohanty, S., Nayak, S., Sukla, L.B., & Das, A.P. (2016). Molecular identification of indigenous manganese solubilising bacterial biodiversity from manganese mining deposits. Journal of Basic Microbiology, 56 (2016), 254-262.

Giesler, R., Ilvesniemi, H., Nyberg, L., van Hees, P., Starr, M., Bishop K, Kareinen, T., & Lundstrom U.S.(2000). Distribution and mobilization of Al, Fe and Si in three podzolic soil profiles in relation to the humus layer. Geoderma, 94, 249–263.

Giller, K.E., Witter, E., & Mcgrath, S.P. (1998). Toxicity of heavy metals to microorganisms and microbial processes in agricultural soils: a review. Soil Biology and Biochemistry, 30, 1389–1414.

Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Filho, E. I. F. (2019). Modelling and mapping soil organic carbon stocks in Brazil. Geoderma, 340, 337–350.

González-Macías, C., Schifter, I., Lluch-Cota, D. B., Méndez-Rodríguez, L., & Hernández-Vázquez, S. (2006). Distribution, enrichment and accumulation of heavy metals in coastal sediments of Salina Cruz Bay, México. Environmental Monitoring and Assessment, 118(1–3), 211–230.

Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. Geoderma, 103(1-2), 3-26.

Graedel, T.E., Van Beers, D., Bertram, M., Fuse, K., Gordon, R.B., Gritsinin, A., Harper, E.M., Kapur, A., Klee, R.J., Lifset, R., Memon, L., & Spatari, S. (2005). The multilevel cycle of anthropogenic zinc. Journal of Industrial Ecology, 9, 67–90.

Grybos, M., Davranche, M., Gruau, G., Petitjean, P., Pedrot, M. (2009). Increasing pH drives organic matter solubilization from wetland soils under reducing conditions. Geoderma, 154, 13–19.

Guagliardi, I., Cicchella, D., & De Rosa, R. (2012). A geostatistical approach to assess concentration and spatial distribution of heavy metals in urban soils. Water, Air, and Soil pollution, 223(9), 5983–5998.

Guo, P.T., Li, M.F., Luo, W., Tang, Q.F., Liu, Z.W., & Lin, Z.M., (2015). Digital mapping of soil organic matter for rubber plantation at regional scale: an application of random forest plus residuals kriging approach. Geoderma, 237–238, 49–59.

Haj Heidary, R., Golzan, S. A., Mirza Alizadeh, A., Hamedi, H., & Ataee, M. (2023). Probabilistic health risk assessment of potentially toxic elements in traditional and industrial olive products. Environmental Science and Pollution Research, *30*(4), 10213-10225.

Han, F.X., Banin, A., Su, Y., David, Monts, L., Plodinec, M. John, Kingery, W.L., & Glover, Triplett, E. (2002). Industrial age anthropogenic inputs of heavy metals into the pedosphere. Naturwissenschaften, 89, 497–504.

Hengl, T., Heuvelink, G. B., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma, 120(1-2), 75-93.

Hettler, J., Irion, G., & Lehmann, B. (1997). Environmental impact of mining waste disposal on a tropical lowland river system: a case study on the Ok Tedi Mine, Papua New Guinea. Mineralium Deposita, 32, 280-291.

Hinton G E, Osindero S, & Teh Y W. (2006). A fast-learning algorithm for deep belief nets. Neural Computation, 18, 1527–1554.

Hjortenkrans, D.S.T., Bergbäck, B.G., &Häggerud, A.V. (2007). Metal emissions from brake linings and tires: case studies of Stockholm, Sweden 1995/1998 and 2005. Environmental Science & Technology, 41,5224–5230.

Hoshyari, E., Hassanzadeh, N., Keshavarzi, B., Jaafarzadeh, N., & Rezaei, M. (2023). Spatial distribution, source apportionment, and ecological risk assessment of elements (PTEs, REEs, and ENs) in the surface soil of shiraz city (Iran) under different land-use types. Chemosphere, 311, 137045.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical statistics, 15(3), 651-674.

Hou, D., O'Connor, D., Nathanail, P., Tian, L., & Ma, Y. (2017). Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review. Environmental Pollution, 231, 1188–1200.

Huang, J., Guo, S., Zeng, G.M., Li, F., Gu, Y., Shi, Y., Shi, L., Liu, W. & Peng, S. (2018). A new exploration of health risk assessment quantification from sources of soil heavy metals under different land use. Environmental Pollution, 243, 49-58.

Iqbal, J., Thomasson, J. A., Jenkins, J. N., Owens, P. R., & Whisler, F. D. (2005). Spatial variability analysis of soil physical properties of alluvial soils. Soil Science Society of America Journal, 69(4), 1338–1350.

Jansen, B., Nierop, K. G., & Verstraten, J. M. (2005). Mechanisms controlling the mobility of dissolved organic matter, aluminium and iron in podzol B horizons. European Journal of Soil Science, 56(4), 537-550.

IUSS Working Group WRB. (2015). World Reference Base for soil resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.

Jansen, B., Nierop, K. G., & Verstraten, J. M. (2005). Mechanisms controlling the mobility of dissolved organic matter, aluminium and iron in podzol B horizons. European Journal of Soil Science, 56(4), 537-550.

Jannetto, P. J., & Cowl, C. T. (2023). Elementary overview of heavy metals. Clinical Chemistry, 69(4), 336-349.

Johari, A., Khani, M., Hadianfard, M.A. & JavidSharifi, B. (2020). System reliability analysis for seismic site classification based on sequential Gaussian co-simulation: A case study in Shiraz, Iran. Soil Dynamics and Earthquake Engineering, 137, 106286.

Jones, B. G., Alyazichi, Y. M., Low, C., Goodfellow, A., Chenhall, B. E., & Morrison, R. J. (2019). Distribution and sources of trace element pollutants in the sediments of the industrialized Port Kembla Harbour, New South Wales, Australia. Environmental Earth Sciences, 78(12).

Kabala, C., Galka, B., & Jezierski, P. (2020). Assessment and monitoring of soil and plant contamination with trace elements around Europe's largest copper ore tailings impoundment. Science of The Total Environment, 738, 139918.

Kabata-Pendias, A., & Szteke, B. (2015). Trace Elements in Abiotic and Biotic Environments. In Trace Elements in Abiotic and Biotic Environments.

Kabata-Pendias, A. (2011). Trace Elements in Soils and Plants. 4th Edition. CRC Press Taylor & Francis Group. NY, USA.

Kalkhajeh, Y.K., Huang, B., Hu, W., Ma, C., Gao, H., Thompson, M.L., & Hansen, H. C. B. (2021). Environmental soil quality and vegetable safety under current greenhouse vegetable production management in China. Agriculture, Ecosystems & Environment, 307, 107230.

Kazapoe, R. W., Amuah, E. E. Y., Dankwa, P., Ibrahim, K., Mville, B. N., Abubakari, S., & Bawa, N. (2021). Compositional and source patterns of potentially toxic elements (PTEs) in soils in southwestern Ghana using robust compositional contamination index (RCCI) and k-means cluster analysis. Environmental Challenges, 5, 100248.

Ke, W., Li, C., Zhu, F., Luo, X., Feng, J., Li, X., Jiang, Y., Wu, C., Hartley, W. and Xue, S. (2023). Effect of potentially toxic elements on soil multifunctionality at a lead smelting site. Journal of Hazardous Materials, 454, 131525.

Kempen, B., Brus, D. J., Stoorvogel, J. J., Heuvelink, G. B. M., & de Vries, F. (2012). Efficiency comparison of conventional and digital soil mapping for updating soil maps. Soil Science Society of America Journal, 76(6), 2097–2115.

Keskin, H. & Grunwald, S. (2018). Regression kriging as a workhorse in the digital soil mapper's toolbox. Geoderma, 326, 22-41.

Kharuk, V.I., Ponomarev, E.I., Ivanova, G.A., Dvinskaya, M.L., Coogan, S.C., Flannigan, M.D.(2021). Wildfires in the Siberian Taiga. Ambio, 2021, 50, 1953–1974.

Kim, D.M., Kwon, H.L., & Im, D.G. (2023). Determination of contamination sources and geochemical behaviors of metals in soil of a mine area using Cu, Pb, Zn, and S isotopes and positive matrix factorization. Journal of Hazardous Materials, 447, 130827.

Klapstein, S. J., Walker, A. K., Saunders, C. H., Cameron, R. P., Murimboh, J. D., & O'Driscoll, N. J. (2020). Spatial distribution of mercury and other potentially toxic elements using epiphytic lichens in Nova Scotia. Chemosphere, 241, 125064.

Kozák, J., Němeček, J., Borůvka, L., Lérová, Z., Němeček, K., Kodešová, R., Janků, J., Jacko, K., Hladík, J. & Zádorová, T., (2010). Atlas půd České republiky.[Soil Atlas of the Czech Republic.]. Prague, Czech University of Life Sciences Prague, 150.

Keskin, H., Grunwald, S. 2018. Regression kriging as a workhorse in the digital soil mapper's toolbox. Geoderma, 326, 22–41.

Kozák, J. (2010). Soil Atlas of the Czech Republic. 150

Kuesel, K., Roth, U., & Drake, H.L. (2001). Microbial reduction of Fe(III) in the presence of oxygen under low pH conditions. Abstracts of the General Meeting of the American Society for Microbiology. Environmental Microbiology, 101:519.

Kuhn, M., Weston, S., Keefer, C., Coulter, N., & Quinlan, R. (2018). Cubist: Rule-and Instance-based Regression Modeling. (R package version 0.2.2).

Kumar, S., Lal, R., & Liu, D. (2012). A geographically weighted regression kriging approach for mapping soil organic carbon stock. Geoderma, 189, 627-634.

Lagacherie, P., & McBratney, A. B. (2006). Chapter 1 spatial soil Information systems and spatial soil inference systems: Perspectives for digital soil mapping. Developments in Soil Science, 31(C), 3–22.

Lago, B. C., Silva, C. A., Melo, L. C. A., & Morais, E. G. de. (2021). Predicting biochar cation exchange capacity using Fourier transform infrared spectroscopy combined with partial least square regression. Science of the Total Environment, 794, 148762.

Lake, D.L., Kirk, P.W.W., & Lester, J.N. (1984). Fractionation, characterization, and speciation of heavy metals in sewage sludge and sludge-amended soils: A Review. Journal of Environmental Quality, 13(2), 175–183.

Lalonde, K., Mucci, A., Ouellet, A., & Gelinas, Y. (2012). Preservation of organic matter in sediments promoted by iron. Nature, 483:198–200.

Lamichhane, S., Kumar, L., & Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. Geoderma 395–413.

Landeweert, R., Hoffland E., Finlay R.D., Kuyper T.W., & van Breemen, N. (2001). Linking plants to rocks: ectomycorrhizal fungi mobilize nutrients from minerals. Trends in Ecology & Evolution 16:248–254.

Lenntech. (2008). Manganese (Mn) - Chemical properties, health and environmental effects. http://www.lenntech.com/periodic/elements/mg.htm.

Liang-yan, S., & Li, C. (2009). A fast and scalable fuzzy-rough nearest neighbor algorithm. In 2009 WRI Global Congress on Intelligent Systems, 4, 311-314.

Li, J.S., Beiyuan, J., Tsang, D.C.W., Wang, L., Poon, C.S., Li, X. D., & Fendorf, S. (2017). Arsenic-containing soil from geogenic source in Hong Kong: Leaching characteristics and stabilization/solidification. Chemosphere, 182, 31–39.

Li, K., Liang, T., Wang, L. & Yang, Z., (2016). Contamination and health risk assessment of heavy metals in road dust in Bayan Obo Mining Region in Inner Mongolia, North China. Journal of Geographical Sciences, 25(12), 1439-1451.

Lian, G., Guo, X., Fu, B., & Hu, C. (2009). Prediction of the spatial distribution of soil properties based on environmental correlation and geostatistics. Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering, 25(7), 237–242.

Liang, Q., Tian, K., Li, L., He, Y., Zhao, T., Liu, B., & Teng, Y. (2022). Ecological and human health risk assessment of heavy metals based on their source apportionment in cropland soils around an e-waste dismantling site, Southeast China. Ecotoxicology and Environmental Safety, 242, 113929.

Linnik, V. G., Bauer, T. V., Minkina, T. M., Mandzhieva, S. S., & Mazarji, M. (2020). Spatial distribution of heavy metals in soils of the flood plain of the Seversky Donets River (Russia) based on geostatistical methods. Environmental Geochemistry and Health, 1–15.

Liu, W. X., Li, X. D., Shen, Z. G., Wang, D. C., Wai, O. W. H., & Li, Y. S. (2003). Multivariate statistical study of heavy metal enrichment in sediments of the Pearl River Estuary. Environmental Pollution, 121(3), 377–388.

Liu, X., Shi, H., Bai, Z., Zhou, W., Liu, K., Wang, M., & He, Y. (2020). Heavy metal concentrations of soils near the large opencast coal mine pits in China. Chem, 244, 125360.

Liu, Y., Chen, Z., Xiao, T., Zhu, Z., Jia, S., Sun, J., Ning, Z., Gao, T. and Liu, C., (2022). Enrichment and environmental availability of cadmium in agricultural soils developed on Cd-rich black shale in southwestern China. Environmental Science and Pollution Research, 29(24), 36243-36254.

López-Granados, F., Jurado-Expósito, M., Peña-Barragán, J. M., & García-Torres, L. (2005). Using geostatistical and remote sensing approaches for mapping soil properties. European Journal of Agronomy, 23(3), 279–289.

Lu, A.X., Wang, J.H., Qin, X.Y., Wang, K.Y., Han, P., & Zhang, S.Z., (2012). Multivariate and geostatistical analyses of the spatial distribution and origin of heavy metals in the agricultural soils in Shunyi, Beijing, China. Science of The Total Environment, 425, 66–74.

Luo, L., Ma, Y., Zhang, S., Wei, D., & Zhu, Y. G. (2009). An inventory of trace element inputs to agricultural soils in China. Journal of Environmental Management, 90(8), 2524–2530.

Luo, X., Wu, C., Lin, Y., Li, W., Deng, M., Tan, J., & Xue, S. (2023). Soil heavy metal pollution from Pb/Zn smelting regions in China and the remediation potential of biomineralization. Journal of Environmental Sciences, 125, 662-677.

Łyszczarz, S., Błońska, E., & Lasota, J. (2020). The application of the geo-accumulation index and geostatistical methods to the assessment of forest soil contamination with heavy metals in the Babia Góra National Park (Poland). Archives of Environmental Protection, 46(3), 69–79.

Ma, J., Chen, L., Chen, H., Wu, D., Ye, Z., Zhang, H., & Liu, D. (2023). Spatial distribution, sources, and risk assessment of potentially toxic elements in cultivated soils using isotopic tracing techniques and Monte Carlo simulation. Ecotoxicology and Environmental Safety, 259, 115044.

Ma, Y., Minasny, B., & Wu, C. (2017). Mapping key soil properties to support agricultural production in Eastern China. Geoderma Regional, 10, 144–153.

Maas, S., Scheifler, R., Benslama, M., Crini, N., Lucot, E., Brahmia, Z., Benyacoub, S. & Giraudoux, P. (2010). Spatial distribution of heavy metal concentrations in urban, su burban, and agricultural soils in a Mediterranean city of Algeria. Environmental pollution, 158(6), 2294- 2301.

Mahmoudabadi, E., Sarmadian, F., & Nazary Moghaddam, R. (2015). Spatial distribution of soil heavy metals in different land uses of an industrial area of Tehran (Iran). International Journal of Environmental Science and Technology, 12(10), 3283–3298.

Majhi P., S.M. Samantaray (2020). Effect of hexavalent chromium on paddy crops (*Oryza sativa*) Journal of Pharmacognosy and Phytochemistry, 9 (2), 1301-1305

Mantovi, P., Bonazzi, G., Maestri, E., & Marmiroli, N. (2003). Accumulation of copper and zinc from liquid manure in agricultural soils and crop plants. Plant and Soil, 250, 249-257.

Martin, A. P., Lim, C., Kah, M., Rattenbury, M. S., Rogers, K. M., Sharp, E. L., & Turnbull, R. E. (2023). Soil pollution driven by duration of urbanisation and dwelling quality in urban areas: An example from Auckland, New Zealand. Applied Geochemistry, 148, 105518.

Martínez, C.E., & Motto, H.L. (2000). Solubility of lead, zinc and copper added to mineral soils. Environmental Pollution, 107(1), 153–158.

Marx, A., Hintze, S., Sanda, M., Jankovec, J., Oulehle, F., Dusek, J., Vitvar, T., Vogel, T., van Geldern, R., Barth, J.A.C. (2017). Acid rain footprint three decades after peak deposition: long-term recovery from pollutant sulphate in the Uhlirska catchment (Czech Republic). Science of The Total Environment, 598, 1037–1049.

Maynard, J.J., & Levi, M.R. (2017). Hyper-temporal remote sensing for digital soil mapping: characterizing soil-vegetation response to climatic variability. Geoderma, 285, 94–109.

McBratney, A.B., Mendonça Santos, M.L., & Minasny, B. (2003). On digital soil mapping. Geoderma, 117(1–2), 3–52.

Meinshausen, N. (2006). Quantile regression forests. Journal of Machine Learning Research, 7, 983–999. https://www.jmlr.org/papers/volume7/meinshausen06a/meinshausen06a.pdf.

Minasny, B., & McBratney, A.B. (2016). Digital soil mapping: A brief history and some lessons. Geoderma, 264, 301–311.

Mishra, P. & Nikzad-Langerodi, R. (2020). Partial least square regression versus domain invariant partial least square regression with application to near-infrared spectroscopy of fresh fruit. Infrared Physics & Technology, 111, 103547.

Mohanty, S., Benya, A., Hota, S., Kumar, M. S., & Singh, S. (2023). Eco-toxicity of hexavalent chromium and its adverse impact on environment and human health in Sukinda Valley of India: A review on pollution and prevention strategies. Environmental Chemistry and Ecotoxicology, 5, 46-54.

Mohanty, S., Bal, B., & Das, A.P. (2014). Adsorption of hexavalent chromium onto activated carbon. Austin Journal of Biotechnology & Bioengineering, 1(2), 5.

Molla, A., Zhang, W., Zuo, S., Ren, Y., & Han, J. (2023). A machine learning and geostatistical hybrid method to improve spatial prediction accuracy of soil potentially toxic elements. Stochastic Environmental Research and Risk Assessment, 37(2), 681-696.

Mondal, S. & Singh, G. (2021). Pollution evaluation, human health effect and tracing source of trace elements on road dust of Dhanbad, a highly polluted industrial coal belt of India. Environmental Geochemistry and Health, 43, 2081-2103.

Moreno-Seco, F., Micó, L. and Oncina, J. (2003). A modification of the LAESA algorithm for approximated k-NN classification. Pattern Recognition Letters, 24(1-3), 47-53.

Mwelwa, S., Chungu, D., Tailoka, F., Beesigamukama, D., & Tanga, C. (2023). Biotransfer of heavy metals along the soil-plant-edible insect-human food chain in Africa. Science of The Total Environment, 881, 163150.

Naidu, R., Biswas, B., Willett, I.R., Cribb, J., Singh, B.K., Nathanail, C.P., Coulon, F., Semple, K.T., Jones, K.C., Barclay, A., & Aitken, R.J. (2021). Chemical pollution: A growing peril and potential catastrophic risk to humanity. Environment International, 156, 106616.

Nicodemus, K. K., Malley, J. D., Strobl, C., & Ziegler, A. (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. BMC Bioinformatics, 11(1), 1–13.

Nozari, S., Pahlavan-Rad, M. R., Brungard, C., Heung, B., & Borůvka, L. (2024). Digital soil mapping using machine learning-based methods to predict soil organic carbon in two different districts in the Czech Republic. Soil and Water Research, 19(1), 32-49.

Ntzoufras, I. (2011). Bayesian modeling using WinBUGS. https://doi.org/10.1080/09332480.2012.685377

Nwaogu, C., & Cherubin, M.R. (2023). Integrated Agricultural Systems: the 21st Century nature-based solution for resolving the global FEEES challenges. Advances in Agronomy. In Press.

Nwaogu, C., Ogbuagu, H.O., Abrakasa, S., Olawoyin, M.A., & Pavlů, V. (2017). Assessment of the impacts of municipal solid waste dumps on soils and plants, Chemistry and Ecology, 33:7, 589-606.

Nziguheba, G., & Smolders, E. (2008). Inputs of trace elements in agricultural soils via phosphate fertilizers in European countries. Science of The total environment, 390(1), 53-57.

Ogunlalu, O., Oyekunle, I.P., Iwuozor, K.O., Aderibigbe, A.D., Emenike, E.C. (2021). Trends in the mitigation of heavy metal ions from aqueous solutions using unmodified and chemically-modified agricultural waste adsorbents. Current Research in Green and Sustainable Chemistry, 2021, 4, 100188.

Olarewaju, E., & Obeng-Gyasi, E. (2023). Cadmium, lead, chronic physiological stress and endometrial cancer: How environmental policy can alter the exposure of at-risk women in the United States. Healthcare, 11(9), 1278.

Ottoy, S., De Vos, B., Sindayihebura, A., Hermy, M., & Van Orshoven, J. (2017). Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. Ecological Indicators, 77, 139–150.

Paatero, P., Eberly, S., Brown, S.G., & Norris, G.A. (2014). Methods for estimating uncertainty in factor analytic solutions. Atmospheric Measurement Techniques, 7(3), 781-797.

Pacifico, L. R., Pizzolante, A., Guarino, A., Iannone, A., Esposito, M., & Albanese, S. (2023). Wildfires as a source of potentially toxic elements (PTEs) in soil: A case study from Campania Region (Italy). International Journal of Environmental Research and Public Health, 20(5), 4513.

Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning for digital soil mapping. Soil, 5(1), 79–89.

Pavlů, L., Drábek, O., Stejskalová, Š., Tejnecký, V., Hradilová, M., Nikodem, A., & Borůvka, L., (2018). Distribution of aluminium fractions in acid forest soils: influence of vegetation changes. iForest - Biogeosciences and Forestry, For. 11, 721–727.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: machine learning in Python. Journal of Machine Learning Research, 12 (10), 2825–2830.

Peli, M., Bostick, B.C., Barontini, S., Lucchini, R.G., & Ranzi, R. (2021). Profiles and species of Mn, Fe and trace metals in soils near a ferromanganese plant in Bagnolo Mella (Brescia, IT). Science of The Total Environment, 755, 143123.

Peng, G., Bing, W., Guangpo, G., & Guangcan, Z. (2013). Spatial distribution of soil organic carbon and total nitrogen based on GIS and geostatistics in a small watershed in a hilly area of northern China. PLoS ONE, 8(12), e83592.

Penizek, V., & Boruvka, L. (2008). The digital terrain model as a tool for improved delineation of alluvial soils. Digital Soil Mapping with Limited Data, 319–326.

Petryshen, W. (2023). Spatial distribution of selenium and other potentially toxic elements surrounding mountaintop coal mines in the Elk Valley, British Columbia, Canada. Heliyon, 9(7).

Phippen, B. Horvath, C. Nordin, R., & Nagpal, N. (2008). Ambient water quality guidelines for iron: overview. Ministry of Environment Province of British Columbia.

PlantProbs.net.(2019). Manganese in plants and soil. https://plantprobs.net/plant/nutrientImbalances/sodium.html

Qing, X., Yutong, Z. & Shenggao, L. (2015). Assessment of heavy metal pollution and human health risk in urban soils of steel industrial city (Anshan), Liaoning, Northeast China. Ecotoxicology and Environmental Safety, 120, 377-385.

Qu, M., Chen, J., Huang, B., & Zhao, Y. (2020). Enhancing apportionment of the point and diffuse sources of soil heavy metals using robust geostatistics and robust spatial receptor model with categorical soil-type data. Environmental Pollution, 265, 114964.

Quinlan, J.R. (1992). Learning with continuous classes. In 5th Australian joint conference on artificial intelligence, 92, 343-348.

Rao, R. A. K., & Kashifuddin, M. (2016). Adsorption studies of Cd (II) on Ball Clay: comparison with other natural clays. Arabian Journal of Chemistry, 9, S1233-S1241.

Ray, R.R. (2016). Adverse hematological effects of hexavalent chromium: an overview. Interdisciplinary Toxicology, 9 (2), 55-65.

Reyes, A., Cuevas, J., Fuentes, B., Fernández, E., Arce, W., Guerrero, M., & Letelier, M. V. (2021). Distribution of potentially toxic elements in soils surrounding abandoned mining waste located in Taltal, Northern Chile. Journal of Geochemical Exploration, 220, 106653.

Rouhani, A., Gusiatin, M. Z., & Hejcman, M. (2023). An overview of the impacts of coal mining and processing on soil: Assessment, monitoring, and challenges in the Czech Republic. Environmental Geochemistry and Health, 45, 7459–7490.

Minasny, B., Setiawan, B.I., Saptomo, S.K. & McBratney, A.B. (2018). Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. Geoderma, 313, 25-40.

Salvador-Blanes, S., Cornu, S., Bourennane, H., & King, D. (2006). Controls of the spatial variability of Cr concentration in topsoils of a central French landscape. Geoderma, 132, 143–157.

Sánchez, M., Sabio, L., Gálvez, N., Capdevila, M., & Dominguez-Vera, J. M. (2017). Iron chemistry at the service of life. IUBMB life, 69(6), 382-388.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.

Sepúlveda, C.H., Sotelo-Gonzalez, M.I., Osuna-Martínez, C.C., Frías-Espericueta, M.G., Sánchez-Cárdenas, R., Bergés-Tiznado, M.E., Góngora-Gómez, A.M. & García-Ulloa, M. (2023). Biomonitoring of potentially toxic elements through oysters (Saccostrea palmula and Crassostrea corteziensis) from coastal lagoons of Southeast Gulf of California, Mexico: health risk assessment. Environmental Geochemistry and Health, 45(5), 2329-2348.

Shar, S., Reith, F., Ball, A. S., & Shahsavari, E. (2021). Long-term impact of gold and platinum on microbial diversity in australian soils. Microbial Ecology, 81, 977-989.

Shi, W., Liu, J., Du, Z., Stein, A., & Yue, T. (2011). Surface modelling of soil properties based on land use information. Geoderma, 162(3–4), 347–357.

Smiljanic, S., Tesan-Tomic, N., & Perusic, M. (2019). The main sources of heavy metals in the soil and pathways intake. In VI international congress engineering, Environment and Materials in Processing Industry, 453-465.

Smith, L.A. & Brauning, S.E., (1995). Remedial options for metals-contaminated sites 17-122. Boca Raton: CRC Press.

Song, X.D., Brus, D.J., Liu, F., Li, D.C., Zhao, Y.G., Yang, J.L., & Zhang, G.L. (2016). Mapping soil organic carbon content by geographically weighted regression: a case study in the Heihe River Basin, China. Geoderma, 261, 11–22.

Sowlat, M. H., Hasheminassab, S., & Sioutas, C. (2016). Source apportionment of ambient particle number concentrations in central Los Angeles using positive matrix factorization (PMF). Atmospheric Chemistry and Physics, 16(8), 4849-4866.

Steinnes, E., Rühling, Å., Lippo, H., & Mäkinen, A. (1997). Reference materials for large-scale metal deposition surveys. Accreditation and Quality Assurance, 2, 243-249.

Stemweis, P.C.J., Buss, J.E., Mumby, S.M., Casey, P.J., Gilman, A.G., Sefton, B.M., & Sugimoto, K., (1988). Quantitative assessment of worldwide contamination of air, water and soils by trace metals. Nature, 333, 134–139.

Suleymanov, A., Polyakov, V., Kozlov, A., Abakumov, E., Kuzmenko, P., & Telyagissov, S. (2023). Mapping of potentially toxic elements in the urban topsoil of St. Petersburg (Russia) using regression kriging and random forest algorithms. Environmental Earth Sciences, 82(23), 561.

Sun, L., Guo, D., Liu, K., Meng, H., Zheng, Y., Yuan, F., & Zhu, G. (2019). Levels, sources, and spatial distribution of heavy metals in soils from a typical coal industrial city of Tangshan, China. Catena, 175, 101–109.

Sun, R. & Chen, L. (2016). Assessment of heavy metal pollution in topsoil around Beijing metropolis. PLoS One, 11(5)

Sungur, A., Kavdir, Y., Özcan, H., İlay, R., & Soylak, M., (2021). Geochemical fractions of trace metals in surface and core sections of aggregates in agricultural soils, CATENA. 197, 104995.

Swartjes, F. A., & Siciliano, S. (2012). Dealing with contaminated sites: From theory towards practical application. Soil Science Society of America Journal, 76(2), 748–748

Taghizadeh-Mehrjardi, R., Fathizad, H., Hakimzadeh Ardakani, M. A., Sodaiezadeh, H., Kerry, R., Heung, B., & Scholten, T. (2021). Spatio-temporal analysis of heavy metals in arid soils at the

catchment scale using digital soil assessment and a random forest model. Remote Sensing, 13(9), 1698.

Tao, H., Al-Hilali, A.A., Ahmed, A.M., Mussa, Z.H., Falah, M.W., Abed, S.A., Deo, R., Jawad, A.H., Maulud, K.N.A., Latif, M.T. & Yaseen, Z.M. (2023). Statistical and spatial analysis for soil heavy metals over the Murray-Darling River basin in Australia. Chemosphere, 317, 137914.

Tchagang, A. B., & Valdés, J. J. (2019). Prediction of the atomization energy of molecules using coulomb matrix and atomic composition in bayesian regularized neural networks. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 11731 LNCS, 793–803.

Teng, Y., Li, J., Wu, J., Lu, S., Wang, Y., & Chen, H. (2015). Environmental distribution and associated human health risk due to trace elements and organic compounds in soil in Jiangxi province, China. Ecotoxicology and Environmental Safety, 122, 406–416.

Tomlinson, D.L., Wilson, J.G., Harris, C.R., & Jeffrey, D. W. (1980). Problems in the assessment of heavy-metal levels in estuaries and the formation of a pollution index. Helgoländer Meeresuntersuchungen, 33, 566-575.

Tóth, G., Hermann, T., Szatmári, G., & Pásztor, L. (2016). Maps of heavy metals in the soils of the European Union and proposed priority areas for detailed assessment. Science of the Total Environment, 565, 1054–1062.

Tu, C., Fan, W., Yang, S., & Luo, Y. (2024). Copper in the Soil. Inorganic Contaminants and Radionuclides, 95-111.

Tudor, V.C., Stoicea, P., Chiurciu, I.A., Soare, E., Iorga, A.M., Dinu, T.A., David, L., Micu, M.M., Smedescu, D.I. & Dumitru, E.A. (2023). The use of fertilizers and pesticides in wheat production in the main European Countries. Sustainability, 15(4), 3038.

Tyler, G., Påhlsson, A. M. B., Bengtsson, G., Bååth, E., & Tranvik, L. (1989). Heavy-metal ecology of terrestrial plants, microorganisms and invertebrates: a review. Water, Air, and Soil Pollution, 47, 189-215.

USDA-NRCS. (2018). Soil Health. https://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/health/

US DHHS. (2007). Toxicological profile for lead. ATSDR's Toxicological Profiles. https://doi.org/10.1201/9781420061888_ch106.

US DOE, (2011). The risk assessment information system (RAIS). U.S. Department of Energy's Oak Ridge Operations Office (ORO).

US EPA. 1989. Risk assessment guidance for superfund. Human Health Evaluation Manual, (Part A) Vol. 1. Office of Emergency and Remedial Response, Washington, DC, USA (EPA/540/1–89/002)

US EPA (U.S. Environmental Protection Agency), (1991). Risk Assessment Guidance for Superfund: Volume I Human Health Evaluation Manual (Part B, Development of Risk-based Preliminary Remediation Goals). EPA/540/R-92/003.

US EPA (U.S. Environmental Protection Agency), (2002). Risk-based concentration table. United States Environmental Protection Agency, Washington DC.

US EPA, (2010). EPA positive matrix factorization (PMF) 3.0 Model; U.S. Environmental Protection Agency: Research Triangle Park, NC, USA, Available online: http://www.epa.gov/heasd/products/pmf/pmf.html (accessed on 20 March 2022).

US EPA (U. S. Environmental Protection Agency), (2011). Exposure Factors Handbook: 2011 Edition. National Center for Environmental Assessment, Office of Research and Development, US Environmental Protection Agency, Washington, DC 20460, EPA/600/R-09/052F.

US EPA (U.S. Environmental Protection Agency), (2013). Region IX, regional screening levels (formerly PRGs). San Francisco, CA 94105. Available at: http://www.epa.gov/region9/superfund/prg/.

US EPA, (2014). Positive matrix factorization (PMF) 5.0-fundamentals and user guide. Washington: U.S. Environmental Protection Agency. Available online: https://www.epa.gov/sites/production/files/2015-02/documents/pmf_5.0_user_guide.pdf

Usoh, G. A., Isiguzo Edwin Ahaneku, E.C. Ugwu, E.O. Sam, D.H. Itam, George Uwadiegwu Alaneme, & Ndamzi, T.C. (2023). Mathematical modeling and numerical simulation technique for selected heavy metal transport in MSW dumpsite. Scientific Reports, 13 (1), 5674.

Vacek, O., Vašát, R., & Borůvka, L. (2020). Quantifying the pedodiversity-elevation relations. Geoderma, 373, 114441.

Vácha, R. (2021). Heavy Metal Pollution and Its Effects on Agriculture. Agronomy, 11(9), 1719.

Valko MMHCM, Morris H, Cronin MTD. (2005). Metals, toxicity and oxidative stress. Current Medicinal Chemistry, 12(10), 1161–1208.

Van der Meer, F. (2012). Remote-sensing image analysis and geostatistics. International Journal of Remote Sensing, 33(18), 5644–5676.

van der Westhuizen, D., Howlett-Downing, C., Molnár, P., Boman, J., Wichmann, J., & von Eschwege, K. G. (2024). Source apportionment of fine atmospheric particles in Bloemfontein, South Africa, using positive matrix factorization. Environmental Monitoring and Assessment, 196(2), 188.

Vašát, R., Pavlů, L., Borůvka, L., Drábek, O., & Nikodem, A. (2013). Mapping the topsoil pH and humus quality of forest soils in the north bohemian Jizerské hory mts. region with ordinary, universal, and regression kriging: Cross-validation comparison. Soil and Water Research, 8 (3), 97-104.

Velayatzadeh, M. (2023). Heavy metals in surface soils and crops. In Eds. Almayyahi, B.A. Heavy Metals-Recent Advances. IntechOpen. DOI: 10.5772/intechopen.108824.

Veronesi, F., Corstanje, R., & Mayr, T. (2014). Landscape scale estimation of soil carbon stock using 3D modelling. Science of The Total Environment, 487 (1), 578–586.

Vieira, E.V.R., do Rosario, N.E., Yamasoe, M.A., Morais, F.G., Martinez, P.J.P., Landulfo, E., & Maura de Miranda, R. (2023). Chemical Characterization and Optical Properties of the Aerosol in São Paulo, Brazil. Atmosphere, 14(9), 1460.

von Wachenfeldt, E., Sobek, S., Bastviken, D., Tranvik, L.J. (2008). Linking allochthonous dissolved organic matter and boreal lake sediment carbon sequestration: the role of light-mediated flocculation. Limnol Oceanogr, 53:2416–2426.

Walker, D.J., Clemente, R., Roig, A. & Bernal, M.P. (2003). The effects of soil amendments on heavy metal bioavailability in two contaminated Mediterranean soils. Environmental Pollution, 122(2), 303-312.

Wang, G., Liu, H.Q., Gong, Y., Wei, Y., Miao, A.J., Yang, L.Y. & Zhong, H., (2017). Risk assessment of metals in urban soils from a typical industrial city, Suzhou, Eastern China. International Journal of Environmental Research and Public Health, 14(9), 1025.

Wang, Z., Xiao, J., Wang, L., Liang, T., Guo, Q, Guan, Y., & Rinklebe, J. (2020). Elucidating the differentiation of soil heavy metals under different land uses with geographically weighted regression and self-organizing map. Environmental Pollution, 260, 114065.

Wang, J., Yang, R., & Bai, Z. (2015). Spatial variability and sampling optimization of soil organic carbon and total nitrogen for Minesoils of the Loess Plateau using geostatistics. Ecological Engineering, 82, 159–164.

Wang, J., Yuan, J., Hou, Q., Yang, Z., You, Y., Yu, T., Ji, J., Dou, L., Ha, X., Sheng, W. & Liu, X. (2023). Distribution of potentially toxic elements in soils and sediments in Pearl River Delta, China: Natural versus anthropogenic source discrimination. Science of The Total Environment, 903, 166573.

Wang, J., Yu, D.,Wang, Y., Du, X., Li, G., Li, B., Zhao, Y., Wei, Y., & Xu, S. (2021).Source analysis of heavy metal pollution in agricultural soil irrigated with sewage in Wuqing, Tianjin. Scientific Reports , 11, 17816.

Weather Spark. (2016). Average Weather in Frýdek-Místek, Czechia, Year-Round - Weather Spark. https://weatherspark.com/y/83671/Average-Weather-in-Frýdek-Místek-Czechia-Year-Round.

Webster, R., & Oliver, M.A. (2007). Geostatistics for Environmental Scientists. Second edition. John Wiley & Sons Ltd., Chichester.

Wedepohl, K. H. (1995). The composition of the continental crust. Geochimica et Cosmochimica Acta, 59, 1217–1223.

Wen, Y., Wang, Y., Ji, W., Wei, N., Liao, Q., Huang, D., Meng, X. & Song, Y. (2023). Influencing factors of elevated levels of potentially toxic elements in agricultural soils from typical Karst Regions of China. Agronomy, 13(9), 2230.

Were, K., Bui, D. T., Dick Ø. B., & Singh, B.R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecological Indicators, 52, 394–403.

Wu, H., Yang, F., Li, H., Li, Q., Zhang, F., Ba, Y., Cui, L., Sun, L., Lv, T., Wang, N. & Zhu, J. (2020). Heavy metal pollution and health risk assessment of agricultural soil near a smelter in an industrial city in China. International Journal of Environmental Health Research, 30(2), 174-186.

Wuana, R. A., & Okieimen, F. E. (2011). Heavy Metals in Contaminated Soils: A Review of Sources, Chemistry, Risks and Best Available Strategies for Remediation. ISRN Ecology, 2011.

Xu, X., Cao., Z., Zhang, Z., Li, R., Hu, B. (2016). Spatial distribution and pollution assessment of heavy metals in the surface sediments of the Bohai and Yellow Seas. Marine Pollution Bulletin, 110(1), 596–602.

Xue, S., Korna, R., Fan, J., Ke, W., Lou, W., Wang, J., & Zhu, F. (2023). Spatial distribution, environmental risks, and sources of potentially toxic elements in soils from a typical abandoned antimony smelting site. Journal of Environmental Sciences, 127, 780-790.

Yang, D., Yang, Y., & Hua, Y. (2023). Source analysis based on the positive matrix factorization models and risk assessment of heavy metals in agricultural soil. Sustainability, 15(17), 13225.

Yang, X., & Yang, Y. (2023). Spatiotemporal patterns of soil heavy metal pollution risk and driving forces of increment in a typical industrialized region in central China. Environmental Science: Processes & Impacts, 25(3), 554-565.

Yigini, Y., & Panagos, P. (2016). Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. Science of The Total Environment, 557–558, 838–850.

Yigini, Y., Olmedo, G.F., Reiter, S., Baritz, R., Viatkin, K., &Vargas, R. (Eds.), (2018). Soil Organic Carbon Mapping Cookbook. FAO, Rome (220).

Yuan, B., Cao, H., Du, P., Ren, J., Chen, J., Zhang, H., Zhang, Y. & Luo, H. (2023). Source-oriented probabilistic health risk assessment of soil potentially toxic elements in a typical mining city. Journal of Hazardous Materials, 443, 130222.

Zaib, M., Ibrahim, M., Aryan, M., Mustafa, R., Zubair, M., Mumtaz, S., & Hussain, T. (2023). Long-term efficacy of biochar-based immobilization for remediation of heavy metal-contaminated soil and environmental factors impacting remediation performance. International Journal of Scientific Research and Engineering Development, 6(5):58-71.

Zawadzki, J., Cieszewski, C.J., Zasada, M., & Lowe, R.C. (2005). Applying geostatistics for investigations of forest ecosystems using remote sensing imagery. Silva Fennica Monographs, 39(4), 599–617.

Zhang, G. lin, LIU, F., & Song, X. D. (2017). Recent progress and future prospect of digital soil mapping: A review. Journal of Integrative Agriculture, 16(12), 2871–2885.

Zhang, J., Li, H., Zhou, Y., Dou, L., Cai, L., Mo, L., & You, J. (2018). Bioavailability and soil-to-crop transfer of heavy metals in farmland soils: A case study in the Pearl River Delta, South China. Environmental Pollution, 235, 710–719.

Zhang, M.K., Liu, Z.Y., & Wang, H. (2010). Use of single extraction methods to predict bioavailability of heavy metals in polluted soils to rice communications in soil. Science and Plant Analysis, 41(7), 820–831.

Zhang, W. & Goh, A.T. (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geoscience Frontiers, 7(1), 45-52.

Zhao, L., Xu, Y., Hou, H., Shangguan, Y. & Li, F., (2017). Source identification and health risk assessment of metals in urban soils around the Tanggu chemical industrial district, Tianjin, China. Science of the Total Environment, 468, 654-662.

Zhou, L., Zhao, X., Meng, Y., Fei, Y., Teng, M., Song, F., & Wu, F. (2022). Identification priority source of soil heavy metals pollution based on source-specific ecological and human health risk analysis in a typical smelting and mining region of South China. Ecotoxicology and Environmental Safety, 242, 113864.

Zhu, A.X., Hudson, B., Burt, J., Lubich, K., & Simonson, D. (2001). Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal, 65(5), 1463–1472.

Zhu, A.X., Lu, G., Liu, J., Qin, C.Z., & Zhou, C. (2018). Spatial prediction based on the Third Law of Geography. Annals of GIS, 24(4), 225–240.

Zimmerman, D.L., & Homer, K.E. (1991). A network design criterion for estimating selected attributes of the semivariogram. Environmetrics, 2(4), 425-441.

6.0 LIST OF PUBLICATIONS

1.Ahado, S.K., & Nwaogu, C. (2023). Spatial modelling and quantification of soil potentially toxic elements based on variability in sample size and land use along a toposequence at a district scale. International Journal of Environmental Science and Technology,1-20. *Impact factor 3.2 / Q3*

2.Ahado, S.K., Nwaogu, C., Sarkodie, V.Y.O. & Borůvka, L. (2021). Modeling and assessing the spatial and vertical distributions of potentially toxic elements in soil and how the concentrations differ. Toxics, 9(8), 181. *Impact factor 4.8 / Q2*

3. Ahado, S. K., Agyeman, P. C., Borůvka, L., Kanianska, R., & Nwaogu, C. (2023). Using geostatistics and machine learning models to analyze the influence of soil nutrients and terrain attributes on lead prediction in forest soils. Modeling Earth Systems and Environment, 1-14. *Impact factor 3.2 / Q3*

4. Agyeman, P.C., Ahado, S.K., John, K., Kebonye, N.M., Vašát, R., Borůvka, L., Kočárek, M. & Němeček, K. 2021. Health risk assessment and the application of CF-PMF: A pollution assessment–based receptor model in an urban soil. Journal of Soils and Sediments, 21(9), 3117-3136. *Impact factor 3.8 Q2*

5. Agyeman, P.C., Ahado, S.K., Borůvka, L., Biney, J.K.M., Sarkodie, V.Y.O., Kebonye, N.M. & Kingsley, J. (2021). Trend analysis of global usage of digital soil mapping models in the prediction of potentially toxic elements in soil/sediments: a bibliometric review. Environmental Geochemistry and Health, 43, 1715-1739. *Impact factor 4.4 Q2*

6. Agyeman, P.C., Ahado, S.K., Kingsley, J., Kebonye, N.M., Biney, J.K.M., Borůvka, L., Vašát, R. & Kočárek, M. (2021). Source apportionment, contamination levels, and spatial prediction of potentially toxic elements in selected soils of the Czech Republic. Environmental geochemistry and health, 43, 601-620. *Impact factor 4.4 Q2*

7. Kebonye, N.M., Eze, P.N., Ahado, S.K. & John, K. (2020). Structural equation modeling of the interactions between trace elements and soil organic matter in semiarid soils. International Journal of Environmental Science and Technology, 17, 2205-2214. *Impact factor 3.2 Q3*

8. John, K., Kebonye, N.M., Agyeman, P.C. & Ahado, S.K. (2021). Comparison of Cubist models for soil organic carbon prediction via portable XRF measured data. Environmental monitoring and assessment, 193, 1-15. *Impact factor 3.1 Q3*

9. Kebonye, N.M., John, K., Chakraborty, S., Agyeman, P.C., Ahado, S.K., Eze, P.N., Němeček, K., Drábek, O. & Borůvka, L. (2021). Comparison of multivariate methods for arsenic estimation and mapping in floodplain soil via portable X-ray fluorescence spectroscopy. Geoderma, 384, 114792. . *Impact factor 7 Q1*

10. Agyeman, P.C., John, K., Kebonye, N.M., Ahado, S.K., Borůvka, L., Němeček, K. & Vašát, R. (2021). Multi-geochemical background comparison and the identification of the best normalizer for the estimation of PTE contamination in agricultural soil. Environmental Geochemistry and Health, 1-17. *Impact factor 4.4 Q2*

11. Kanianska, R., Drimal, M., Varga, J., Komárek, M., Ahado, S.K., Šťastná, M., Kizeková, M. & Jančová, Ľ. (2023). Critically raw materials as potential emerging environmental contaminants, their distribution patterns, risks and behaviour in floodplain soils contaminated by heavy metals. Scientific Reports, 13(1). **Impact factor 4.9 Q2**

12. Asare, A.O., Sarpong, E.O., Truong holds, N., Osei-Bonsu, P., Ahado, S. & Mensah, W.G. (2022). COVID-19 pandemic and African innovation: Finding the good from the bad using Twitter data and text mining approach. International Social Science Journal, 73(250), 959-978. **Impact factor 0.36 Q2**

13. Nwaogu, C., Olawoyin, M.A., Ahado, S.K., Wallace, E. & Gardiner, R. (2016). Socio-economic Implications of Climate Seasonality: a comparative assessment of Road Transport inequalities between Rural and Urban areas. International Journal of Technology Enhancements and Emerging Engineering Research, 4(1). **Impact factor 0.40**

14. Olawoyin, M.A., Nwaogu, C., Nakashole, P., Ahado, S., Toseafa, H.K. & Gardiner, R. (2015). Environmental and socio-economic impacts of solid waste dumping and burning on the major motorways in developing countries. International Journal of Recent Advances in Multidisciplinary Research, 2(12), 999-1007. **Impact factor 4.9**

15. Kebonye, N.M., Eze, P.N., Agyeman, P.C., John, K. & Ahado, S.K. (2021). Efficiency of the t-distribution stochastic neighbor embedding technique for detailed visualization and modeling interactions between agricultural soil quality indicators. Biosystems Engineering, 210, 282-298. **Impact factor 5.5 Q1**

7.0 PUBLISHED PAPERS ATTACHED

1. Agyeman, Prince Chapman, **Samuel Kudjo Ahado**, Luboš Borůvka, James Kobina Mensah Biney, Vincent Yaw Oppong Sarkodie, Ndiye M. Kebonye, Kingsley John. (2021). *Trend analysis of global usage of digital soil mapping models in the prediction of potentially toxic elements in soil/sediments: a bibliometric review.* **Environmental Geochemistry and Health,** *43: 1715-1739.*

2. **Samuel Kudjo Ahado**, *Agyeman, Prince Chapman, Luboš Borůvka, Radoslava Kanianska, Chukwudi Nwaogu. (2023). Using geostatistics and machine learning models to analyze the influence of soil nutrients and terrain attributes on lead prediction in forest soils.* **Modeling Earth Systems and Environment,** *10:2099-2112.*

3. **Samuel Kudjo Ahado**, *Chukwudi Nwaogu. (2023). Spatial modelling and quantification of soil potentially toxic elements based on variability in sample size and land use along a toposequence at a district scale.* **International Journal of Environmental Science and Technology**, *20: 1-20.*

4. Agyeman, Prince Chapman, **Samuel Kudjo Ahado**, Kingsley John, Ndiye Michael Kebonye, James Kobina Mensah Biney, Luboš Borůvka, Radim Vašát, Martin Kočárek. (2021). *Source apportionment, contamination levels, and spatial prediction of potentially toxic elements in selected soils of the Czech Republic.* **Environmental Geochemistry and Health,** *43: 601-620.*

5. **Samuel Kudjo Ahado**, *Chukwudi Nwaogu, Vincent Yaw Oppong Sarkodie, Luboš Borůvka. (2021). Modeling and assessing the spatial and vertical distributions of potentially toxic elements in soil and how the concentrations differ.* **Toxics**, *9: 181-187.*

6. Agyeman, Prince Chapman, **Samuel Kudjo Ahado**, Kingsley John, Ndiye Michael Kebonye, Radim Vašát, Luboš Borůvka, Martin Kočárek, Karel Němeček. (2021). *Health risk assessment and the application of CF-PMF: A pollution assessment–based receptor model in an urban soil.* **Journal of Soils and Sediments,** *21 (9): 3117-3136*

REVIEW PAPER

Trend analysis of global usage of digital soil mapping models in the prediction of potentially toxic elements in soil/sediments: a bibliometric review

**Prince Chapman Agyeman · Samuel Kudjo Ahado · Luboš Borůvka ·
James Kobina Mensah Biney · Vincent Yaw Oppong Sarkodie · Ndiye M. Kebonye ·
John Kingsley**

Abstract The rising and continuous pollution of the soil from anthropogenic activities is of great concern. Owing to this concern, the advent of digital soil mapping (DSM) has been a tool that soil scientists use in this era to predict the potentially toxic element (PTE) content in the soil. The purpose of this paper was to conduct a review of articles, summarize and analyse the spatial prediction of potentially toxic elements, determine and compare the models' usage as well as their performance over time. Through Scopus, the Web of Science and Google Scholar, we collected papers between the year 2001 and the first quarter of 2019, which were tailored towards the spatial PTE prediction using DSM approaches. The results indicated that soil pollution emanates from diverse sources. However, it provided reasons why the authors investigate a piece of land or area, highlighting the uncertainties in mapping, number of publications per journal and continental efforts to research as well as published on trending issues regarding DSM. This paper reveals the complementary role machine learning algorithms and the geostatistical models play in DSM. Nevertheless, geostatistical approaches remain the most preferred model compared to machine learning algorithms.

Keywords Digital soil mapping · Spatial prediction · Geostatistics · Machine learning · Algorithms · Potentially toxic elements · Soil pollution

Introduction

Potentially toxic elements (PTEs) are abundant natural components of the earth's crust soils (Kabata-Pendias and Mukherjee 2007; Iñigo et al. 2011). PTE is a generic nomenclature given to poisonous metal(loid)s that are detrimental to either human well-being or sustainable environment or both. The term soil contamination alludes to the presence of a chemical(s) or strange substance higher than the average concentration that has adverse effects on any non-targeted organism (FAO and ITPS 2015). Although part of PTEs has anthropogenic sources, some contaminants can happen naturally in soils as components of minerals and can be toxic at high concentrations. Soil contamination cannot be regularly evaluated or outwardly seen, making it a concealed threat. The diverse variety of contaminants is continuously advancing due to agrochemical and industrial developments. Nevertheless, the impacts of soil contamination also depend upon soil properties since these controls the mobility, bioavailability and residence

P. C. Agyeman (✉) · S. K. Ahado ·
L. Borůvka · J. K. M. Biney · V. Y. O. Sarkodie ·
N. M. Kebonye · J. Kingsley
Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamýcká 129, 165 00 Praha 6, Suchdol, Czech Republic
e-mail: agyeman@af.czu.cz

time of PTEs (FAO and ITPS 2015). Industrialization, wars, mining and intensification in farming have left a legacy of contaminated soils around the entire world (Luo et al. 2009a, b; Bundschuh et al. 2012). Since the urban expansion, the soil has been utilized as a repository for the dumping of solid and liquid waste. It was well thought out that once buried and out of sight, the contaminants would not pose any danger to human health or the environment and that they would somehow disappear (Swartjes and Siciliano 2012). The fundamental sources of soil contamination are anthropogenic, resulting in the sowing of contaminants in soils that may reach levels of concern (Cachada et al. 2017). Source of pollutant occurs from different sources, namely natural enrichment, agricultural activities (land application of fertilizers, animal manures, composts, pesticides), industrial activities, transportation system, atmospheric deposition, waste management and treatment and mining (Basta et al. 2005; Khan et al. 2008; Zhang et al. 2010; Ballesta et al. 2017). PTEs of anthropogenic sources are typically more mobile and bioavailable in soil than PTEs of lithogenous or pedogenic origin (Kuo et al. 1983; Kaasalainen and Yli-Halla 2003). According to Seaward and Richardson (1990) and Thevenon et al. (2011), natural processes such as weathering of rocks, erosion, rock formation and volcanic eruption play a major role in the emission and exposure of enormous quantity of PTEs such as Al, Cu, Hg, Mn, Ni and Zn into the environment particularly soil. Scragg (2006) reported that agricultural production was the foremost human influence exerted on the soil. The ever-growing human population is the fulcrum that pushes farmers to produce more as well as apply agrochemicals such as fertilizers, pesticides or animal manure to enhance yield and productivity. Application of agrochemicals like foliar sprays that is rich in PTEs, for instance, Co, Cu, Fe, Mn, Mo, Ni and Zn, to soil essentially for plant growth (Lasat 1999), successively during every crop season elevates the PTE concentration in the soil. However, recent publications by Nicholson et al. (2003), Luo et al. (2009a, b) and Liang et al. (2017) suggested that anthropogenic activities related to agronomic practices such as the use of fertilizers, fungicides and fossil fuel combustion practices have contributed more to the high accumulation of Cu, Hg, Mn, Pb or Zn in soils. For example, lead arsenate and arsenate compounds used to control pests in fruit orchards in New Zealand and Australia are rich in Cu,

Cr and As (Wuana and Okieimen, 2014) these elements are likely to increase the concentration of PTEs in soil beyond the tolerable limits. Basta et al. (2005) recounted that the application of biosolids like sewage sludge, industrial waste and compost to agricultural fields results in the increment of PTEs such as As, Cd, Cr, Cu, Pb, Hg, Ni, Se, Mo, Zn, Tl and Sb in the soil. Industrial activity coupled with mining as well as tailings discharges a large amount of PTEs that contaminates the soil. For instance, huge lead (Pb) and zinc (Zn) ore mining, as well as metal smelting, has the propensity in contaminating the soil and poses an ecological risk. According to the FAO and ITPS (2018) reports, the United Nations Environmental Assembly (UNEA-3) agreed on a resolution calling for expedite actions and collaboration to address and manage soil pollution worldwide. The report further articulated that a total consensus was reached by all the member nations acknowledging the global significance of soil pollution and the readiness of these countries to develop concrete solutions to address the causes and impacts of this significant threat. The United Nations Environment Programme (UNEP) and the International Soil Research and Information Centre (ISRIC) estimated that 22 million hectares had been affected by soil pollution (Oldeman 1991). However, the latest data indicated that this number might underestimate the nature and extent of the problem (FAO and ITPS 2018). The frightening rate of soil pollution has been determined as the third most vital danger to soils in Europe, fourth in North Africa, fifth in Asia, seventh in the Northwest Pacific, eighth in North America and ninth in sub-Saharan Africa and Latin America (FAO and ITPS 2015). The occurrence of certain pollutants may likely create nutrient imbalances and soil acidification. Approximately 3 million potentially polluted sites that are located in the European Economic Area and cooperating countries in the Western Balkans (EEA-39) (EEA 2014) and more than 1,300 contaminated sites in the USA are included in the Superfund National Priorities List (US GAO 2015).

A crucial role in soil science, environmental science and ecology is to research spatial or geotemporal trends of ecosystem variables, e.g. climate dynamics (Appelhans et al. 2015), soil properties variability (Gasch et al. 2015) or vegetation form distribution (Juel et al. 2015). Spatially continuous datasets of soil/sediment variables are required to

examine spatial trends and dynamics. The application of a rich geospatial dataset plays a critical role in solving several major societal issues; but, due to the specific characteristics of spatial data, it is also technically challenging. Jiang (2018) stated that there is an urgent need for effective and accurate prediction methods that could unlock the values of such rich geospatial data properties.

Digital soil mapping (DSM) or predictive soil mapping is presently the most effective way to predict the spatial variation of soil/sediment over an area (McBratney et al. 2003). According to Minasny and McBratney (2016), DSM or predictive soil mapping has become a successful subdiscipline of soil science. Iqbal et al. (2005) stated that spatial variability of soil physical properties within or between soils is, at most times, inherent in nature due to geological and pedological soil formation factors, although some of the variability may be caused by other management practices. The factors work together on a temporal and spatial scale, and the content, is further adjusted by the spatial heterogeneity deposition of soil properties. Zhu et al. (2018) reported that environmental covariates and soil relationship in spatial predictions are fitted with a model as well as the learned nexus and are subsequently applied to spaces or locations that data (soil/sediment data) are unknown. Usually, DSM forms a quantitative soil environment relationship centred on the modelling points or sample observation points to characterize the nexus between soil and environmental covariates such as climate variables, geological variables, slope and topographic wetness index. DSM applies models to compute soil property values at unknown locations (Zhu et al. 2001; McBratney et al. 2003; Minasny and McBratney 2016; Heung et al. 2016). Globally, the soil science communities have adopted DSM for mapping soil properties and classes (Arrouays et al. 2014) as well as to a significant extent predict the concentration of PTEs in the soil/sediments. Due to its high accuracy compared to conventional mapping, many stakeholders (FAO) have embraced the usage of DSM. DSM is consistent for sustainable land management (Padarian et al. 2019), and by extension, it is useful and efficient in the spatial prediction of PTEs. Significant to the success and applicability of spatial predictions are the underlying assumptions employed in describing the relationships and how these relationships are characterized (Zhu et al. 2018). Soil mapping techniques

have generally improved by the progression of geographical information technology and computational technology (Zhang et al. 2017). Lagacherie and McBratney (2006) defined digital soil mapping as the creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from soil observation and knowledge from related environmental variables. The accumulation of PTEs in the soils/sediments has been a worldwide concern (Liu et al. 2003; Gonz´alez et al. 2006), as it poses an utmost threat to human health (Chen et al. 2015). According to Chen et al. (2009), one of the feasible roles of studies is the inhibition of PTEs in the soil. On the other hand, spatial prediction of PTEs provides an avenue to delineate the distribution of potentially toxic elements, their concentration, occurrence and possibly knowing their source of pollution. There are substantial proven research and published papers showing the prediction of PTEs in soils/sediments using DSM approaches.

The successes of DSM at this present day compared to conventional soil mapping are not in doubt. We hypothesized that the option of spatial prediction or DSM models for the prediction of PTEs depends primarily on environmental factors, the preference of the author and the prevailing conditions accessible. However, what is the global reach of DSM? How accessible is DSM in the developing and least developed countries? This paper is seeking to assess the trend and the global usage of DSM models in the prediction of PTEs in soils and sediments. It brings together a list of articles published on the distribution of PTEs from various journals using DSM approaches. DSM has provided a more significant deal of correlation and has a superior effect of spatially predicting PTEs in the soil and sediment. This work will submit a quantitative review of well-proven and tested scientific algorithms that have been used to predict the distribution of PTEs in soils and sediments spatially. The emphasis is on summarizing the spatial prediction of PTEs between 2001 and 2019 to analyse, determine and compare model usage as well as their performance.

Materials and methods

We collected various articles from different databases such as Scopus, Web of Science and Google Scholar. We identified some keywords that were geared towards the topic under review, and the keywords were tuned consistently to ascertain articles that are relevant to our search. The number of keywords that were combined includes Soils*/sediments* AND ("digital soil mapping" OR "digital mapping" OR "spatial prediction" OR "predictive mapping") AND ("toxic elements" OR "risk elements" OR "heavy metals" OR pollution OR Arsenic OR Beryllium OR Cobalt OR Cadmium OR Copper OR Chromium OR Mercury OR Molybdenum OR Nickel OR Pb OR Antimony OR Zinc"). This paper made use of only completed articles that were duly published within the ambit of prominent journals of which all the required information needed was collected for the analysis. No unpublished papers or articles were included in this review. The period for this review span between 2001 and the first quarter of 2019. Over 523 articles were downloaded, and we thoroughly read through to ascertain 319 articles that were relevant to the study based on these following criteria:

- Spatial prediction or interpolation by predictive models
- Potentially toxic elements (PTEs)
- Soil or sediment-based analysis
- Geostatistical model or
- Remote sensing or
- Machine learning model

Data filtering and mining

In the data mining and filtering process, many parameters were considered, such as spatial prediction using either machine learning algorithm or geostatistics. Furthermore, under each parameter, we additionally provided subtopics that aided us in collecting the needed information allowing us to have diverse as well as copious data to work with. We carefully reviewed the 319 articles and eliminated some articles based on the quality and relevance to our topic. Finally, we reduced it to 238 papers that are highly relevant to our research, comprising 208 soil-based analyses and 30 sediment-based researches. Below are the headings and information used in the data filtering and mining process (Table 1).

Table 1 List of information/options on which information was collected from papers on PTE modelling and mapping

Headings	Information collected, option
Paper identification	Authors, journal and year
Auxiliary data remote sensing	Remote sensing (RS), Remote sensing airborne (RS airbone), Remotes sensing drone(RS drones), laboratory spectroscopy (VisNIR), laboratory spectroscopy (MIR), field spectroscopy (VisNIR), field spectroscopy (MIR), others)
Region	Europe w/o Russia, Northern America, Southern America, Africa, Australia, Russia, China, Asia (without China)
Objective	Mapping, temporal changes/development, spatial distribution/variation assessment, comparison of methods, others
Studied PTEs	Arsenic, beryllium, cobalt, cadmium, chromium, mercury, molybdenum, nickel, lead, zinc, antimony, others
PTE source	Geology, atmospheric deposition, mining, metallurgy, industry, transportation, urban, agriculture (pesticides, fertilizers, etc.), flooding, alluvial areas, others
Modelling methods	Inverse distance weighting (IDW), geostatistics, kriging (ordinary, simple, universal), regression kriging, fuzzy methods, multiple linear regression (MLR), cubist, polynomial models, Bayesian methods, (boosted) regression trees (RTs), classification/decision trees, random forests (RFs), partial least squares regression (PLSR), principal component regression (PCR), artificial neural networks (ANNs), multiple additive regression splines (MARSs), support vector machines (SVMs), other machine learning techniques
Uncertainty assessment	Leave-one/group-out, validation subsample, independent validation samples, number of validations, maximum R^2 obtained (validation), minimum root mean square error (RMSE) validation (wt), minimum root mean square error (RMSE) validation ((t/ha)

Results and discussion

Geographical distribution of study areas

Two hundred and thirty-eight articles were collected from various journals across the globe, featuring countries on the six continents. Out of the six continents, there were 46 countries. The largest number of countries included came from Europe (without Russia), 21 countries, Asia (without China) 11, Africa 5, North America 3, South America 2, Australia and New Zealand, China and Russia. The percentage distribution of the number of papers from each continent on a global scale is shown on a pie chart (see Fig. 1). The articles reviewed presented countries from different continents, and countries like Russia as well as China were segregated to stand on its own to ascertain the influx of articles coming from those countries. China was isolated based on both population and land, and Russia was separated based on the land area in Europe. Out of 238 articles, 9 articles were from Africa accounted for just 3.78% of the chart, 49 from Asia (without China) for 20.59%, 77 from Europe (without Russia) for 32.35%, 5 from Southern America for 2.10%, 33 from North America for 13.89%, 8 from Australia and New Zealand for 3.36%, 8 from Russia for 3.36% and 8 from Russia for 3.3%, and lastly, from China 49 articles represent 20.59% of the chart. Even though Europe had the largest share of the chart (32.35%), it took 21 countries to get that share. China, on the other hand, accounted for 20.59% of the chart, which is the same as the percentage of Asia (without China). It is important to note that China alone contributed more than one-fifth

of the articles collected for this review, juxtaposing with Europe (without Russia), which had accumulated 32.4%. Similarly, other countries such as the USA, Iran, Spain, Italy and Mexico made a significant contribution to 20, 19, 14, 14 and 11 papers respectively.

Global concern about PTEs has necessitated a great deal of research to be carried out in other areas with higher levels of contaminants. The current distribution of the studies and publications coming from continents indicated that European countries place premium in investigating PTE spatial distribution. The prevalence of PTE research in Europe is widespread, not limited to a few countries, but quite several countries (Austria, Belgium, Croatia, Czechia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, the Netherlands, Norway, Poland, Portugal, Serbia, Spain, Slovenia, Sweden Switzerland and UK). Although Europe has the highest number of countries doing research and publishing on PTEs, Asia (without China) has made a significant contribution to PTE research.

The countries that featured in Asia (without China) include Hong Kong, India, Iran, South Korea, Malaysia, Mongolia, Pakistan, Qatar, Taiwan, Thailand and Turkey. Among the countries that made publications, China published 49 articles, which is equal to the number of publications from the other Asian countries. This presupposes that the rate of research and publication on PTE spatial distribution in China is very high. Gunson et al. (2001) reported that China is a global producer and a consumer of metal(loids) such as antimony (Sb), iron (Fe), lead (Pb), manganese (Mn), tin (Sn), tungsten (W) and zinc (Zn) as well as resources like coal. The extraction and

Fig. 1 Pie chart representing the distribution of studies among countries per continents
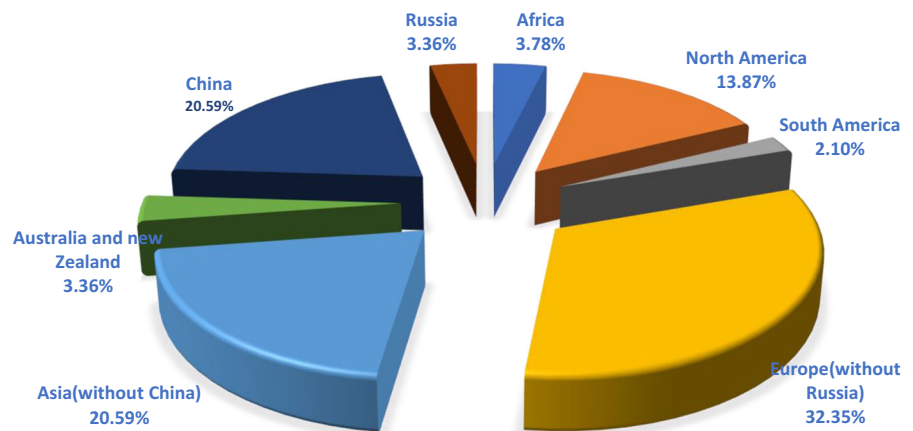
consumption of these minerals will always incur damages to the soil and the environment. However, Chen et al. (1999) recounted that the primary sources of soil pollution from PTEs in China are sewage irrigation, sludge use, mining and smelting operations for metallic ores. Yang et al. (2018) noted that the study (six-year soil pollution study) conducted by the Chinese government found that the country's soil was heavily polluted by anthropogenic activities, such as human activities in the industrial, mining and agricultural sectors.

North America and Australia and New Zealand have made a significant contribution to soil science and DSM in general. However, having access to many published papers from these continents (particularly countries such as Australia and New Zealand and the USA) is quite complicated. DSM has made a lot of progress, but a lot of challenges limits its use in developing countries. Zijl (2019) reported that the exploitation of the tools provided by DSM has been low in developing countries such as Africa. It presupposes that the developed nations have embraced the use of the tools provided by the DSM. On the contrary, the number of developing nations according to the International Statistical Institute Report (ISI 2020) stands at 137 countries, which implies that the global reach of the DSM is low. The slow nature of most developing nations to adopt DSM tools is likely to be due to the lack of technical expertise, lack of logistics and minimal interest in DSM. For example, the West

African subregion may hold periodic meetings and conferences to engage with experts, encourage TOT (Trainer of Trainers), build confidence in practitioners and work together to enhance knowledge transfer across developing countries in other ways to enhance efficiency and modern ways of mapping.

Publishing journals.

Articles were drawn from soil science, soil contamination or pollution, spatial prediction and related journals repository. Although there were 80 journals in which all the papers were taken from, the bar chart (Fig. 2) indicates 13 individual journals with the highest number of papers in this review and the 13th is labelled with others. The 13th bar on the chart consists of 68 different journals in which the number of articles reviewed is below six. Geoderma journal gave 25 articles. Also, the science of the Total Environment, Environmental Monitoring & Assessment, the Journal of Geochemical Exploration, Environmental Earth Science and Environmental Pollution, have published 10 articles and more, respectively. The 68 journals that were labelled others, 5 articles were taken from one journals, 4 articles were taken each from 3 journals, 3 articles were taken from 5 journals, 2 articles were taken from 10 journals and one article each was taken from 49 journals. Geoderma is a journal that is part of Elsevier's multifaceted journals, which focuses on soil science and encircles all aspects of soil science and related pedometrics articles. The journal

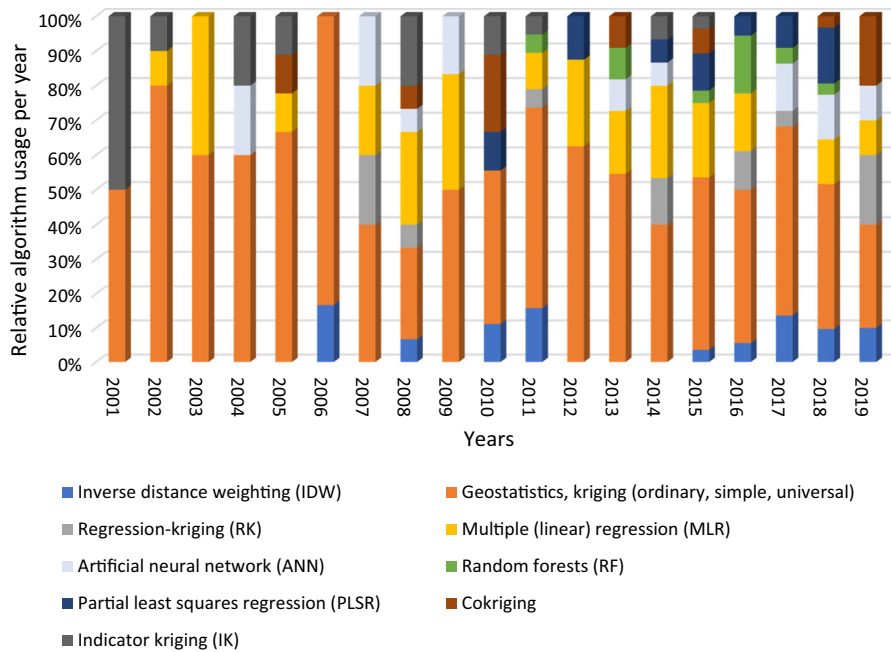**Fig. 2** Number of papers from selected journals
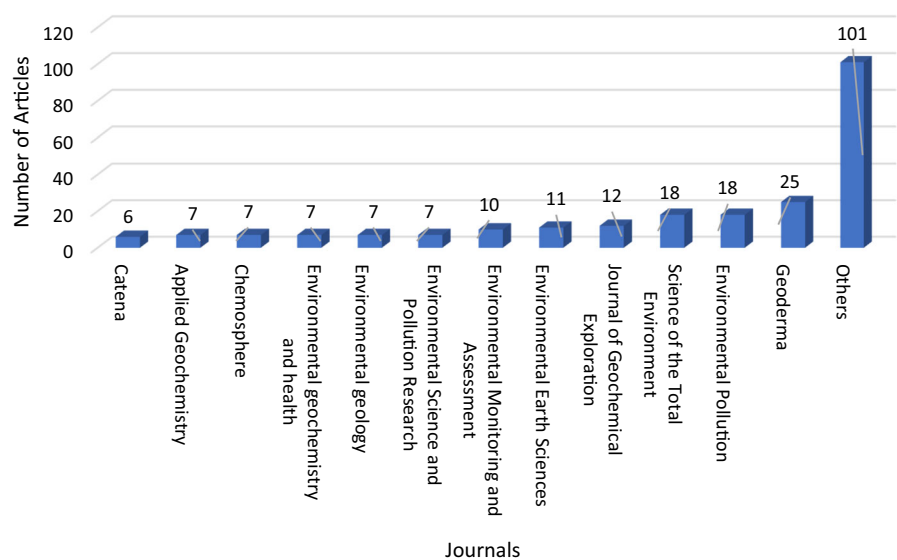
Fig. 3 Relative usage of DSM algorithms for PTE spatial prediction

(Geoderma) also provides technical reports on soils from a variety of environments, such as agricultural, urban and forest soils. Alternatively, the Science of the Total Environment, which has given us 20 articles for this review, is also from Elsevier. Its scope of research cuts across environmental issues related to the atmosphere, the hydrosphere, the biosphere, the lithosphere and the anthroposphere.

Modelling methods

The data mining process captured a range of modelling techniques used to predict contamination levels in a variety of ways and to express or interpret the intent on how the soil has been polluted. The DSM models included in this paper include inverse distance weighting, kriging (simple, ordinary, universal), indicator kriging, regression, kriging, fuzzy methods, multiple (linear) regression, cubist, classification and regression trees, random forest, partial least square regression, principal component regression, artificial neural network, multivariate adaptive regression splines and support vector machine. Sixteen modelling techniques that are widely used in DSM were captured between 2001 and the first quarter of 2019. Considering the performance of the algorithms, only 9

models included in the 16 algorithms were worthy of use in the bar chart. These 9 models consisted of 4 geostatistical models (kriging (simple, ordinary, universal), regression kriging, coking and indicator kriging), MLA (random forest, partial least square regression, artificial neural network), multiple linear regression and inverse distance weighting. The kriging (simple, ordinary, universal) model was the most preferred model of choice by most authors to the other algorithms. Over the 18 years of this review, it became apparent that kriging (simple, ordinary, universal) model usage exceeded all algorithm use, as shown in Fig. 3. Conversely, regression kriging, multiple linear regression and artificial neural networks have been used quite a few times. Random forest, partial least square regression, multiple linear regression and inverse distance weighting were relatively used.

Geostatistics is statistics that relies on spatial/temporal modelling data as well as focuses on the high-precision estimation with quantifiable uncertainty. Machine learning, on the other hand, is relatively a new technique, other than geostatistics, which some schools of thought assume that it gives a high rate of predictive accuracy concerning the type of model used. The spread of MLA use has increased in recent years, and some of the DSM experts believe that

there is a gradual paradigm shift from the use of the geostatistical model to machine learning techniques. Although MLA is becoming more popular in DSM due to the availability of computational power (Rossiter 2018), geostatistical kriging is the most commonly used in prediction (Webster and Oliver 2007). Although the geostatistical model remains the citadel of the DSM juxtaposing it with the papers reviewed, it is evident that geostatistics is still widely used in the soil sciences community. From 2011 until now, the soil science community is having a flair in the blending of the usage of MLA and geostatistical models. There is no doubt about the popularity of the kriging model and its usage in spatial prediction in DSM, as Veronesi and Schillaci (2019) have recently argued. Conversely, to Veronesi and Schillaci earlier arguments, they went on to report that practitioners relied less on interpolation (kriging models), but rather more on MLA. This assertion reinforces the fact that algorithms could be used based on the preference and diverse conditions available at the time. Numerous studies have been conducted to compare the two algorithms to determine the contrast and superiority between geostatistics and MLA. It is important to note that, for academic findings, it is possible to compare algorithms within DSM to find the best results that fit the modelling that a practitioner or author is trying to model.

Recent papers published by different authors have revealed in various articles that the author may conclude a comparative analysis based on the prevailing exigencies. For example, Beguin et al. (2017) conducted a comparative study of 8 algorithms and, following their research, concluded that ordinary kriging performed better than MLA (random forest, boosted regressing tree). In the same year in South Korea, Rhee and Im (2017) reported in their research a piece of conclusive findings that MLA (random forests, decision tree) showed impressive results compared to their counterpart kriging. The essence of the selection of algorithms enables the modeller or author to decide on the best-fit model for a study area that is mainly dependent on location, environmental factors and the author's decision on the algorithm to choice. The intercourse between geostatistics and MLA has resulted in a lot of versatility in the DSM space. It is also worth not divorcing the fact that both algorithms have some degree of uncertainty. However, the soil science scientific community argues that MLA

has gained more popularity than geostatistics in recent times, regardless of the lack of evidence to show a significant difference in accuracy (Veronesi and Schillaci 2019).

The geostatistical model enables scientists or authors to work and play with several assumptions. Geostatistics is mathematics that deals with finding a link between variables to predict the outcome. A machine learning algorithm, on the other hand, is just an algorithm that can learn and work on iterations, where the computer tries to find patterns hidden in the data (Srivastava et al. 2015). However, geostatistical modelling is usually applicable to data with fewer attributes or ends up being overfitting. Many modellers may prefer geostatistics to MLA because geostatistics is based on intensive mathematics and coefficient estimation. For example, for a modeller to use geostatistics, the modeller must understand the relationship between the variables before entering them. There is a significant human effort in geostatistics that allows humans to play a vital role in prediction, while minimal human effort is needed in MLA.

Land-use type and sources of PTES

Soil pollutants are introduced into the soil by natural processes or by anthropogenic means. Most of the areas studied by the respective authors in their articles ranged from cropland, grassland, forests, urban areas, mining areas, industrial sites, recreation areas, peri-urban areas and other host areas. In certain countries, the type of pollutants introduced into the soil/sediments arose predominantly from anthropogenic activities. The clustered bar (Fig. 4) shows that the pollution that is introduced into the soil has occurred through a multiplicity of events. A high number of articles have focused their attention on cropland or arable land because it presents a high percentage of land, and its pollution with PTEs poses a threat to the food chain. However, an increase in the number of publications on urban, industrial, forestry and mining is likely to be due to an increase in pollution levels and most likely being the most polluted areas. Cropland or arable land has been cited 125 times as being affected by PTEs, followed by urban areas, and recreation areas are the least contaminated areas on the chart. It is worth noting that urban land or areas are polluted often, but the attention paid by soil scientists to urban
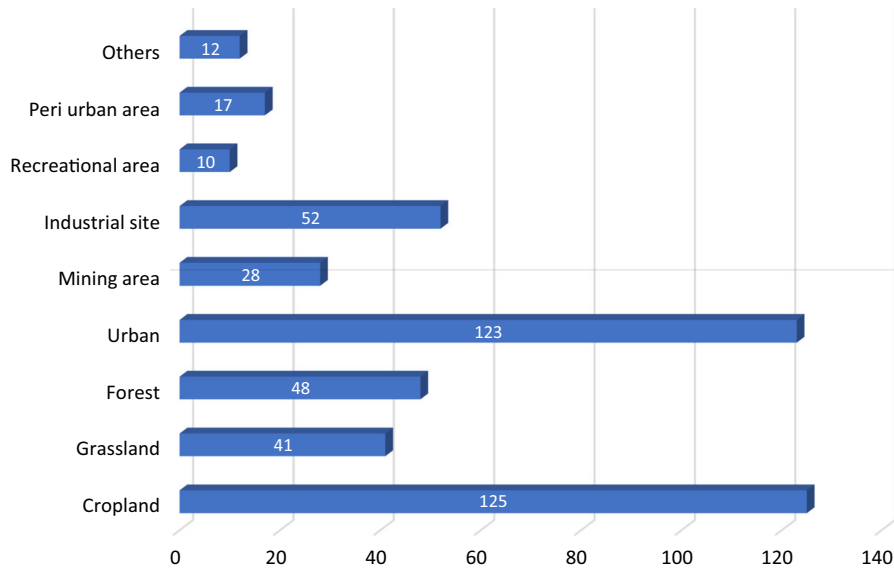
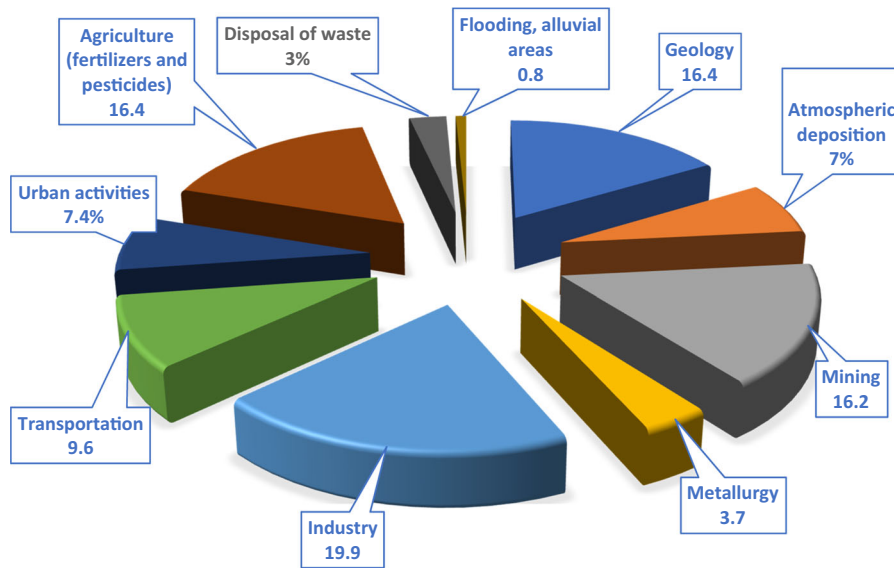**Fig. 4** Proportion of land-use categories studied in the papers

Fig. 5 Proportion of pollution sources studied in the reviewed papers

land is minimal. On the contrary, due to the attention being paid to these sectors, it is relatively easy to find substantial research and reportage on cropland or agricultural land. Urban areas are more heterogeneous, fragmented and not used for food production, which is why less attention is being paid to it.

Figure 5 displays the various sources for which the soil has been polluted. From the papers reviewed, 102 articles referred to the industry as a significant soil polluter, accounting for 19.9% of the pie chart. Other sources of pollution being analysed are agricultural (84), geology (84), mining (83), transport (49), construction activities (38), atmospheric deposition (36), metallurgy (19), waste disposal (14) and flooding (4). In some land-use areas, the sources of pollution used have been a multiplicity of sources, and industrial pollution (19.9%) is very high in soil pollution. These numbers indicate the relative attention paid to the

sources in the reviewed articles, which may reflect on the importance of pollution; however, they may also be influenced by the easiness of mapping.

According to Bundschuh et al. (2012), EEA (2014) and FAO and ITPS (2015) anthropogenic activities such as waste disposal, mining, industrial activities, transport, agriculture (fertilizers and pesticides) and other hosts of anthropogenic activities have been a significant problem of global soil pollution for centuries. FAO and ITPS (2015) argued that mining had had a dominant impact on soil, water and biota since prehistoric times. Alloway (2013) reported earlier before FAO and ITPS (2015) communicated that there is a documented evidence all over the world, highlighting many instances of heavy soil contamination by mining activities. Substantial quantities of PTEs and other related elements are released into the environment by mining and smelting facilities, which have continued for an extended period even after these activities (Ogundele et al. 2017).

Armah et al. (2014) catalogued that different PTEs were found in higher levels around urban areas, peri-urban areas, hinterland, foodstuffs, soil and water bodies around a close range of mining areas in Ghana. According to the report by FAO and ITPS (2018), the development of infrastructure in the form of housing, roads and railways has contributed primarily to environmental degradation. Furthermore, the situation has shown that the negative impact on the soil is the sealing of the soil and land consumption. Apart from these imminent threats, infrastructural soil pollution has received a great deal of attention for a long time in terms of impact assessment and planning (FAO and ITPS 2018). Moving humans from one point to the other using the transport system has also been one of the focal points for soil pollution in a variety of ways, such as emissions and combustion of engines that reach more than 100 m of soil through atmospheric deposition as well as oil spills (Mirsal 2008). Transportation systems using rail, road and highway are some of the significant soil pollutants, particularly in the peri-urban, urban and hinterland regions. On the other hand, paints rich in Pb are a significant legacy of Pb contaminant in urban and peri-urban areas, apart from fuel knocking. According to Mielke and Reagan (1998), the soil is contaminated with paint rich in Pb when it is pulverized into smaller particles or dust during the demolition or renovation process that percolates into the environment. Mielke and Reagan

(1998) further pointed out that approximately the exact tonnage of Pb used in leaded gasoline between 1929 and 1989 was used in white paint pigments between 1884 and 1989 with the maximum use of paints rich in Pb in the year 1920 to 1929. Sometimes contaminants from landfill sites (landfill leachates) also pollute the soil and groundwater. It is, therefore, incumbent to deduce based on the papers reviewed that attention is being paid to anthropogenic pollution; nevertheless, the human is at the centre of these pollutions. In particular, land use related to PTEs (study areas) were studied because most of the authors placed a premium on the critical role arable land plays in food production. Zhang et al. (2014), Tóth et al. (2016) and Adagunodo et al. (2018) confirmed that the most research paper focuses on arable land and the investigation of PTEs, as mentioned earlier.

Studied potentially toxic elements

Various PTEs above the normal threshold have the propensity to contribute to the soil's insalubrious state. A large number of articles focused on PTEs such as lead (Pb), chromium (Cr), arsenic (As), zinc (Zn), cadmium (Cd), copper (Cu), mercury (Hg), molybdenum (Mo), cobalt (Co), vanadium (V) manganese (Mn) and nickel (Ni). Other elements such as thorium (Th) and antimony (Sb) were also mentioned in some of the articles. But the elements mentioned above were mostly those that were very popular in the articles. Figure 6 shows clearly that, from 2001 to 2018, As, Cd, Cr, Cu, Pb and Zn were found to be predominant in most of the soils examined by the scientists. During 2018, the PTE survey captured in Fig. 6 showed a progressive increase in studies of all elements except vanadium. It was evident that from 2010 to 2014, ongoing studies showed the presence of As, Cd, Cr, Co, Cu, Ni, Pb, Zn, V and Mn in soil annually. However, a large number of As, Cd, Cu, Ni, Pb, Zn and Cr assessments were conducted in the soil during 2011, 2013 and 2014. In 2015 to 2018, the number of articles publishing the presence of As, Cd, Cr, Co, Cu, Ni, Pb, Zn, V and Mn in soil increased. During 2015, 2016 and 2018, there has been a steady increase in the number of publications, indicating the presence of most of the elements captured in this paper. During 2019, we reviewed papers from the first quarter only, and Fig. 8 clearly shows that there has been a steady influx of PTE publications. The data obtained showed
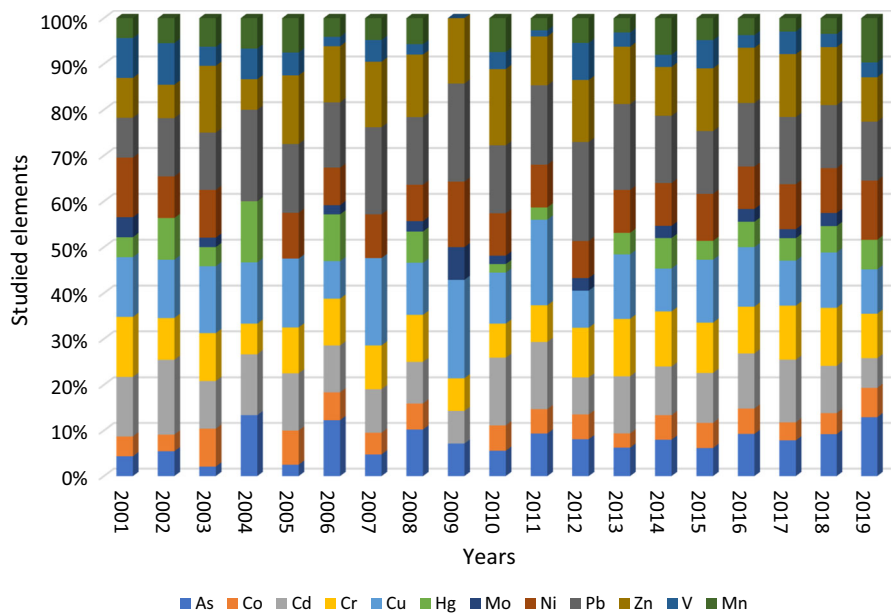
Fig. 6 Proportion of papers focused on DSM of particular PTEs

that many papers mentioned the study of Pb (181 times), Cu (153 times), Zn (153 times) and Cd (141 times). Other PTEs such as Cr, Ni, As, Co, Hg, Mn, V and Mo have also been mentioned several times in various papers accounting for the following figures: 130, 116, 95, 62, 61, 58, 49 and 20, respectively. Much of the study centred on Pb because of the danger it presents to society and people in general. In October every year, the World Health Organization (WHO) spends a week on raising awareness and reducing Pb poisoning to the barest minimum. In 2019, the WHO objectives for the International Pb Poisoning Prevention Week of Action set out the following objectives: to raise awareness of the health effects of lead poisoning; to highlight the efforts of countries and partners to prevent childhood lead poisoning; and to urge further action to eliminate lead paint through country-level regulatory action (WHO 2019). Member states have adopted the objectives by ensuring that there is a reduction in the fight against Pb contamination in soil and the environment. Studies conducted in the Mediterranean (Spain) by Iñigo et al. (2014) and Marín et al. (2016) have shown that human activities, such as soil management practices, increase the concentration of PTEs (such as Pb and Cd) in soil. A great deal of attention has been given to them (mainly Pb) and hence necessitated soil scientists given it the needed attention to curtail its release into the soil and

the environment. The need for these elements from the earth crust for our industry remains essential, as there may be no alternative means of extraction or processing of these earth metals. Mining is known to generate direct and indirect employment as well as meet the needs of industries, but considering the impact it has had, soil scientists, policymakers and other stakeholders need to find a more effective and environmentally sustainable way to mine metals.

Objectives for using DSM models

There are different motives behind the modelling of a piece of land, area or settings. Most DSM users have a variety of reasons for using a model or algorithm, and their intention aligns with the creation of a geographical soil database at a specified resolution with the help of a laboratory and field observation method (either one or both). This objective is achieved by means of environmental data obtained quantitatively and through a quantifiable relationship.

Figure 7 shows various reasons why soil scientists and experts in the field of pedometrics or soil science have applied or used a model to provide information on the state of health of the soil/sediment out there to policymakers or end-users. The authors of the reviewed papers indicated that they applied a model or algorithm for various reasons, and based on the

Fig. 7 Objectives of the authors to use the DSM algorithm
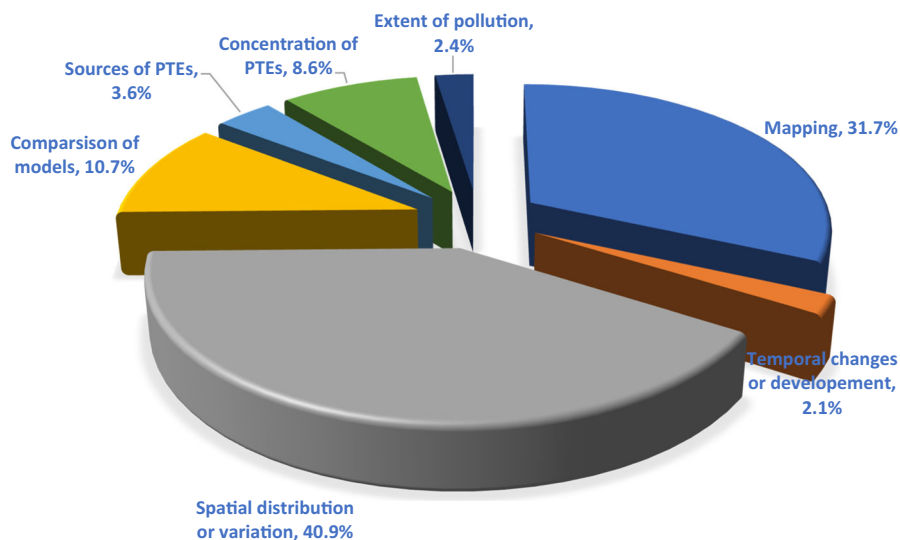
information accrued, it was apparent that their motives aligned with the choice of the model applied were justifiable since it provided accurate results.

Among the most frequent reasons, some of the authors referred to mapping as a reason for modelling; soil mapping was cited 133 times in the articles (representing 31.7% in the pie chart). On the other hand, some of the reasons were for determining the extent of the pollution and 10 papers were collected for this purpose; 172 articles indicated spatial distribution or variation as a reason for using the algorithms. Comparison of the models as an objective accounted for 45 articles, indicating sources of PTEs for 15 papers, concentration of PTEs for 36 papers, and finally, the least mentioned reason was temporal changes or developments for 9 articles. In some of the papers, more than one objective was mentioned as a reason for the use of the algorithm. The primary motivation of the DSM is to emphasize the modelling of the landscape in which it quantifies the relation between the environmental variable and the soil properties (Scull et al. 2003). The papers reviewed brought together a collection of predictive models or algorithms in the field of soil mapping, tailored towards the prediction of PTEs, using either geostatistical models, machine learning or a combination of both models. Many of these models identified in this review are documented in a paper published by McBratney et al. (2003) underlining the application process, the underpinnings and principles of DSM or soil science by different authors. These DSM models

in this paper provided useful information and practical experience from other published articles, highlighting the reasons as well as some limitations encountered in their research. Many of these publications allow new users, as well as old users, who want to use a different algorithm, to have access to information and sometimes serve as a guideline.

Uncertainty and validation

The quality of soil map can be related to measures of accuracy (Finke 2006). According to Cressie and Kornak (2003), if there is an error sampled, there is a
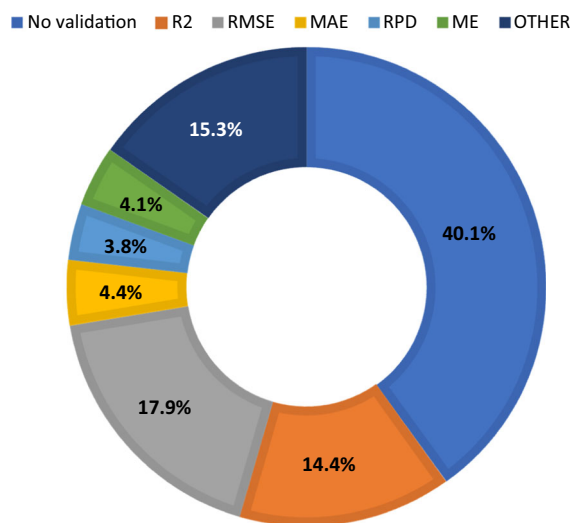

Fig. 8 Percentage of papers using various validation criteria

likelihood that it will contribute to the relationship inferred by the fact that covariates and soil observation will not actually correspond. Based on such error, estimating the uncertainty that will occur when making maps is acceptable. One of the essences of research is to keep people informed and disseminate information to people, end-users and policymakers to enhance the well-being of humans in society or the world. Scientist or researcher ensures that the information provided is well checked, validated and of higher quality. In this context, there is a certain level of uncertainty in DSM due to unknown information, such as specific uncertainties stemming from the prediction of soil properties generated by the DSM, which is essential for the use of soil function assessments. Based on this presumption, it is incumbent upon the authors to use a quantifiable level of uncertainty and to make a comparison to assess the performance of the method. As a result, validation provides an orifice to authenticate and address these issues. Several validation criteria have been presented in the articles reviewed. Some of the maps created by the authors in the articles went on further to validate it, while others (40.1%) did not validate it. Figure 8 shows the proportion of validation methods used to address the issue of uncertainty and outlines some articles that did not use any validation methods. No validation was carried out on 128 articles reviewed, 46 papers used R square (R2) validation method, 57 root mean square error (RMSE), 14 mean absolute error (MAE), 12 residual prediction deviation (RPD), 13 mean error (ME) and the rest of the validation criteria put together 45 papers. In some of the articles, a multiplicity of validation indicators was used. In the year 2001 to 2006, many published articles measured the quality of maps as well as proposed different ways of classifying, comparing and validating soft maps (Woodcock and Gopal 2000; Pontius and Cheuk 2006). Grunwald (2009) argued that conventional and digital soil maps are not perfect and contain some errors, and therefore, it is crucial that they are frequently validated. Congalton (1991) captured the most critical measurements of map quality, outlining that the quality map should include the following: overall accuracy (overall purity, map purity), user accuracy, producer accuracy, and kappa coefficient of agreement. In another vein, Lark (1995) raised questions about the appropriateness of the term's user and producer accuracy, as well as the incorporation of quality

measures, which may be crucial for users as well as producers. In recent times, the use of deep learning for DSM has been shown to be useful and produces a higher level of accuracy in the spatial prediction model. Recent research conducted by Behrens et al. (2018) has shown that the use of intermediate scales in DSM has been found to produce high predictive accuracy and, as such, serve as a multipurpose approach to spatial prediction and modelling of any size of area at any resolution. However, it is imperative to point out that the uncertainty associated with prediction and map creation should be independently validated to address mapping issues in other areas to improve the quality of the map.

Conclusions

This study quantitatively reviewed DSM models that are used spatially to predict PTEs in soil and sediment. Europe, Asia, North America and Australia and New Zealand have always been hot spot regions leading to research and publication on PTEs, as well as providing end-user materials and policy for several activities. Research on PTEs in Europe using the DSM model is relatively prevalent. The difficulty is that it is sometimes challenging to access articles, especially in the USA and Australia, despite their enormous contribution to the DSM. In Africa, some of the articles collected were collaborations with other institutions beyond the shores of Africa. However, the African continent has become a hub for all kinds of exploration and mining activities. However, the use of DSM in developing countries is relatively low; therefore, we advocate a decentralized approach for the soil science community to ensure that DSM reaches globally and is user-friendly. The soil science community should animate clubs in different universities to adapt and learn modern mapping methods. Moreover, subregional integration should inform and update scientists on emerging trends in disadvantaged parts of the world, as well as support for the uptake of DSM tools. Our research confirms that kriging (simple, ordinary, universal) approaches remain the most preferred algorithms. Nonetheless, the intervention of machine learning algorithms has resulted in resilience and versatility in the use of models to predict PTEs spatially. There should be further research to deepen the complementing role knitted between geostatistical

approach and machine learning algorithm to enhance the finest hybrid model to decrease the level of uncertainty clothe with both algorithms. Geoderma gave us more of the articles than any other journals captured in this study. It was evident that many researchers were focusing on cropland, but again it was also essential that a great deal of attention be given to other land uses. Besides, the increasing level of publicity and the campaign against anthropogenic activities that release PTEs into the soil, more significant concerns should be placed on the developing countries, particularly in Africa, Southern America and some parts of Asia. The root mean square error (RMSE) criterion was the most preferred validation criterion used by the authors. Several papers worked on distinct PTEs such as Pb, Cr, As, Zn, Cd, Cu, Hg, Mo, Co, V, Mn and Ni. Among the PTEs, Pb was the most mentioned element.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict interest.

References

Adagunodo, T. A., Sunmonu, L. A., & Emetere, M. E. (2018). Heavy metals' data in soils for agricultural activities. *Data in Brief, 18,* 1847–1855.

Alloway, B. (2013). *Heavy metals in soils: trace metals and metalloids in soils and their bioavailability*. Berlin: Springer.

Appelhans, T., Mwangomo, E., Hardy, D. R., Hemp, A., & Nauss, T. (2015). Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro Tanzania. *Spatial Statistics, 14,* 91–113.

Armah, F. A., Quansah, R., & Luginaah, I. (2014). A systematic review of heavy metals of anthropogenic origin in environmental media and biota in the context of gold mining in Ghana. *International Scholarly Research Notices, 2014,* 1–37.

Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B. M., Hong, S. Y., et al. (2014). GlobalSoilMap. Toward a Fine-Resolution Global Grid of Soil Properties. In *Advances in Agronomy* (Vol. 125, pp. 93–134).

Basta, N. T., Ryan, J. A., & Chaney, R. L. (2005). Trace element chemistry in residual-treated soil: key concepts and metal bioavailability. *Journal of Environmental Quality, 34*(1), 49–63.

Beguin, J., Fuglstad, G. A., Mansuy, N., & Paré, D. (2017). Predicting soil properties in the Canadian boreal forest with limited data: Comparison of spatial and non-spatial statistical approaches. *Geoderma, 306,* 195–205.

Behrens, T., Schmidt, K., MacMillan, R. A., & Viscarra Rossel, R. A. (2018). Multi-scale digital soil mapping with deep learning. *Scientific Reports*. https://doi.org/10.1038/s41598-018-33516-6

Bundschuh, J., Litter, M. I., Parvez, F., Román-Ross, G., Nicolli, H. B., Jean, J. S., et al. (2012). One century of arsenic exposure in Latin America: A review of history and occurrence from 14 countries. *Science of the Total Environment, 429,* 2–35.

Cachada, A., Rocha-Santos, T. A. P., & Duarte, A. C. (2017). Soil and pollution: An introduction to the main issues. *Soil Pollution: From Monitoring to Remediation*. https://doi.org/10.1016/B978-0-12-849873-6.00001-7

Chen, T., Liu, X., Li, X., Zhao, K., Zhang, J., Xu, J., et al. (2009). Heavy metal sources identification and sampling uncertainty analysis in a field-scale vegetable soil of Hangzhou China. *Environmental Pollution, 157*(3), 1003–1010.

Chen, H., Teng, Y., Lu, S., Wang, Y., & Wang, J. (2015). Contamination features and health risk of soil heavy metals in China. *Science of the Total Environment, 512–513,* 143–153.

Chen, H., Zheng, C., Tu, C., & Zhu, Y. G. (1999). Heavy metal pollution in soils in China: Status and countermeasures. *Ambio, 28*(2), 130–134.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*(1), 35–46.

Cressie, N., & Kornak, J. (2003). Spatial statistics in the presence of location error with an application to remote sensing of the environment. *Statistical Science, 18*(4), 436–456.

EEA. (2014). Progress in Management of Contaminated Sites (CSI 015/LSI 003).

FAO and ITPS. (2015). Status of the World's Soil Resources. *Intergovernmental Technical Panel on Soils*, 123–126.

FAO and ITPS. (2018). Status of the World's Soil Resources. *Intergovernmental Technical Panel on Soils*, 123–126.

Finke, P. A. (2006). Chapter 39 quality Assessment of Digital Soil Maps: Producers and Users Perspectives. *Developments in Soil Science*. https://doi.org/10.1016/S0166-2481(06)31039-2

GAO, U. (2015). United States Government Accountability Office. (2015). Trends in Federal Funding and Cleanup of EPA's Nonfederal National Priorities List Sites (GAO Publication No. GAO-15-812). Washington, DC: U.S. Government Printing Office.

Gasch, C. K., Hengl, T., Gräler, B., Meyer, H., Magney, T. S., & Brown, D. J. (2015). Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. *Spatial Statistics, 14,* 70–90.

Gonzalez, C., Gonzalez-Macias, G., Macias, M., Schifter, I., Lluch-Cota, D. B., Endez-Rodriguez, M., et al. (2006).

Distribution, enrichment and accumulation of heavy metals in coastal sediments of salina cruz bay, m´exico. *Environmental Monitoring and Assessment, 118*(1–3), 211–230.

Grunwald, S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma, 152*(3–4), 195–207.

Gunson, A. J., Gunson, A. J., & Jian, Y. (2001). Artisanal Mining in The People's Republic of China Communities and Small-scale of Mines (CASM) under the World Bank View project Artisanal Mining in The People's Republic of China. researchgate.net.

Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma, 265,* 62–77.

ISI. (2020). Developing Countries - ISI. https://www.isi-web.org/index.php/capacity-building/developing-countries.

Iqbal, J., Thomasson, J. A., Jenkins, J. N., Owens, P. R., & Whisler, F. D. (2005). Spatial variability analysis of soil physical properties of alluvial soils. *Soil Science Society, 69*(4), 1338–1350.

Iñigo, V., Andrades, M., Alonso-Martirena, J. I., Marín, A., & Jiménez-Ballesta, R. (2011). Multivariate statistical and GIS-based approach for the identification of Mn and Ni concentrations and spatial variability in soils of a humid mediterranean environment: La Rioja, Spain. *Water, Air, and Soil Pollution, 222*(1–4), 271–284.

Iñigo, V., Andrades, M. S., Alonso-Martirena, J. I., Marín, A., & Jiménez-Ballesta, R. (2014). Background values and distribution trends of Cu and Zn in soils of humid Mediterranean environment. *Chemistry and Ecology, 30*(3), 252–266.

Jiang, Z. (2018). A survey on spatial prediction methods. *IEEE Transactions on Knowledge and Data Engineering., 31,* 164.

Jiménez-Ballesta, R., García-Navarro, F. J., Bravo, S., Amorós, J. A., Pérez-de-los-Reyes, C., & Mejías, M. (2017). Environmental assessment of potential toxic trace element contents in the inundated floodplain area of Tablas de Daimiel wetland (Spain). *Environmental Geochemistry and Health, 39*(5), 1159–1177.

Juel, A., Groom, G. B., Svenning, J. C., & Ejrnæs, R. (2015). Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. *International Journal of Applied Earth Observation and Geoinformation, 42,* 106–114.

Kaasalainen, M., & Yli-Halla, M. (2003). Use of sequential extraction to assess metal partitioning in soils. *Environmental Pollution, 126*(2), 225–233.

Kabata-Pendias, A., & Mukherjee, A. B. (2007). Trace elements from soil to human. *Trace Elements from Soil to Human.* https://doi.org/10.1007/978-3-540-32714-1

Khan, S., Cao, Q., Zheng, Y. M., Huang, Y. Z., & Zhu, Y. G. (2008). Health risks of heavy metals in contaminated soils and food crops irrigated with wastewater in Beijing China. *Environmental Pollution, 152*(3), 686–692.

Kuo, S., Heilman, P. E., & Baker, A. S. (1983). Distribution and forms of copper, zinc, cadmium, iron, and manganese in soils near a copper smelter. *Soil Science., 135,* 101.

Lagacherie, P., & McBratney, A. B. (2006). Chapter 1 spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping. *Developments in Soil Science.* https://doi.org/10.1016/S0166-2481(06)31001-X

Lark, R. M. (1995). A reappraisal of unsupervised classification, II: Optimal adjustment of the map legend and a neighbourhood approach for mapping legend units. *International Journal of Remote Sensing, 16*(8), 1445–1460.

Lasat, M. M. (1999). Phytoextraction of metals from contaminated soil: A review of plant/soil/metal interaction and assessment of pertinent agronomic issues. *Journal of Hazardous Substance Research, 2*(1), 1–1.

Liang, J., Feng, C., Zeng, G., Gao, X., Zhong, M., Li, X., et al. (2017). Spatial distribution and source identification of heavy metals in surface soils in a typical coal mine city, Lianyuan, China. *Environmental Pollution, 225,* 681–690.

Liu, W. X., Li, X. D., Shen, Z. G., Wang, D. C., Wai, O. W. H., & Li, Y. S. (2003). Multivariate statistical study of heavy metal enrichment in sediments of the Pearl River Estuary. *Environmental Pollution, 121*(3), 377–388.

Luo, L., Ma, Y., Zhang, S., Wei, D., & Zhu, Y. G. (2009a). An inventory of trace element inputs to agricultural soils in China. *Journal of Environmental Management, 90*(8), 2524–2530.

Luo, Y., Wu, L., Liu, L., Han, C., & Li, Z. (2009b). *Heavy metal contamination and remediation in Asian agricultural land. National Institutes for Agro-Environmental Sciences* (Vol. 1).

Marín, A., Andrades, M., Iñigo, V., & Jiménez-Ballesta, R. (2016). Lead and Cadmium in Soils of La Rioja Vineyards Spain. *Land Degradation and Development, 27*(4), 1286–1294.

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*(1–2), 3–52.

Mielke, H. W., & Reagan, P. L. (1998). Soil is an important pathway of human lead exposure. *Environmental Health Perspectives., 106,* 217.

Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma, 264,* 301–311.

Mirsal, I. (2008). *Soil pollution: Origin monitoring & remediation* (p. 310). Berlin: Springer.

Nicholson, F. A., Smith, S. R., Alloway, B. J., Carlton-Smith, C., & Chambers, B. J. (2003). An inventory of heavy metals inputs to agricultural soils in England and Wales. *Science of the Total Environment, 311*(1–3), 205–219.

Ogundele, L. T., Owoade, O. K., Hopke, P. K., & Olise, F. S. (2017). Heavy metals in industrially emitted particulate matter in Ile-Ife, Nigeria. *Environmental Research, 156,* 320–325.

Oldeman, L. (1991). *World map on status of human - induced soil degradation.* Wageningen: Ref.

Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning for digital soil mapping. *Soil, 5*(1), 79–89.

Pontius, R. G., & Cheuk, M. L. (2006). A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions. *International Journal of Geographical Information Science, 20*(1), 1–30.

Rhee, J., & Im, J. (2017). Meteorological drought forecasting for ungauged areas based on machine learning: Using long-

range climate forecast and remote sensing data. *Agricultural and Forest Meteorology, 237–238,* 105–122.

Rossiter, D. G. (2018). Past, present & future of information technology in pedometrics. *Geoderma., 324,* 131.

Scragg, A. (2006). *Environmental Biotechnology, Oxford University Press, Oxford, UK, 2nd edition, 2006.* Oxford: Oxford University Press.

Scull, P., Franklin, J., Chadwick, O. A., & McArthur, D. (2003). Predictive soil mapping: A review. *Progress in Physical Geography, 27*(2), 171–197.

Seaward, M. R. D, Richardson, D. H. S. (1990). Atmospheric sources of metal pollution and effects on vegetation. *Heavy metal tolerance in plants evolutionary aspects,* 75–92.

Srivastava, P. K., Islam, T., Gupta, M., Petropoulos, G., & Dai, Q. (2015). WRF dynamical downscaling and bias correction schemes for NCEP estimated hydro-meteorological variables. *Water Resources Management, 29*(7), 2267–2284.

Swartjes, F. A., & Siciliano, S. (2012). Dealing with contaminated sites: from theory towards practical application. *Soil Science Society of America Journal, 76*(2), 748–748.

Thevenon, F., Guédron, S., Chiaradia, M., Loizeau, J. L., & Poté, J. (2011). (Pre-) historic changes in natural and anthropogenic heavy metals deposition inferred from two contrasting Swiss Alpine lakes. *Quaternary Science Reviews, 30*(1–2), 224–233.

Tóth, G., Hermann, T., Szatmári, G., & Pásztor, L. (2016). Maps of heavy metals in the soils of the European Union and proposed priority areas for detailed assessment. *Science of the Total Environment, 565,* 1054–1062.

van Zijl, G. (2019). Digital soil mapping approaches to address real world problems in southern Africa. *Geoderma, 337,* 1301–1308.

Veronesi, F., & Schillaci, C. (2019). Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. *Ecological Indicators, 101,* 1032–1044.

WHO (2019). International Lead Poisoning Prevention Week of Action (October 2015): Examples of planned activities by governments, non-governmental organizations and others.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists, 2nd Edition (Statistics in Practice).* Chichester: Wiley.

Woodcock, C. E., & Gopal, S. (2000). Fuzzy set theory and thematic maps: Accuracy assessment and area estimation. *International Journal of Geographical Information Science, 14*(2), 153–172.

Wuana, R. A., & Okieimen, F. E. (2014). Heavy metals in contaminated soils: A review of sources, chemistry, risks, and best available strategies for remediation. *Heavy Metal Contamination of Water and Soil Analysis, Assessment, and Remediation Strategies.* https://doi.org/10.1201/b16566-3

Yang, Q., Li, Z., Lu, X., Duan, Q., Huang, L., & Bi, J. (2018). A review of soil heavy metal pollution from industrial and agricultural regions in China: Pollution and risk assessment. *Science of the Total Environment., 642,* 690.

Zhang, G. L., Liu, F., & Song, X. D. (2017). Recent progress and future prospect of digital soil mapping: A review. *Journal of Integrative Agriculture., 16,* 2871.

Zhang, M. K., Liu, Z. Y., & Wang, H. (2010). Use of single extraction methods to predict bioavailability of heavy metals in polluted soils to rice. *Communications in Soil Science and Plant Analysis, 41*(7), 820–831.

Zhang, X., Zhang, X., & Zhong, T. (2014). Spatial distribution and accumulation of heavy metal in arable land soil of China. *Huan jing ke xue, 35,* 692.

Zhu, A. X., Hudson, B., Burt, J., Lubich, K., & Simonson, D. (2001). Soil Mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal, 65*(5), 1463–1472.

Zhu, A.-X., Lu, G., Liu, J., Qin, C.-Z., & Zhou, C. (2018). Spatial prediction based on Third Law of Geography. *Annals of GIS, 24*(4), 225–240.

List of reference used for the bibliometric review

Maas, S., Scheifler, R., Benslama, M., Crini, N., Lucot, E., Brahmia, Z., & Giraudoux, P. (2010). Spatial distribution of heavy metal concentrations in urban, suburban and agricultural soils in a Mediterranean city of Algeria. *Environmental Pollution, 158*(6), 2294–2301.

Manta, D. S., Angelone, M., Bellanca, A., Neri, R., & Sprovieri, M. (2002). Heavy metals in urban soils: a case study from the city of Palermo (Sicily). *Italy. Science of the Total Environment, 300*(1–3), 229–243.

Al Maliki, A., Bruce, D., & Owens, G. (2015). Spatial distribution of Pb in urban soil from Port Pirie, South Australia. *Environmental Technology & Innovation, 4,* 123–136.

Birch, G. F., Vanderhayden, M., & Olmos, M. (2011). The nature and distribution of metals in the Sydney estuary catchment, Australia. *Water, Air, & Soil Pollution, 216*(1–4), 581–604.

Johnson, L. E., Bishop, T. F. A., & Birch, G. F. (2017). Modeling drivers and distribution of lead and zinc concentrations in soils of an urban catchment (Sydney estuary, Australia). *Science of The Total Environment, 598,* 168–178.

Robinson, T. P., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and electronics in agriculture, 50*(2), 97–108.

Shamsoddini, A., Raval, S., & Taplin, R. (2014). SPECTROSCOPIC ANALYSIS OF SOIL METAL CONTAMINATION AROUND A DERELICT MINE SITE IN THE BLUE MOUNTAINS, AUSTRALIA. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 2(7).

Verstraeten, G., Prosser, I. P., & Fogarty, P. (2007). Predicting the spatial patterns of hillslope sediment delivery to river channels in the Murrumbidgee catchment. *Australia. Journal of Hydrology, 334*(3–4), 440–454.

Wilford, J., de Caritat, P., & Bui, E. (2016). Predictive geochemical mapping using environmental correlation. *Applied geochemistry, 66,* 275–288.

Delbari, M., Afrasiab, P., & Loiskandl, W. (2009). Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. *CATENA, 79*(2), 163–169.

Van Meirvenne, M., & Goovaerts, P. (2001). Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. *Geoderma, 102*(1–2), 75–100.

Burak, D. L., Fontes, M. P., Santos, N. T., Monteiro, L. V. S., de Sousa Martins, E., & Becquer, T. (2010). Geochemistry and spatial distribution of heavy metals in Oxisols in a mineralized region of the Brazilian Central Plateau. *Geoderma, 160*(2), 131–142.

Camargo, L. A., Marques, J., Jr., Barrón, V., Alleoni, L. R. F., Pereira, G. T., Teixeira, D. D. B., & de Souza Bahia, A. S. R. (2018). Predicting potentially toxic elements in tropical soils from iron oxides, magnetic susceptibility and diffuse reflectance spectra. *CATENA, 165,* 503–515.

De Oliveira, M. T. G., Rolim, S. B. A., de Mello-Farias, P. C., Meneguzzi, Á., & Lutckmeier, C. (2008). Industrial pollution of environmental compartments in the Sinos River Valley, RS, Brazil: geochemical–biogeochemical characterization and remote sensing. *Water, air, and soil pollution, 192*(1–4), 183–198.

Lourenço, R. W., Landim, P. M. B., Rosa, A. H., Roveda, J. A. F., Martins, A. C. G., & Fraceto, L. F. (2010). Mapping soil pollution by spatial analysis and fuzzy classification. *Environmental earth science, 60*(3), 495–504.

Bower, J. A., Lister, S., Hazebrouck, G., & Perdrial, N. (2017). Geospatial evaluation of lead bioaccessibility and distribution for site specific prediction of threshold limits. *Environmental Pollution, 229,* 290–299.

Forsythe, K. W., Dennis, M., & Marvin, C. H. (2004). Comparison of Mercury and Lead Sediment Concentrations in Lake Ontario (1968–1998) and Lake Erie (1971–1997/98) using a GIS-Based Kriging Approach. *Water Quality Research Journal, 39*(3), 190–206.

Oyarzun, R., Oyarzún, J., Lillo, J., Maturana, H., & Higueras, P. (2007). Mineral deposits and Cu–Zn–As dispersion–contamination in stream sediments from the semiarid Coquimbo Region. *Chile. Environmental Geology, 53*(2), 283–294.

Cai, L., Xu, Z., Bao, P., He, M., Dou, L., Chen, L., & Zhu, Y. G. (2015). Multivariate and geostatistical analyses of the spatial distribution and source of arsenic and heavy metals in the agricultural soils in Shunde, Southeast China. *Journal of Geochemical Exploration, 148,* 189–195.

Cao, S., Lu, A., Wang, J., & Huo, L. (2017). Modeling and mapping of cadmium in soils based on qualitative and quantitative auxiliary variables in a cadmium contaminated area. *Science of the Total Environment, 580,* 430–439.

Ding, Q., Wang, Y., & Zhuang, D. (2018). Comparison of common spatial interpolation methods used to analyze potentially toxic elements surrounding mining regions. *Journal of Environmental Management, 212,* 23–31.

Dong, J., Yu, M., Bian, Z., Wang, Y., & Di, C. (2011). Geostatistical analyses of heavy metal distribution in reclaimed mine land in Xuzhou. *China. Environmental Earth Sciences, 62*(1), 127–137.

Duan, X., Zhang, G., Rong, L., Fang, H., He, D., & Feng, D. (2015). Spatial distribution and environmental factors of catchment-scale soil heavy metal contamination in the dry-hot valley of Upper Red River in southwestern China. *CATENA, 135,* 59–69.

Gu, Y. G., Li, Q. S., Fang, J. H., He, B. Y., Fu, H. B., & Tong, Z. J. (2014). Identification of heavy metal sources in the reclaimed farmland of the pearl river estuary in China using a multivariate geostatistical approach. *Ecotoxicology and environmental safety, 105,* 7–12.

Guagliardi, I., Buttafuoco, G., Cicchella, D., & De Rosa, R. (2013). A multivariate approach for anomaly separation of potentially toxic trace elements in urban and peri-urban soils: an application in a southern Italy area. *Journal of soils and sediments, 13*(1), 117–128.

Husnizar, H., Wilopo, W., & amp; Yuliansyah, AT, . (2018). The prediction of heavy metals lead (Pb) and zinc (Zn) contents in soil using NIRs technology and PLSR regression method. *Journal of Degraded and Mining Lands Management, 5*(3), 1153.

Jin, Y., O'Connor, D., Ok, Y. S., Tsang, D. C., Liu, A., & Hou, D. (2019). Assessment of sources of heavy metals in soil and dust at children's playgrounds in Beijing using GIS and multivariate statistical analysis. *Environment international, 124,* 320–328.

Johnbull, O., Abbassi, B., & Zytner, R. G. (2018). Risk assessment of heavy metals in soil based on the geographic information system-Kriging technique in Anka. *Nigeria. Environmental Engineering Research, 24*(1), 150–158.

Li, D. A., Jiang, J., Li, T., & amp; Wang, J. . (2016). Soil heavy metal contamination related to roasted stone coal slag: study based on geostatistical and multivariate analyzes. *Environmental Science and Pollution Research, 23*(14), 14405–14413.

Li, F., Huang, J., Zeng, G., Huang, X., Li, X., Liang, J., & amp; Bai, B. . (2014). Integrated source apportionment, risk risk mapping, and heavy metals in surface sediments: a study of Dongting Lake, Middle China. *Human and Ecological Risk Assessment: An International Journal, 20*(5), 1213–1230.

Li, P., Zhi, Y., Shi, J., Zeng, L., & Wu, L. (2015). County-scale temporal–spatial distribution and variability tendency of heavy metals in arable soils influenced by policy adjustment during the last decade: a case study of Changxing. *China. Environmental Science and Pollution Research, 22*(22), 17937–17947.

Li, X., Wu, T., Bao, H., Liu, X., Xu, C., Zhao, Y., & amp; Yu, H. . (2017). Potential toxic element (PTE) contamination in Baoji urban soil (NW China): spatial distribution, mobility behavior, and health risk. *Environmental Science and Pollution Research, 24*(24), 19749–19766.

Liang, J., Feng, C., Zeng, G., Gao, X., Zhong, M., Li, X., & Fang, Y. (2017). Spatial distribution and source identification of heavy metals in surface soils in a typical coal mine city, Lianyuan, China. *Environmental Pollution, 225,* 681–690.

Lin, Y. P., Chang, T. K., Shih, C. W., & Tseng, C. H. (2002). Factorial and indicator kriging methods using a geographic information system to delineate spatial variation and pollution sources of soil heavy metals. *Environmental Geology, 42*(8), 900–909.

Liu, H., Liu, G., Zhou, Y., & He, C. (2017). Spatial Distribution and Influence Analysis of Soil Heavy Metals in Hilly Region of Sichuan Basin. *Polish Journal of Environmental Studies , 26* (3).

Liu, M., Liu, X., Wu, M., Li, L., & Xiu, L. (2011). Integrating spectral indices with environmental parameters for estimating heavy metal concentration in rice using dynamic

fuzzy neural-network model. *Computers & Geosciences, 37*(10), 1642–1652.

Liu, M., Yang, Y., Yun, X., Zhang, M., & Wang, J. (2015). Concentrations, distribution, sources, and ecological risk assessment of heavy metals in the topsoil of the Three Gorges Dam region. *China. Environmental Monitoring and Assessment, 187*(3), 147.

Liu, X., Wu, J., & Xu, J. (2006). Characterizing the risk assessment of heavy metals and sampling uncertainty analysis in paddy field by geostatistics and GIS. *Environmental pollution, 141*(2), 257–264.

Liu, Y., Ma, Z., Lv, J., & Bi, J. (2016). Identifying sources and hazardous risks of heavy metals in East China. *Journal of Geographical Sciences, 26*(6), 735–749.

Lu, J., Jiao, W. B., Qiu, H. Y., Chen, B., Huang, X. X., & Kang, B. (2018). Origin and spatial distribution of heavy metals and carcinogenic risk assessment in mining areas at You'xi County southeast China. *Geoderma, 310,* 99–106.

Lv, J. (2019). Multivariate receptor models and robust geostatistics are the source of heavy metal and soils. *Environmental Pollution, 244,* 72–83.

Lv, J., Liu, Y., Zhang, Z., Dai, J., Dai, B., & Zhu, Y. (2015). Identifying the origins and spatial distributions of heavy metals in Eastern China using a multivariate and geostatistical approach. *Journal of soils and sediments, 15*(1), 163–178.

Lv, J., Liu, Y., Zhang, Z., Zhou, R., & Zhu, Y. (2015). Distinguishing anthropogenic and natural sources of trace elements in soils undergoing recent 10-year rapid urbanization: a case of Donggang, Eastern China. *Environmental Science and Pollution Research, 22*(14), 10539–10550.

Qiu, L., Wang, K., Long, W., Wang, K., Hu, W., & amp; Amable, GS, . (2016). A comparative assessment of influences of human impacts on soil Cd based on stepwise linear regression, classification and regression tree, and random forest models. *PLoS ONE, 11*(3), e0151131.

Sawut, R., Kasim, N., Abliz, A., Hu, L., Yalkun, A., Maihemuti, B., & Qingdong, S. (2018). Possibility of optimized indications for assessment of heavy metal contents in soil around an open pit coal mine area. *International Journal of Applied Earth Observation and Geoinformation, 73,* 14–25.

Sun, C., Liu, J., Wang, Y., Sun, L., & Yu, H. (2013). Multivariate and geostatistical analyses of the spatial distribution and sources of heavy metals in agricultural soil in Dehui. *Northeast China. Chemosphere, 92*(5), 517–523.

Szopka, K., Karczewska, A., Jezierski, P., & Kabała, C. (2013). Spatial distribution of lead in the surface layers of mountain forest soils, an example from the Karkonosze National Park, Poland. *Geoderma, 192,* 259–268.

Tian, K., Huang, B., Xing, Z., & & Hu, W. . (2017). Geochemical baseline establishment and ecological risk evaluation of heavy metals in greenhouse soils from Dongtai, China. *Ecological indicators, 72,* 510–520.

Timofeev, I., Kosheleva, N., & Kasimov, N. (2018). Contamination of soils by potentially toxic elements in the impact zone of tungsten-molybdenum ore mine in the Baikal region: A survey and risk assessment. *Science of the total environment, 642,* 63–76.

Wang, H., & Lu, S. (2011). Spatial distribution, source identification and affecting factors of heavy metals contamination in urban-suburban soils of Lishui city. *China. Environmental Earth Sciences, 64*(7), 1921–1929.

Wang, L., Dai, L., Li, L., & Liang, T. (2018). Multivariable cocriging prediction and source analysis of potentially toxic elements (Cr, Cu, Cd, Pb, and Zn) in surface sediments from Dongting Lake, China. *Ecological Indicators, 94,* 312–319.

Wang, L. M., Wang, Q. S., Wen, H. H., Luo, J., & Wang, S. (2019). Heavy metals in agricultural soils from a typical township in Guangdong Province, China: Occurrences and spatial distribution. *Ecotoxicology and environmental safety, 168,* 184–191.

Wang, Z., Hong, C., Xing, Y., Wang, K., Li, Y., Feng, L., & Ma, S. (2018). Spatial distribution and sources of heavy metals in natural pasture soil around copper-molybdenum mine in Northeast China. *Ecotoxicology and environmental safety, 154,* 329–336.

Wang, Z., Meng, B., Zhang, W., Bai, J., Ma, Y., & Liu, M. (2018). Multi-Target Risk Assessment of Potentially Toxic Elements in Farmland Soil Based on the Environment-Ecological-Health Effect. *International journal of environmental research and public health, 15*(6), 1101.

Wu, C., Huang, J., Minasny, B., & Zhu, H. (2017). Two-dimensional empirical mode decomposition of heavy metal spatial variation in Southeast China. *Environmental Science and Pollution Research, 24*(9), 8302–8314.

Wu, C., Wu, J., Luo, Y., Zhang, H., & Teng, Y. (2008). Statistical and geoestatistical characterization of heavy metal concentrations in a contaminated area taking into account soil map units. *Geoderma, 144*(1–2), 171–179.

Wu, C., Wu, J., Luo, Y., Zhang, H., Teng, Y., & DeGloria, S. D. (2011). Spatial interpolation of severely skewed data with several peak values by the approach of integrating kriging and triangular irregular network interpolation. *Environmental Earth Sciences, 63*(5), 1093–1103.

Xie, Y., Chen, T. B., Lei, M., Yang, J., Guo, Q. J., Song, B., & Zhou, X. Y. (2011). Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis. *Chemosphere, 82*(3), 468–476.

Yang, L., Huang, B., Mao, M., Yao, L., Hickethier, M., & amp; Hu, W. . (2015). Trace metal accumulation in soil and their phytoavailability as affected by greenhouse types in north China. *Environmental science and pollution research, 22*(9), 6679–6686.

Yong, L. I. U., Huifeng, W. A. N. G., Xiaoting, L. I., & Jinchang, L. I. (2015). Heavy metal contamination of agricultural soils in Taiyuan. *China. Pedosphere, 25*(6), 901–909.

Yuan, G. L., Sun, T. H., Han, P., & Li, J. (2013). Environmental geochemical mapping and multivariate geostatistical analysis of heavy metals in a closed steel smelter: Capital Iron & Steel Factory, Beijing, China. *Journal of Geochemical Exploration, 130,* 15–21.

Zhang, J., Wang, Y., Liu, J., Liu, Q., & Zhou, Q. (2016). Multivariate and geostatistical analysis of heavy metals and their distribution in agricultural soil in Gongzhuling, Northeast China. *Journal of soils and sediments, 16*(2), 634–644.

Zhang, X., Wei, S., Sun, Q., Wadood, S. A., & amp; Guo, B. . (2018). Source identification and spatial distribution of arsenic and heavy metals in agricultural soil around Hunan industrial estate by positive matrix factorization model, principle of analysis and geo statistical analysis. *Ecotoxicology and environmental safety, 159,* 354–362.

Zhao, K., Liu, X., Xu, J., & Selim, H. M. (2010). Heavy metal contaminations in a soil–rice system: identification of spatial dependence in relation to soil properties of paddy fields. *Journal of Hazardous Materials, 181*(1–3), 778–787.

Zhao, Y., Wang, Z., Sun, W., Huang, B., Shi, X., & Ji, J. (2010). Spatial interrelations and multi-scale sources of soil heavy metal variability in a typical urban–rural transition area in Yangtze River Delta region of China. *Geoderma, 156*(3–4), 216–227.

Zhong, B., Liang, T., Wang, L., & Li, K. (2014). Applications of stochastic models and geostatistical analyses to study sources and spatial patterns of soil heavy metals in a metalliferous industrial district of China. *Science of the Total Environment, 490,* 422–434.

Zhou, J., Feng, K., Li, Y., & Zhou, Y. (2016). Factorial Kriging analysis and sources of heavy metals in soils of different land-use types in the Yangtze River Delta of Eastern China. *Environmental Science and Pollution Research, 23*(15), 14957–14967.

Romic, M., & Romic, D. (2003). Heavy metals distribution in agricultural topsoils in urban area. *Environmental geology, 43*(7), 795–805.

Romic, M., Hengl, T., Romic, D., & Husnjak, S. (2007). Representing soil pollution by heavy metals using continuous limitation scores. *Computers & Geosciences, 33*(10), 1316–1326.

Sollitto, D., Romic, M., Castrignan, A., Romic, D., & Bakic, H. (2010). Assessing heavy metal contamination in Zagreb region (Northwest Croatia) using multivariate geostatistics. *CATENA, 80*(3), 182–194.

Stazi, S. R., Antonucci, F., Pallottino, F., Costa, C., Marabottini, R., Petruccioli, M., & Menesatti, P. (2014). Hyperspectral visible–near infrared determination of arsenic concentration in soil. *Communications in soil science and plant analysis, 45*(22), 2911–2920.

Boruvka, L., & Vacha, R. (2006). Litavka river alluvium as model heavily polluted with potentially risk elements. In Phytoremediation of Metal-contaminated Soils (pp. 267–298). Springer, Dordrecht.

Gholizadeh, A., Borůvka, L., Vašát, R., Saberioon, M., Klement, A., Kratina, J., & Drábek, O. (2015). Estimation of potential toxic elements contamination in anthropogenic soils on brown coal mining dumpsite by reflectance spectroscopy: A case study. *PLoS ONE, 10*(2), e0117457.

Kváčová, M., Ash, C., Borůvka, L., Pavlí, L., Nikod, A., Němeček, K., & Drábek, O. (2015). Content of potentially toxic elements in forest soils of the Jizera Mountains Region. Environmental Modeling & amp; Evaluation , 20 (3), 183–195.

El Nemr, A., Khaled, A., & El Sikaily, A. (2006). Distribution and statistical analysis of leachable and total heavy metals in the sediments of the Suez Gulf. *Environmental monitoring and assessment, 118*(1–3), 89–112.

Omran, E. S. E. (2016). Inference model to predict heavy metals of Bahr El Baqar soils, Egypt using spectroscopy and chemometrics technique. *Modeling Earth Systems and Environment, 2*(4), 1–17.

Buttafuoco, G., Tarvainen, T., Jarva, J., & Guagliardi, I. (2016). Spatial variability and trigger values of arsenic in the surface urban soils of the cities of Tampere and Lahti. *Finland. Environmental Earth Sciences, 75*(10), 896.

Bourennane, H., Douay, F., Sterckeman, T., Villanneau, E., Ciesielski, H., King, D., & Baize, D. (2010). Mapping of anthropogenic trace elements input in agricultural topsoil from Northern France using enrichment factors. *Geoderma, 157*(3–4), 165–174.

Fritsch, C., Giraudoux, P., Curdassier, M., Douay, F., Raoul, F., Pruvot, C., & amp; Scheifler, R. . (2010). Spatial distribution of metals in smelter-impacted soils of woody habitats: Influence of landscape and soil properties, and risk for wildlife. *Chemosphere, 81*(2), 141–155.

Saby, N., Arrouays, D., Boulonne, L., Jolivet, C., & Pochot, A. (2006). Geostatistical assessment of Pb in soil around Paris. *France. Science of the total environment, 367*(1), 212–221.

Salvador-Blanes, S., Cornu, S., Bourennane, H., & King, D. (2006). Controls of the spatial variability of Cr concentration in topsoils of a central French landscape. *Geoderma, 132*(1–2), 143–157.

Lacarce, E., Saby, NP, Martin, MP, Marchant, BP, Boulonne, L. Mapping soil Pb stocks and availability in mainland France combining regression trees with robust geostatistics. Geoderma , 170 , 359–368.

Altfelder, S., Beyer, C., Duijnisveld, W. H., Schneider, J., & Streck, T. (2002). Distribution of Cd in the vicinity of a metal smelter: Interpolation of soil Cd concentrations with regard to regulative limits. *Journal of Plant Nutrition and Soil Science, 165*(6), 697–705.

Anagu, I., Ingwersen, J., Utermann, J., & Streck, T. (2009). Estimation of heavy metal sorption in German soils using artificial neural networks. *Geoderma, 152*(1–2), 104–112.

Akoto, O., Bortey-Sam, N., Ikenaka, Y., Nakayama, S. M., Baidoo, E., Yohannes, Y. B., & Ishizuka, M. (2017). Contamination Levels and Sources of Heavy Metals and a Metalloid in Surface Soils in the Kumasi Metropolis. *Ghana. Journal of Health and Pollution, 7*(15), 28–39.

Argyraki, A., & Kelepertzis, E. (2014). Urban soil geochemistry in Athens, Greece. *Science of the Total Environment, 482,* 366–377.

Kelepertzis, E. (2014). Accumulation of heavy metals in agricultural soils of Mediterranean: insights from Argolida basin, Peloponnese, Greece. *Geoderma, 221,* 82–90.

Kelepertzis, E., & Argyraki, A. (2015). Geochemical associations for evaluation of potential harmful elements in urban soils: lessons learn from Athens, Greece. *Applied Geochemistry, 59,* 63–73.

Korre, A., Durucan, S., & amp; Koutroumani, A. . (2002). Quantitative-spatial assessment of risks associated with high Pb loads in soils around Lavrio. *Greece. Applied Geochemistry, 17*(8), 1029–1045.

Chan, L. S., & sedimentNg, SL, Davis, AM, Yim, WWS, & Yeung, CH, . (2001). Magnetic properties and heavy-metal contents of contaminated seabed sediments of Penny's Bay. *Hong Kong. Marine Pollution Bulletin, 42*(7), 569–583.

Zhou, F., Guo, H., & Liu, L. (2007). Quantitative identification and source apportionment of anthropogenic heavy metals in marine sediment of Hong Kong. *Environmental geology, 53*(2), 295–305.

Zhou, F., Guo, H., & Hao, Z. (2007). Spatial distribution of heavy metals in Hong Kong's marine sediments and their human impacts: a GIS-based chemometric approach. *Marine Pollution Bulletin, 54*(9), 1372–1384.

Lee, C. S. L., Li, X., Shi, W., Cheung, S. C. N., & Thornton, I. (2006). Metal contamination in urban, suburban, and country park soils of Hong Kong: a study based on GIS and multivariate statistics. *Science of the Total Environment, 356*(1–3), 45–61.

Li, X., Lee, S. L., Wong, S. C., Shi, W., & Thornton, I. (2004). The study of metal contamination in urban soils of Hong Kong using a GIS-based approach. *Environmental pollution, 129*(1), 113–124.

Pásztor, L., Szabó, K. Z., Szatmári, G., Laborczi, A., & Horváth, Á. (2016). Mapping geogenic radon potential by regression kriging. *Science of the total environment, 544,* 883–891.

Chabukdhara, M., & Nema, A. K. (2013). Heavy metals assessment in urban soil around industrial clusters in Ghaziabad, India: probabilistic health risk approach. *Ecotoxicology and Environmental Safety, 87,* 57–64.

Chakraborty, S., Li, B., Deb, S., Paul, S. W., & DC, & Das, BS, . (2017). Predicting soil arsenic pools by visible near diffraction reflectance spectroscopy. *Geoderma, 296,* 30–37.

Chakraborty, S., Weindorf, D. C., Deb, S., Li, B., Paul, S., Choudhury, A., & Ray, D. P. (2017). Rapid assessment of regional soil arsenic pollution risk via diffusion reflectance spectroscopy. *Geoderma, 289,* 72–81.

Amini, MANOUCHEHR; Continuous soil pollution mapping using fuzzy logic and spatial interpolation. *Geoderma, 124* (3–4), 223–233.

Ayoubi, S., Jabbari, M., & amp; Khademi, H. (2018). Multiple linear modeling of soil properties, magnetic susceptibility and heavy metals in various land uses. Modeling Earth Systems and Environment , 1–11.

Bastami, K. D., Bagheri, H., Haghparast, S., Soltani, F., Hamzehpoor, A., & Bastami, M. D. (2012). Geochemical and geo-statistical assessment of selected heavy metals in the surface sediments of the Gorgan Bay. *Iran. Marine pollution bulletin, 64*(12), 2877–2884.

Dayani, M., & Mohammadi, J. (2010). Geostatistical assessment of Pb, Zn and Cd contamination in near-surface soils of the urban-mining transitional region of Isfahan. *Iran. Pedosphere, 20*(5), 568–577.

Delavar, M. A., & Safari, Y. (2016). Spatial distribution of heavy metals in Zinc Town, northwest Iran. *International Journal of Environmental Science and Technology, 13*(1), 297–306.

Ghanbarpour, M. R., Goorzadi, M., & Vahabzade, G. (2013). Spatial variability of heavy metals in surficial sediments: Tajan River Watershed. *Iran. Sustainability of Water Quality and Ecology, 1,* 48–58.

Hani, A., & Pazira, E. (2011). Heavy metals assessment and identification of their sources in agricultural soils of Southern Tehran. *Iran. Environmental monitoring and assessment, 176*(1–4), 677–691.

Hasani, S., Asghari, O., Ardejani, F. D., & Yousefi, S. (2017). Spatial modeling of hazardous elements at a waste dump using a geostatistical approach: a case study of Sarcheshmeh copper mine. *Iran. Environmental Earth Science, 76*(15), 532.

Keshavarzi, B., Abbasi, S., Moore, F., Mehravar, S., Sorooshian, A., Soltani, N., & Najmeddin, A. (2018). Contamination level, source identification and risk assessment of potentially toxic elements (PTEs) and polycyclic aromatic hydrocarbons (PAHs) in street dust of an important commercial center in Iran. *Environmental management, 62*(4), 803–818.

Khosravi, V., Ardejani, F. D., Yousefi, S., & Aryafar, A. (2018). Monitoring soil lead and zinc contents by combination of spectroscopy with extreme learning machine and other data mining methods. *Geoderma, 318,* 29–41.

Krami, L. K., Amiri, F., Sefiyanian, A., Shariff, A. R. B. M., Tabatabaie, T., & Pradhan, B. (2013). Spatial patterns of heavy metals in soil under different geological structures and land uses for assessing metal enrichments. *Environmental monitoring and assessment, 185*(12), 9871–9888.

Najmeddin, A., Keshavarzi, B., Moore, F., & Lahijanzadeh, A. (2018). Source and health risk assessment of potential toxic elements in road dust from urban industrial areas of Ahvaz megacity. *Iran. Environmental Geochemistry and Health, 40*(4), 1187–1208.

Piroozfar, P., Alipour, S., Modabberi, S., & Cohen, D. (2018). Application of Geochemistry and VNIR Spectroscopy in Mapping Heavy Metal Pollution of Stream Sediments in the Takab Mining Area, NW of Iran. *Acta Geologica Sinica-English Edition, 92*(6), 2382–2394.

Sakizadeh, M., Martín, J. A. R., Zhang, C., Sharafabadi, F. M., & Ghorbani, H. (2018). Trace elements concentrations in soil, desert-adapted and non-desert plants in central Iran: Spatial patterns and uncertainty analysis. *Environmental pollution, 243,* 270–281.

Sakizadeh, M., Mirzaei, R., & Ghorbani, H. (2015). The extent and prediction of heavy metal pollution in soils of Shahrood and Damghan. *Iran. Bulletin of environmental contamination and toxicology, 95*(6), 770–776.

Sakizadeh, M., Mirzaei, R., & Ghorbani, H. (2017). Support vector machine and artificial neural network to model soil pollution: a case study in Semnan Province. *Iran. Neural Computing and Applications, 28*(11), 3229–3238.

Sakizadeh, M., Sattari, M. T., & Ghorbani, H. (2017). A new method to consider spatial risk assessment of cross-correlated heavy metals using geo-statistical simulation. *Journal of Mining and Environment, 8*(3), 373–391.

Soffianian, AR, Bakir, HB, & Khodakarami, L(2015)..Evaluation of heavy metals concentration in soil using GIS, RS and Geostatistics

Yousefi, G., Homaee, M., & amp; Norouzi, AA, . (2018). Estimating soil heavy metals concentration at large scale using visible and near-infrared reflectance spectroscopy. *Environmental Monitoring and Assessment, 190*(9), 513.

McGrath, D., Zhang, C., & Carton, O. T. (2004). Geostatistical analysis and harzard assessment on soil lead in Silvermines area. *Ireland. Environmental Pollution, 127*(2), 239–248.

McIlwaine, R., Doherty, R., Cox, S. F., & Cave, M. (2017). The relationship between historical development and

potentially toxic element concentrations in urban soils. *Environmental pollution, 220,* 1036–1049.

Meerschman, E., Cockx, L., & Van Meirvenne, M. (2011). A geostatistical two-phase sampling strategy to map soil heavy metal concentrations in a former war zone. *European journal of soil science, 62*(3), 408–416.

Zhang, C. (2006). Using multivariate analyses and GIS to identify pollutants and their spatial patterns in urban soils in Galway. *Ireland. Environmental pollution, 142*(3), 501–511.

Zhang, C., Tang, Y., Luo, L., & Xu, W. (2009). Outlier identification and visualization for Pb concentrations in urban soils and its implications for potential contaminated land identification. *Environmental Pollution, 157*(11), 3083–3090.

Albanese, S., De Vivo, B., Lima, A., Cicchella, D., Civitillo, D., & Cosenza, A. (2010). Geochemical baselines and risk assessment of the Bagnoli brownfield site coastal sea sediments (Naples, Italy). *Journal of Geochemical Exploration, 105*(1–2), 19–33.

Ballabio, C., & Comolli, R. (2010). Mapping Heavy Metal Content in Multi-Kernel SVR and LiDAR Derived Data. In *Digital Soil Mapping* (pp. 205–216). Springer, Dordrecht.

Bertazzon, S., Micheletti, C., Critto, A., & Marcomini, A. (2006). Spatial analysis in ecological risk assessment: Pollutant bioaccumulation in clams Tapes philipinarum in the Venetian lagoon (Italy). *Computers, environment and urban systems, 30*(6), 880–904.

Covelli, S., Faganeli, J., Horvat, M., & Brambati, A. (2001). Mercury contamination of coastal sediments as the result of long-term cinnabar mining activity (Gulf of Trieste, northern Adriatic sea). *Applied Geochemistry, 16*(5), 541–558.

Guagliardi, I., Cicchella, D., & De Rosa, R. (2012). A geostatistical approach to assessing concentration and spatial distribution of heavy metals in urban soils. *Water, Air, & Soil Pollution, 223*(9), 5983–5998.

Guagliardi, I., Cicchella, D., De Rosa, R., & Buttafuoco, G. (2015). Assessment of lead pollution in topsoils of a southern Italy area: analysis of urban and peri-urban environment. *Journal of Environmental Sciences, 33,* 179–187.

Guagliardi, I., Rovella, N., Apollaro, C., Bloise, A., De Rosa, R., Scarciglia, F., & Buttafuoco, G. (2016). Modelling seasonal variations of natural radioactivity in soils: A case study in southern Italy. *Journal of Earth System Science, 125*(8), 1569–1578.

Imperato, M., Adamo, P., Naimo, D., Arienzo, M., Stanzione, D., & Violante, P. (2003). Spatial distribution of heavy metals in urban soils of Naples city (Italy). *Environmental pollution, 124*(2), 247–256.

Imrie, C. E., Korre, A., Munoz-Melendez, G., Thornton, I., & Durucan, S. (2008). Application of factorial kriging analysis to the FOREGS European topsoil geochemistry database. *Science of the total environment, 393*(1), 96–110.

Lancianese, V., & Dinelli, E. (2015). Different Spatial Methods in Regional Geochemical Mapping at High Density Sampling: An Application on the Sediment of Romagna Apennines, Northern Italy. *Journal of Geochemical Exploration, 154,* 143–155.

Petrik, A., Thiombane, M., Albanese, S., Lima, A., & De Vivo, B. (2018). Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through compositional discrimination analysis: A case study on the Campanian topsoil data. *Geoderma, 331,* 87–99.

Poggio, L., & Vrščaj, B. (2009). A GIS-based human health risk assessment for urban green space planning — An example from Grugliasco (Italy). *Science of the total environment, 407*(23), 5961–5970.

Ungaro, F., Ragazzi, F., Cappellin, R., & Giandon, P. (2008). Arsenic concentration in the soils of the Brenta Plain (Northern Italy): mapping the probability of exceeding contamination thresholds. *Journal of Geochemical Exploration, 96*(2–3), 117–131.

Vaccaro, S., Sobiecka, E., Contini, S., Locoro, G., Free, G., & Gawlik, B. M. (2007). The application of positive matrix factorization in the analysis, characterisation and detection of contaminated soils. *Chemosphere, 69*(7), 1055–1063.

Choe, E., Kim, K. W., Bang, S., Yoon, I. H., & Lee, K. Y. (2009). Qualitative analysis and mapping of heavy metals in abandoned Au – Ag mines using NIR spectroscopy. *Environmental geology, 58*(3), 477–482.

Hwang, C. K., Cha, J. M., Kim, K. W., & Lee, H. K. (2001). Application of multivariate statistical analysis and geographic information system to trace element in Chungnam Coal Mine area. *Korea. Applied Geochemistry, 16*(11–12), 1455–1464.

Suh, J., Lee, H., & Choi, Y. (2016). A rapid, accurate, and efficient method to map heavy metal-contaminated soils of abandoned mine sites using converted portable XRF data and GIS. *International journal of environmental research and public health, 13*(12), 1191.

Praveena, S. M. (2019). Spatial eco-risk assessment and prediction of heavy metal pollution in surface soil: a preliminary assessment of an urban area from a developing country. *Toxin Reviews, 38*(2), 135–142.

Castro-Larragoitia, J., Kramar, U., Monroy-Fernández, M. G., Viera-Décida, F., & García-González, E. G. (2013). Heavy metal and arsenic dispersion in a copper-skarn mining district in a Mexican semi-arid environment: sources, pathways and fate. *Environmental earth sciences, 69*(6), 1915–1929.

Chiprés, J. A., Salinas, J. C., Castro-Larragoitia, J., & Monroy, M. G. (2008). Geochemical mapping of major and trace elements in soils from the Altiplano Potosino, Mexico: a multi-scale comparison. Geochemistry: Exploration, Environment, Analysis, 8(3–4), 279–290.

Cortés, J. L., Bautista, F., Delgado, C., Quintana, P., Aguilar, D., García, A., ... & Gogichaishvili, A. (2017). Spatial distribution of heavy metals in urban dust from Ensenada, Baja California, Mexico. Revista Chapingo. Serie Ciencias Forestales y del Ambiente, 23(1), 47–60.

Covarrubias, S. A., Flores de la Torre, J. A., Maldonado Vega, M., Avelar González, F. J., & Peña Cabriales, J. J. (2018). Spatial Variability of Heavy Metals in Soils and Sediments of "La Zacatecana" Lagoon, Mexico. Applied and Environmental Soil Science, 2018.

Ihl, T., Bautista, F., Cejudo, R., Delgado, C., Quintana, P., Aguilar, D., & Gogichaishvili, A. (2015). Concentration of toxic elements in topsoils of the metropolitan area of

Mexico City: a spatial analysis using Ordinary kriging and Indicator kriging. *Rev. Int. Contam. Ambie, 31*(1), 47–62.

Meza-Montenegro, M. M., Gandolfi, A. J., Santana-Alcántar, M. E., Klimecki, W. T., Aguilar-Apodaca, M. G., Del Río-Salas, R., & Meza-Figueroa, D. (2012). Metals in residential soils and cumulative risk assessment in Yaqui and Mayo agricultural valleys, northern Mexico. *Science of the total environment, 433,* 472–481.

Morton-Bermea, O., Hernández-Álvarez, E., González-Hernández, G., Romero, F., Lozano, R., & Beramendi-Orosco, L. E. (2009). Assessment of heavy metal pollution in urban topsoils from the metropolitan area of Mexico City. *Journal of Geochemical Exploration, 101*(3), 218–224.

Razo, I., Carrizales, L., Castro, J., Díaz-Barriga, F., & Monroy, M. (2004). Arsenic and heavy metal pollution of soil, water and sediments in a semi-arid climate mining area in Mexico. *Water, Air, and Soil Pollution, 152*(1–4), 129–152.

Rodríguez-Salazar, M. T., Morton-Bermea, O., Hernández-Alvarez, E., Lozano, R., & Tapia-Cruz, V. (2011). The study of metal contamination in urban topsoils of Mexico City using GIS. *Environmental Earth Sciences, 62*(5), 899–905.

Salas-Luevano, M. A., Manzanares-Acuna, E., Letechipia-de Leon, C., Hernandez-Davila, V. M., & Vega-Carrillo, H. R. (2011). Lead concentration in soil from an old mining town. *Journal of Radioanalytical and Nuclear Chemistry, 289*(1), 35–39.

Santos-Santos, E., Yarto-Ramírez, M., Gavilán-García, I., Castro-Díaz, J., Gavilán-García, A., Rosiles, R., & López-Villegas, T. (2006). Analysis of arsenic, lead and mercury in farming areas with mining contaminated soils at Zacatecas, Mexico. *Journal of the Mexican Chemical Society, 50*(2), 57–63.

Timofeev, I. V., Kosheleva, N. E., Kasimov, N. S., Gunin, P. D., & Sandag, E. A. (2016). Geochemical transformation of soil cover in copper-molybdenum mining areas (Erdenet, Mongolia). *Journal of soils and sediments, 16*(4), 1225–1237.

El Badaoui, H., Abdallaoui, A., Manssouri, I., & Lancelot, L. (2013). Application of artificial neural networks of MLP type for the prediction of heavy metals in Moroccan aquatic sediments. *International Journal of Computational Engineering Research, 3*(6), 75–81.

Khalil, A., Hanich, L., Bannari, A., Zouhri, L., Pourret, O., & Hakkou, R. (2013). Assessment of soil contamination around an abandoned mine in a semi-arid environment using geochemistry and geostatistics: pre-work of geochemical process modeling with numerical models. *Journal of Geochemical Exploration, 125,* 117–129.

Brus, D. J., De Gruijter, J. J., Walvoort, D. J. J., De Vries, F., Bronswijk, J. J. B., Römkens, P. F. A. M., & De Vries, W. (2002). Mapping the probability of exceeding critical thresholds for cadmium concentrations in soils in the Netherlands. *Journal of Environmental Quality, 31*(6), 1875–1884.

Cattle, J. A., McBratney, A., & Minasny, B. (2002). Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. *Journal of Environmental Quality, 31*(5), 1576–1588.

Nickel, S., Hertel, A., Pesch, R., Schröder, W., Steinnes, E., & Uggerud, H. T. (2014). Modeling and mapping of spatio-temporal trends in heavy metal accumulation in moss and natural surface soil monitored 1990–2010 throughout Norway by multivariate generalized linear models and geostatistics. *Atmospheric Environment, 99,* 85–93.

Alam, N., Ahmad, S. R., Qadir, A., Ashraf, M. I., Lakhan, C., & Lakhan, V. C. (2015). Use of statistical and GIS techniques for assessment and prediction of heavy metals in soils of Lahore City. *Pakistan. Environmental Monitoring and Assessment, 187*(10), 636.

Ali, S. M., & Malik, R. N. (2011). Spatial distribution of metals in top soils of Islamabad City. *Pakistan. Environmental monitoring and assessment, 172*(1–4), 1–16.

Kostka, A., & Leśniak, A. (2020). Spatial and geochemical aspects of heavy metal distribution in lacustrine sediments, using the example of Lake Wigry (Poland). *Chemosphere, 240,* 124879.

Antunes, I. M. H. R., Albuquerque, M. T. D., & Roque, N. (2018). Spatial environmental risk evaluation of potential toxic elements in stream sediments. *Environmental geochemistry and health, 40*(6), 2573–2585.

Candeias, C., Da Silva, E. F., Salgueiro, A. R., Pereira, H. G., Reis, A. P., Patinha, C., & Ávila, P. H. (2011). Assessment of soil contamination by potentially toxic elements in the Aljustrel mining area in order to implement soil reclamation strategies. *Land Degradation & Development, 22*(6), 565–585.

Reis, A. P., Da Silva, E. F., Sousa, A. J., Matos, J., Patinha, C., Abenta, J., & Fonseca, E. C. (2005). Combining GIS and stochastic simulation to estimate spatial patterns of variation for lead at the Lousal mine. *Portugal. Land degradation & development, 16*(2), 229–242.

Reis, A. P., Sousa, A. J., Silva, E. F. D., & Fonseca, E. C. (2005). Application of geostatistical methods to arsenic data from soil samples of the Cova dos Mouros mine (Vila Verde-Portugal). *Environmental geochemistry and health, 27*(3), 259–270.

Salgueiro, A. R., Ávila, P. F., Pereira, H. G., & Oliveira, J. S. (2008). Geostatistical estimation of chemical contamination in stream sediments: The case study of Vale das Gatas mine (northern Portugal). *Journal of Geochemical Exploration, 98*(1–2), 15–21.

Tavares, M. T., Sousa, A. J., & Abreu, M. M. (2008). Ordinary kriging and indicator kriging in the cartography of trace elements contamination in São Domingos mining site (Alentejo, Portugal). *Journal of Geochemical Exploration, 98*(1–2), 43–56.

Peng, Y., Kheir, R., Adhikari, K., Malinowski, R., Greve, M., Knadel, M., & Greve, M. (2016). Digital mapping of toxic metals in Qatari using remote sensing and ancillary data. *Remote Sensing, 8*(12), 1003.

Chalov, S. R., Jarsjö, J., Kasimov, N. S., Romanchenko, A. O., Pietroń, J., Thorslund, J., & Promakhova, E. V. (2015). Spatio-temporal variation of sediment transport in the Selenga River Basin. *Mongolia and Russia. Environmental Earth Sciences, 73*(2), 663–680.

Komnitsas, K., & Modis, K. (2006). Soil risk assessment of As and Zn contamination in a coal mining region using geostatisretics. *Science of the Total Environment, 371*(1–3), 190–196.

Sergeev, A. P., Buevich, A. G., Baglaeva, E. M., & Shichkin, A. V. (2019). Combining spatial autocorrelation with machine

learning increases prediction accuracy of soil heavy metals. *CATENA, 174,* 425–435.

Sergeev, A. P., Tarasov, D. A., Buevich, A. G., Subbotina, I. E., Shichkin, A. V., Sergeeva, M. V., & Lvova, O. A. (2017, June). High variation subarctic topsoil pollutant concentration prediction using neural network residual kriging. In AIP Conference Proceedings (Vol. 1836, No. 1, p. 020023). AIP Publishing.

Shichkin, AV, Buevich, AG, & amp; Sergeev, AP (2018, July). Comparison of artificial neural network, random forest and random perceptron forest. In AIP Conference Proceedings (Vol. 1982, No. 1, p. 020005). AIP Publishing.

Subbotina, I. E., Buevich, A. G., Shichkin, A. V., Sergeev, A. P., Tarasov, D. A., Tyagunov, A. G., ... & Baglaeva, E. M. (2018, July). Multilayer perceptron, generalized regression neural network, and hybrid model in predicting the spatial distribution of impurity in the topsoil of urbanized area. In AIP Conference Proceedings (Vol. 1982, No. 1, p. 020004). AIP Publishing.

Tarasov, D. A., Buevich, A. G., Sergeev, A. P., Shichkin, A. V., & Baglaeva, E. M. (2017, June). Topsoil pollution forecasting using artificial neural networks on the example of the abnormally distributed heavy metal at Russian subarctic. In AIP Conference Proceedings (Vol. 1836, No. 1, p. 020024). AIP Publishing.

Tarasov, D. A., Buevich, A. G., Sergeev, A. P., & Shichkin, A. V. (2018). High variation topsoil pollution forecasting in the Russian Subarctic: Using artificial neural networks combined with residual kriging. *Applied Geochemistry, 88,* 188–197.

Dragovic, R., Gajic, B., Dragovic, S., Kordevic, M., Kordevic, M., Mihailovic, N., & Onjia, A. (2014). Assessment of the Impact of Geographical Factors on the Spatial Distribution of Heavy Metals in the Steel Production Facility in Smederevo (Serbia). *Journal of cleaner production, 84,* 550–562.

Dragovic, S., Iuji, M., Slavkovic-Beškoski, L., Gajic, B., Bajat, B., Kilibarda, M., & Onjia, A. (2013). Trace element distribution in surface soils from coal burning power production area: A power plant site in Serbia. *CATENA, 104,* 288–296.

Finzgar, N., Jez, E., Voglar, D., & Lestan, D. (2014). Spatial distribution of metal contamination before and after remediation in the Meza Valley, Slovenia. *Geoderma, 217,* 135–143.

De Villiers, S., Thiart, C., & Basson, N. C. (2010). Identification of sources of environmental lead in South Africa from surface soil geochemical maps. *Environmental geochemistry and health, 32*(5), 451–459.

Kim, H. R., Kim, K. H., Yu, S., Moniruzzaman, M., Hwang, S. I., Lee, G. T., & Yun, S. T. (2019). Better assessment of the distribution of As and Pb in soils in a former smelting area, using ordinary co-kriging and sequential Gaussian co-simulation of portable X-ray fluorescence (PXRF) and ICP-AES data. *Geoderma, 341,* 26–38.

Kim, S. M., Choi, Y., Yi, H., & amp; Park, HD, . (2017). Geostatistical prediction of heavy metal concentrations in streams considering the stream networks. *Environmental Earth Sciences, 76*(2), 72.

Acosta, J. A., Faz, A., Martínez-Martínez, S., Zornoza, R., Carmona, D. M., & Kabas, S. (2011). Multivariate statistical and GIS-based approach to evaluate heavy metals behavior in mine sites for future reclamation. *Journal of Geochemical Exploration, 109*(1–3), 8–17.

Choe, E., van der Meer, F., van Ruitenbeek, F., van der Werff, H., de Smeth, B., & Kim, K. W. (2008). Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and remote sensing: A case study of Rodalquilar mining area. *SE Spain. Remote Sensing of Environment, 112*(7), 3222–3233.

Delgado, J., Nieto, J. M., & Boski, T. (2010). Analysis of the spatial variation of heavy metals in the Guadiana Estuary sediments (SW Iberian Peninsula) based on GIS-mapping techniques. *Estuarine, Coastal and Shelf Science, 88*(1), 71–83.

Fernández, S., Cotos-Yáñez, T., Roca-Pardiñas, J., & Ordóñez, C. (2018). Geographically weighted principal components analysis to assess diffuse pollution sources of soil heavy metal: application to rough mountain areas in Northwest Spain. *Geoderma, 311,* 120–129.

Franco, C., Soares, A., & Delgado, J. (2006). Geostatistical modeling of heavy metal contamination in the topsoil of Guadiamar river margins (S Spain) using a stochastic simulation technique. *Geoderma, 136*(3–4), 852–864.

Gabarron, M., Faz, A., Martínez-Martínez, S., Zornoza, R., & Acosta, J. A. (2017). Assessment of metals behavior in industrial soil using sequential extraction, multivariable analysis and geostatistical approach. *Journal of Geochemical Exploration, 172,* 174–183.

Gallego, J. L., Ordóñez, A., & Loredo, J. (2002). Investigation of trace element sources from an industrialized area (Aviles, northern Spain) using multivariate statistical methods. *Environment International, 27*(7), 589–596.

Kemper, T., & Sommer, S. (2002). Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environmental science & technology, 36*(12), 2742–2747.

Martín, J. A. R., Arias, M. L., & Corbí, J. M. G. (2006). Heavy metals contents in agricultural topsoils in the Ebro basin (Spain). Application of the multivariate geostatistical methods to study spatial variations. Environmental pollution, 144(3), 1001–1012.

Martín, J. R., Ramos-Miras, J. J., Boluda, R., & Gil, C. (2013). Spatial relations of heavy metals in arable and greenhouse soils of a Mediterranean environment region (Spain). *Geoderma, 200,* 180–188.

Morillo, J., Usero, J., & Gracia, I. (2004). Heavy metal distribution in marine sediments from the southwest coast of Spain. *Chemosphere, 55*(3), 431–442.

Nadal, M., Schuhmacher, M., & Domingo, J. L. (2004). Metal pollution of soils and vegetation in an area with petrochemical industry. *Science of the total environment, 321*(1–3), 59–69.

Rodriguez, J. A., Nanos, N., Grau, J. M., Gil, L., & Lopez-Arias, M. (2008). Multiscale analysis of heavy metal contents in Spanish agricultural topsoils. *Chemosphere, 70*(6), 1085–1096.

Santos-Francés, F., Martínez-Graña, A., Zarza, C. Á., Sánchez, A. G., & Rojo, P. A. (2017). Spatial distribution of heavy metals and the environmental quality of soil in the Northern

Plateau of Spain by geostatistical methods. *International journal of environmental research and public health, 14*(6), 568.

Kishné, A. S., Bringmark, E., Bringmark, L., & Alriksson, A. (2003). Comparison of ordinary and lognormal kriging on skewed data of total cadmium in forest soils of Sweden. *Environmental monitoring and assessment, 84*(3), 243–263.

Hofer, C., Borer, F., Bono, R., Kayser, A., & Papritz, A. (2013). Predicting topsoil heavy metal content of parcels of land: An empirical validation of customary and constrained lognormal block kriging and conditional simulations. *Geoderma, 193,* 200–212.

Schnabel, U., & Tietje, O. (2003). Explorative data analysis of heavy metal contaminated soil using multidimensional spatial regression. *Environmental Geology, 44*(8), 893–904.

Chu, H. J., Lin, Y. P., Jang, C. S., & Chang, T. K. (2010). Delineating the gambling zone of multiple soil pollutants by multivariate indicator kriging and conditioned hypercube sampling. *Geoderma, 158*(3–4), 242–251.

Juang, K. W., Chen, Y. S., & Lee, D. Y. (2004). Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils. *Environmental Pollution, 127*(2), 229–238.

Lin, Y. P., Cheng, B. Y., Chu, H. J., Chang, T. K., & Yu, H. L. (2011). Assessing how heavy metal pollution and human activity are related by using logistic regression and kriging methods. *Geoderma, 163*(3–4), 275–282.

Lin, Y. P. (2002). Multivariate geostatistical methods to identify and map spatial variations of soil heavy metals. *Environmental geology, 42*(1), 1–10.

Lin, Y. P., Cheng, B. Y., Shyu, G. S., & Chang, T. K. (2010). Combining a finite mixture distribution model with indicator kriging to delineate and map the spatial patterns of soil heavy metal pollution in Chunghua County, central Taiwan. *Environmental Pollution, 158*(1), 235–244.

Simasuwannarong, B., Satapanajaru, T., Khuntong, S., & Pengthamkeerati, P. (2012). Spatial distribution and risk assessment of As, Cd, Cu, Pb, and Zn in topsoil at Rayong Province, Thailand. *Water, Air, & Soil Pollution, 223*(5), 1931–1943.

Gannouni, S., Rebai, N., & Abdeljaoued, S. (2012). A spectroscopic approach to assess heavy metals. *Journal of Geographic Information System, 4,* 242–253.

Agca, N. (2015). Spatial distribution of heavy metal content in soils around an industrial area in Southern Turkey. *Arabian Journal of Geosciences, 8*(2), 1111–1123.

Ağca, N., & Özdel, E. (2014). Assessment of spatial distribution and possible sources of heavy metals in the Sariseki-Dörtyol District in Hatay Turkey. *Environmental earth science, 71*(3), 1033–1047.

CoKun, M., Steinnes, E., Frontasyeva, M. V., Sjobakk, T. E., & Demkina, S. (2006). Heavy metal pollution of the soil in the Thrace region. *Turkey. Environmental monitoring and assessment, 119*(1–3), 545–556.

Dindaroğlu, T. (2014). The use of the GIS Kriging technique to determine the spatial changes of natural radionuclide concentrations in soil and forest cover. *Journal of Environmental Health Science and Engineering, 12*(1), 130.

Karanlık, S., Ağca, N., & Yalçın, M. (2011). Spatial distribution of heavy metals content in Amik Plain (Hatay, Turkey). *Environmental Monitoring and Assessment, 173*(1–4), 181–191.

Ersoy, A., Yunsel, T. Y., & Atici, Ü. (2008). Geostatistical conditional simulation for the assessment of contaminated land by abandoned heavy metal mining. *Environmental toxicology, 23*(1), 96–109.

Ersoy, A., Yunsel, T. Y., & Cetin, M. (2004). Characterization of land contaminated by heavy metal mining using geostatistical methods. *Archives of Environmental Contamination and Toxicology, 46*(2), 162–175.

Gay, J. R., & Korre, A. (2006). A spatially-evaluated methodology for assessing risk to a population from contaminated land. *Environmental Pollution, 142*(2), 227–234.

Kirkwood, C., Cave, M., Beamish, D., Grebby, S., & Ferreira, A. (2016). A machine learning approach to geochemical mapping. *Journal of Geochemical Exploration, 167,* 49–61.

Lark, R. M., Bellamy, P. H., & Rawlins, B. G. (2006). Spatiotemporal variability of some metal concentrations in the soil of eastern England, and implications for soil monitoring. *Geoderma, 133*(3–4), 363–379.

Marchant, B. P., Tye, A. M., & Rawlins, B. G. (2011). The assessment of point-source and diffuse soil metal pollution using robust geostatistical methods: a case study in Swansea (Wales, UK). *European Journal of Soil Science, 62*(3), 346–358.

Rawlins, B. G., Lark, R. M., & O'donnell, KE, Tye, AM, & Lister, TR, . (2005). The assessment of point and diffuse metal pollution of soils from an urban geochemical survey of Sheffield. *England. Soil use and management, 21*(4), 353–362.

Zhen, J., Pei, T., & Xie, S. (2019). Kriging methods with auxiliary nighttime lights data to detect potentially toxic metals concentrations in soil. *Science of The Total Environment, 659,* 363–371.

Abel, M. T., Suedel, B., Presley, S. M., Rainwater, T. R., Austin, G. P., Cox, S. B., & Leftwich, B. D. (2010). Spatial distribution of lead concentrations in urban surface soils of New Orleans. *Louisiana USA. Environmental geochemistry and health, 32*(5), 379–389.

Aelion, C. M., Davis, H. T., McDermott, S., & Lawson, A. B. (2008). Metal concentrations in rural topsoil in South Carolina: potential for human health impact. *Science of the total environment, 402*(2–3), 149–156.

Bugdalski, L., Lemke, L. D., & McElmurry, S. P. (2014). Spatial Variation of Soil Lead in an Urban Community Garden: Implications for Risk-Based Sampling. *Risk Analysis, 34*(1), 17–27.

Diawara, M. M., Litt, J. S., Unis, D., Alfonso, N., Martinez, L., Crock, J. G., & Carsella, J. (2006). Arsenic, cadmium, lead, and mercury in surface soils, Pueblo, Colorado: implications for population health risk. *Environmental Geochemistry and Health, 28*(4), 297–315.

Gallagher, F. J., Pechmann, I., Bogden, J. D., Grabosky, J., & Weis, P. (2008). Soil metal concentrations and vegetative assemblage structure in an urban brownfield. *Environmental Pollution, 153*(2), 351–361.

Griffith, D. A. (2002). The geographic distribution of soil lead concentration: description and concerns. *URISA Journal, 14*(1), 5–14.

Ha, H., Olson, J. R., Bian, L., & Rogerson, P. A. (2014). Analysis of heavy metal sources in soil using kriging interpolation on principal components. *Environmental science & technology, 48*(9), 4999–5007.

Machemer, S. D., & Hosick, T. J. (2004). Determination of soil lead variability in residential soil for remediation decision making. *Water, Air, and Soil Pollution, 151*(1–4), 305–322.

Mielke, H. W., Berry, K. J., Mielke, P. W., Powell, E. T., & Gonzales, C. R. (2005). Multiple metal accumulation as a factor in learning achievement within various New Orleans elementary school communities. *Environmental Research, 97*(1), 67–75.

Mielke, H. W., Gonzales, C. R., Powell, E. T., & Mielke, P. W. (2013). Environmental and health disparities in residential communities of New Orleans: The need for soil lead intervention to advance primary prevention. *Environment international, 51*, 73–81.

Mielke, H. W., Gonzales, C., Powell, E., & Mielke, P. W. (2005). Changes of multiple metal accumulation (MMA) in New Orleans soil: Preliminary evaluation of differences between survey I (1992) and survey II (2000). *International Journal of Environmental Research and Public Health, 2*(2), 308–313.

Pandit, C. M., Filippelli, G. M., & Li, L. (2010). Estimation of heavy-metal contamination in soil using reflectance spectroscopy and partial least-squares regression. *International Journal of Remote Sensing, 31*(15), 4111–4123.

Paul, R., & Cressie, N. (2011). Lognormal block kriging for contaminated soil. *European journal of soil science, 62*(3), 337–345.

Reeves, M. K., Perdue, M., Munk, L. A., & Hagedorn, B. (2018). Predicting risk of trace element pollution from municipal roads using site-specific soil samples and remotely sensed data. *Science of the Total Environment, 630,* 578–586.

Schwarz, K., Weathers, K. C., Pickett, S. T., Lathrop, R. G., Pouyat, R. V., & Cadenasso, M. L. (2013). A comparison of three empirically based, spatially explicit predictive models of residential soil Pb concentrations in Baltimore, Maryland, USA: understanding the variability within cities. *Environmental geochemistry and health, 35*(4), 495–510.

Solt, M. J., Deocampo, D. M., & Norris, M. (2015). Spatial distribution of lead in Sacramento, California, USA. *International Journal of Environmental Research and Public Health, 12*(3), 3174–3187.

Webster, J. P., Kane, T. J., Obrist, D., Ryan, J. N., & Aiken, G. R. (2016). Estimating mercury emissions resulting from wildfire in forests of the Western United States. *Science of the Total Environment, 568,* 578–586.

Woodruff, L., Cannon, W. F., Smith, D. B., & Solano, F. (2015). The distribution of selected elements and minerals in the soil of the United States. *Journal of Geochemical Exploration, 154,* 49–60.

Wu, J., Edwards, R., He, X. E., Liu, Z., & Kleinman, M. (2010). Spatial analysis of bioavailable soil lead concentrations in Los Angeles. *California. Environmental research, 110*(4), 309–317.

Yesilonis, I. D., Pouyat, R. V., & Neerchal, N. K. (2008). Spatial distribution of metals in soils in Baltimore, Maryland: role of native parent material, proximity to major roads, housing age and screening guidelines. *Environmental Pollution, 156*(3), 723–731.

Using geostatistics and machine learning models to analyze the influence of soil nutrients and terrain attributes on lead prediction in forest soils

Samuel Kudjo Ahado[1] · Prince Chapman Agyeman[2] · Luboš Borůvka[1] · Radoslava Kanianska[3] · Chukwudi Nwaogu[4]

Abstract

The study aimed at investigating the possibility of predicting lead (Pb) in forest soils by combining terrain attributes and soil nutrients using geostatistics and machine learning algorithms (MLAs). The study was partitioned into three categories: predicting Pb in forest soil using terrain attributes and RK (Context 1); predicting Pb in forest soil using soil nutrients and RK (Context 2); and lastly predicting Pb in forest soils using a combination of soil nutrients, terrain attributes, and RK (Context 3). Stochastic Gradient Boosting-regression kriging (SGB-RK), cubist regression kriging (CUB_RK), quantile regression forest kriging(QRF_RK) and k nearest neighbour regression kriging (KNN_RK) were the modeling approaches used in the estimation of lead (Pb) concentration in forest soil. The results showed that combining the terrain attribute as an auxiliary dataset with QRF_RK proved to be the most effective method for predicting Pb in forest soil (context 1). The most effective method for predicting Pb in forest soil was to combine soil nutrients as an auxiliary dataset with SGB_RK (context 2). Combining cubist_RK with an ancillary dataset of soil nutrients and terrain attributes is the most effective method for predicting Pb in forest soils (context 3). In addition, combining ancillary variables such as soil nutrients and terrain attributes with cubist_RK generated the best results for estimating Pb concentration in forest soils. It was found that applying a robust digital soil mapping (DSM) model in combination with terrain attributes and soil nutrients is efficient in predicting the spatial distribution and estimation of uncertainty levels of lead (Pb) in forest soils based on the model's accuracy parameters.

Keywords Regression kriging · Soil nutrient · Terrain attributes · Uncertainty assessment

Intoduction

Digital soil mapping (DSM) is the application of mathematical and statistical models to combine soil measurements with spatially explicit environmental variables to generate soil spatial information (van der Westhuizen et al. 2023; Takoutsing and Heuvelink 2022; Sanchez et al. 2009). One of the benefits of DSM resides in its capacity to resolve the uncertainty of the final products (Arrouays et al. 2020). Geostatistical based methods such as regression kriging (RK) have been for several years the major DSM Approach. The main robustness of RK lies in its capacity to account for spatial correlation, benefit from relationships with terrain attributes (i.e., explanatory environmental variables), reduce the prediction error variance and give prediction uncertainties through the kriging standard deviation (Hengl et al. 2015; Hengl et al. 2007; Odeh et al. 1995). The presupposition of RK is hinge on a linear mathematical relationship between soil nutrients and terrain attributes (Wadoux et al. 2018).

✉ Samuel Kudjo Ahado
ahados@af.czu.cz

1 Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, 16500 Prague, Czech Republic

2 Institute for Environmental Studies and SOWA RI, Faculty of Science, Charles University, Benátská 2, 12800 Praha, Czech Republic

3 Faculty of Natural Sciences, Matej Bel University Banská Bystrica, Tajovského 40, 974 01 Banská Bystrica, Slovakia

4 Department of Environmental Management, School of Environmental Sciences, Federal university of Technology, Owerri, Nigeria

However, in reality soil properties are often governed by complex relationships with soil forming factors (Behrens et al. 2014) and landscape attributes (Takoutsing et al. 2018) that compel them non-linear and entail complex relationship between terrain attributes (Lamichhane et al. 2019).

The application of machine learning algorithms (MLAs) approaches for spatial prediction of soil parameters has increased in recent years (Hengl et al. 2018). MLAs can handle a high number of cross-correlated covariates as predictors and may depict complex nonlinear interactions between dependent variables and covariates (Nussbaum et al. 2018). Furthermore, unlike geostatistical approaches, which frequently need transformation of the original data to satisfy the normality assumption (Wadoux et al. 2020), MLAs make no assumptions about the underlying distribution of the data. Because ML algorithms do not have to satisfy stringent statistical assumptions, they are more easily implemented. Furthermore, because of their adaptability to the nature of the connection between the soil nutrients of interest and terrain attributes, they are frequently more accurate than regression kriging (Hengl et al. 2015). While there are numerous MLAs, one that has attracted a lot of attention particularly for regression purposes is the random forest (RF) algorithm (Breiman 2001). Several studies (Hengl et al. 2015, Hengl et al. 2018, Nussbaum et al. 2018, Vaysse and Lagacherie 2017) have shown that RF is an efficient and valuable DSM technique for soil spatial prediction.

However, MLAs are not as excellent as kriging in quantifying prediction uncertainty, and the success of a DSM model must also be rated on its capacity to quantify prediction uncertainty reliably and realistically (Heuvelink, 2018; Arrouays et al. 2014). In this backdrop, Meinshausen (2006) improved the RF technique to create quantile regression forests (QRF), which yield all quantiles of the conditional distribution. As a result, it can quantify prediction uncertainty at all prediction locations, such as the upper and lower bounds of a 90% prediction interval. It does not quantify the spatial correlation of prediction uncertainty since it does not model spatial dependency (Heuvelink and Webster 2022). As a result, it is unable to quantify the prediction uncertainty of spatial averages and totals. A geostatistical technique is necessary to do this (Szatmári et al. 2021). Previous studies have employed RF and QRF and other MLAs to successfully forecast various soil parameters (Forkuor et al. 2017, Hengl et al. 2015, Hengl et al. 2021).

Several studies have examined RK and RF performance and found significant variations in prediction maps. Fouedjio (2020) discovered that RF can make forecasts that are as accurate and unbiased as RK. Hengl et al. (2015) proved that the RF algorithm outperformed RK across a wide range of soil parameters. Veronesi and Schillaci (2019) found that kriging outperformed RF in a study conducted on the semiarid island of Sicily in Italy. However, only a few research (Szatmári and

Pásztor 2019, Vaysse and Lagacherie 2017) have been conducted to compare RK and RF and other MLAs prediction uncertainty maps and determine if these maps are genuine representations of the prediction uncertainty. This necessitates additional case studies to gain a better understanding of RK ,QRF, Cubist, and Stochastic gradient boosting capacity to predict lead (Pb) and quantify uncertainty in forest soil.

The development of soil information at the global, national and regional levels is frequently hindered by low data availability, and many algorithms rely on extrapolation to obtain the essential information. In contrast to spatial interpolation, which uses point values in a study area to predict values at other points enclosed by observation points and located within that same area, spatial extrapolation refers to transferring the model beyond the area from which the training data were taken (i.e., to new geographic space). Specifically, whereas spatial interpolation predicts using data from all geographic directions, spatial extrapolation predicts using data from one or a few directions only, possibly from a long distance away. Due to changes in the kind and intensity of soil-forming variables, extrapolating a model from one place to another presents various issues (Angelini et al. 2020). Therefore, a detailed understanding of soil nutrients is critical to the decision-making process to assess lead contamination of forest soils. However, often the data is either unavailable or may not be well-structured, comprehensive, or spatially robust. However, collecting large numbers of field samples, processing and testing them in the laboratory is expensive and time-consuming. A viable alternative, digital soil mapping (DSM), uses soil and readily available environmental factors to spatially predict soil parameters (e.g., McBratney et al. 2003). DSM employs a variety of computational methods, including Machine learning algorithms that can make accurate predictions of soil properties, but they typically use point or raster data and require many environmental inputs (predictors). The difficulty in communicating the imprecision (uncertainty) associated with forecasting is a practical limitation of these methods for decision-making, despite the increasing availability of supporting data. Therefore, the objectives of the study were to: determine whether a combination of soil nutrients and terrain attributes can influence predictions of lead (Pb) in forest soils; determine whether combining terrain attributes and soil micronutrients in the estimation of Pb in forest soil will improve prediction accuracy and ascertain the level of uncertainty that will be associated in both contexts.

Materials and methods

Research area and sampling design

The Frydek Mistek district has a total area of 1208 km2 and is made up of 39 km2 of agricultural land and 49 km2 of
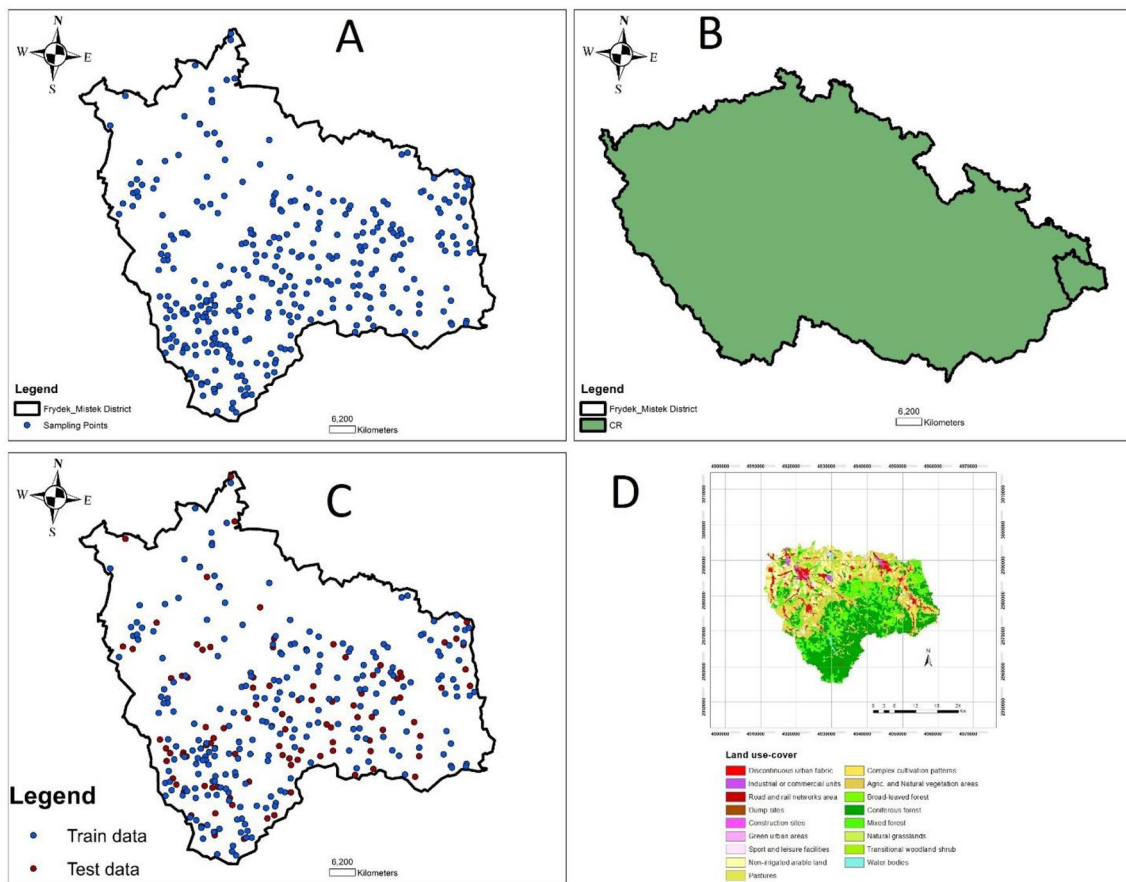
Fig. 1 **A** Study area map with sampling locations, **B** Frydek Mistek district, **C** Research location coupled with the partitioned dataset employed (i.e., training and test), **D** Sampling locations and land use–land cover of the Frýdek-Místek district with coordinates

forest land. The latitude and longitude of the study area are respectively 49° 41' 0" North and 18° 20' 0" East. It undulates from a height of 225 to 327 m above sea level. Due to its chilly, temperate climate, the study area experienced significant rainfall even in the dry months. The summers in Frydek Mistek are hot and partly rainy, while the winters are typically icy, dry, and windy Fig. 1. Temperatures range from 8 to 86 degrees throughout the year, rarely dropping below or rising above 24 degrees, and frequently fluctuating between these two ranges (Spark 2016). Precipitation ranges from 685 to 752 millimetres annually on average. The geomorphology of the studied area reveals rugged terrain with the highest peaks believed to be part of the Moravian-Silesian Beskydy and Outer Carpathians Mountains. 1208 km2 is the approximate survey area for the district, with lower terrain in the northern region. Because of the nearby steel and metal industries, as well as a large amount of farmland, Frydek Mistek is considered a relatively polluted area. Although the Czech Republic has a wide variety of soil types, Cambisols are the most common (Kozák et al., 2010). Fluvisols and stagnosols are two additional minor soil types. The variety of materials used in soil substrates reflects

the region's complex geological structure, with sediments from a variety of solid rocks forming gradient sediments at higher elevations and sediments from aeolian, fluvial, and lacustrine sources at lower elevations (Kozák et al. 2010). Nevertheless, the most common soil types in research environments are Cambisols and stagnosols (Kozák et al., 2010). These soils are found throughout the Czech Republic at elevations between 160.6 and 455.1 m for retentive soils and 59.6–493.5 m for Cambisols Vacek et al. (2020).

Data selection and preprocessing

The legacy database of forest soil properties comprises thousands of institutional samples collected throughout the Czech Republic between 1998 and 2019, as well as topographic attributes of the sampled forests (Borůvka et al. 2020). The depth of topsoil samples used for legacy data was 0–30 cm with a total number of 337 sampling points. Environmental covariates and pedological data derived from forest sampled stands within the forest indicate the forest type (broadleaf, coniferous, mixed forest) and other local land uses and land covers.

The contents of Pb, Cd, As, Cr, Cr, Cu, Mn, Zn, Fe and other elements in the soil were obtained by the standard aqua regia method (ISO 11466:1995, 1995) (Melo et al. 2016). Their pseudo-total content was determined using inductively coupled plasma optical emission spectroscopy (ICP-OES). Standard calibration methods were used to ensure method quality (QC). Standard material SRM 2711 (Montana II soil) (National Institute of Standards and Technology, Gaithersburg, MD, USA) was used to ensure quality control of concentration determinations.

Geostatistics

Geostatistics is a statistical discipline that examines and predicts the values related to spatial heterogeneity of physical processes. This includes the spatial and temporal coordinates of the datasets in the assessments. Several geostatistical approaches have been generated as a prudent approach to distinguish spatial features and linear interpolation values for areas in which samples have not been collected. In this research, Regression kriging (RK) was used to hybridize with machine learning approaches for mapping and predicting lead (Pb) contents in soils. Ordinary kriging is an interpolation method that enables the user to quantify the spatial variability of soil properties at the investigational site (van der Westhuizen et al. 2023; John et al. 2021; Bishop and Geoderma 2001). The equation is as follows:

$$Z^*(X_0) = \sum_{i}^{n} \lambda_i ZX_I \tag{1}$$

where $Z^*(X_0) =$ represents the predicted value at the unquantified location (X_0), $Z(X_I)$ denotes the known or observed value at the location (X_I), λ_i is the coefficient weighting at the observed location (X_O) and n is the number of locations within the area under observation.

Regression kriging (RK) is a form of interpolation technique where there is a mixture of linear models of variables that are dependent and auxiliary variables, such as terrain attributes of variables in which the residuals are kriged alongside (Agyeman et al. 2022b; Odeh et al. 1995). The RK approach was used in this research to spatially interpolate the distribution of lead (Pb) in the following order to: evaluate the lead (Pb) prediction technique approach by utilizing the regression technique in reciprocal directions, quantifying the lead (Pb) prediction modeling approach with residuals at every calibration position, modeling the covariance structure of the lead (Pb) residuals and spatially interpolating the lead (Pb)residuals using the variogram model parameters and obtaining the predicted map and combining the lead (Pb) prediction approach surface on the interpolated residual surface.

Introduction of machine learning algorithms (MLAs)

Machine learning algorithms (MLA) used include Cubist, Stochastic gradient boosting, Quantile regression forest, and K-Nearest-Neighbor. Based on these MLAs, each dataset is randomly divided into two parts: a test set (25%) and a training set (75%). The training set generates regression models that show the relationship between the response variable (i.e., Lead) and the predictor variables (i.e., Terrain attributes and soil nutrients data), while the test data evaluates the performance of each model. A description of the model and packages used by the R software is given below.

Cubist (CUB)

The Cubist algorithm (Rulequest 2016a,b) is one of the rule-based algorithms used to build predictive models based on the analysis of input information. See5/C5.0 can solve classification problems (Quinlan 2004), and Cubist can solve regression problems very well. Cubism model results are preferred over linear regression model results. Also, it is simpler than the ANN model (Rulequest 2016a, b). The Cubist model is a tree extension of Quinlan's M5 model (Quinlan 1992) with the ability to use thousands of input features (Rulequest 2016a, b). In the Cubist model, the goal depends on the input and is calculated according to the rules. These rules are combined with various conditions and linear functions. If the rule considers the general requirements to evaluate the result correctly, the appropriate linear function is used. The cubist algorithm can run multiple instances at once and then test different linear functions to evaluate the output. Therefore, Cubist can generate different models and mix them according to the rules. This was determined before the development of several models with different rules, the combination of which could help the Cubist model achieve a higher level of accuracy. More details on Cubist can be found in Ref. (Nguyen et al. 2019). In this study, the cubist was performed in R using the caret package cubist.

Stochastic gradient boosting (SGB)

Stochastic gradient boosting is associated with mixture boosting and bagging. Many microscale categorization or regression trees are progressively generated from pseudo-residuals (the gradient of the loss function of the previous tree) (Friedman 2001, 2002). At each iteration, a tree is built from a random subsample of the data (selected without substitution), which causes the model to progress incrementally. Using only part of the training data increases computational efficiency and prediction accuracy while trying to avoid data overfitting. An advantage of stochastic gradient enhancement is that no predictor variable needs to be pre-selected or transformed. It is also resistant to outliers, as the steepest

gradient method highlights points related to their correct categorization. In this study, the stochastic gradient boosting was performed in R using the caret package gbm.

Quantile regression forest (QRF)

QRF is a variant of the RF algorithm (Meinshausen and Ridgeway 2006). It records the mean and spread of all observed samples at each node of the decision tree. It uses this information to test the conditional probability of the expected outcome (Dharumarajan et al. 2019). In this study, the Quantile Regression Forest was performed in R using the caret package qrf.

K-nearest-neighbor (KNN)

K-Nearest Neighbors is a popular data mining strategy for solving classification problems (Moreno et al. 2003). It is an excellent choice due to its flexibility and relatively fast computational efficiency. However, the main disadvantage of the K-Nearest Neigbour classifier is that it requires a lot of memory to store the entire sample. If the sample size is large, the response time of a sequential computer is also high (Alpaydin 1997). Despite the memory requirement, it performs well in classification tasks on various datasets (Liang and Li 2009). In this study, the K-Nearest-Neighbor was performed in R using the caret package KNN.

Terrain attributes and soil properties

By integrating several topographic variables, Pb is associated with topographic features. Environmental covariates were extracted from the Aster dataset using a digital elevation model (DEM) at a spatial resolution of 30 m (http://earthdata.nasa.gov/search/) and processed for terrain analysis using the SAGA-GIS toolkit. In any case, a 30 m DEM with a spatial resolution of 30 m is resampled to a spatial resolution of 20 m using the bilinear resampling method in ArcGIS. The topographic parameters used in this study are slope, elevation, valley depth, LS factor, aspect and relative slope location. Terrain attributes were chosen because of their association with the response variable.

Soil condition, based on several physico-chemical and biological properties and their nutrient content, is crucial in agricultural production. Assessing and improving soil sustainability and health, which can contribute to sustainable agricultural systems, has recently received increased attention worldwide (Monther et al. 2020). Plant development and nutrient cycling in both terrestrial and aquatic environments are controlled by phosphorus (P) and nitrogen (N), which are an essential component of the biomacromolecules of all living organisms (Vrede et al. 2004; Sista and Schimel 2012). Therefore, changes that alter N:P stoichiometry can lead to changes in species diversity and vegetation composition, which in turn affect ecosystem functioning (Vitousek et al. 2010; Yuan and Chen 2015). By identifying regional and temporal trends in N and P stoichiometry, we can better understand the dynamics of nutrient cycling on lead distribution in terrestrial ecosystems and their potential impact on ecosystem functioning in a rapidly changing environment. The elements such as potassium (K), calcium (Ca), copper (Cu), iron (Fe), and phosphorus (P) were selected based on their association with the response variable and their conflicting and reinforcing effects on the response variable.

Assessment of the accuracy and validation of the models

In order to determine the prediction accuracy of models, the dataset was split into calibration (75%) and validation (25%) datasets. The calibration dataset is applied to establish the prediction models, and validation dataset is adopted for independently evaluating the prediction accuracy. The coefficient of determination (R^2), root mean square error (RSME), mean error (MAE), RPIQ (ratio of performance to interquartile range) and median absolute error (MdAE) were Used to evaluate model efficiency and accuracy of the methods applied in this research. R2 describes the variability in response proportions represented by the regression model. MdAE confirms the true measured value, while RMSE and MAE determine the magnitude of multiple versions in a given measurement, allowing estimates to approach expected accuracy. The value of R2 should be high and the accuracy should be near to 1 to select the best model method using the validation criteria. According to Li et al.(2016) A standard R2 value of 0.75 should be considered excellent predictions, while 0.50 to 0.75 should be considered satisfactory predictions. The lower obtained values are suitable and considered the best choice for model selection for the validated standard estimation methods RMSE, MdAE and MAE. The RPIQ is calculated by dividing the interquartile range (IQ = Q3–Q1) by the RMSE and depicts the largest spread of population residuals (shi et al. 2023; Bellon-Maurel et al. 2010).

To analyze the uncertainty of digital soil mapping, a bootstrap approach was utilized (Wu et al. 2023; Zeraatpisheh et al.2022). The sample dataset is randomly divided into calibration and validation datasets for N times, for example, 200 times. As a consequence, N spatial distribution maps are generated. The uncertainty map is the difference between the 95th and 5th percentiles of the N maps, then divided by the mean value map.

$$\text{Uncertainty} = \left(CI_{95\%} - CI_{5\%} \right) \big/ Mean \qquad (2)$$

where $CI_{95\%}$ and $CI_{5\%}$ are the upper and lower of the 95% and 5% confidence intervals of the predicted values, and Mean is the mean predicted value of N maps.

Table 1 The summary statistics of lead (Pb) and soil nutrients

	Pb	P	K	Mg	Ca	Cu	Fe
Mean (mg/kg)	54.72	97.82	235.42	189.98	322.17	5.33	5805.55
Geometric mean (mg/kg)	44.18	67.26	214.57	151.28	167.58	4.25	5074.65
Median (mg/kg)	45.75	69.50	218.71	166	130.87	4.20	5307.00
Minimum (mg/kg)	4.00	5.00	67.39	10.00	15.40	0.60	101.00
Maximum (mg/kg)	312.3	611	1389.6	1190	8386.33	101	20760.43
Lower quartile	28.64	37.52	168	93.57	88.17	2.98	4134.27
upper quartile	72.84	121.67	280.76	253.36	252.57	6.15	7114.3
10 percentiles	18.57	21.85	123.25	58.62	65.5	1.96	2860
90 percentiles	101	199.38	342	345.1	627.31	8.88	9102.08
Std.Dev	37.33	95.02	119.41	131.66	677.28	6.38	2785.96
Coef. Var	68.22	97.14	50.72	69.3	210.22	119.6	47.99
Skewness	1.87	2.45	3.93	2.37	7	10.9	1.30
Kurtosis	6.84	7.76	30.31	12.59	66.97	154.71	3.92

Results and discussion

Description of data and variable relative importance of soil nutrients and terrain attributes to Pb

Statistical details of Pb levels and selected environmental covariates used in this study are shown in Table 1. Maximum and minimum values for PTE and environmental covariates ranged from 4.00 to 312.30 (mg/kg), Pb from 5.00 to 611 P (mg). /kg), K from 67.39 to 1389.60 (mg/kg), Mg from 10.00 to 1190 (mg/kg), Ca from 15.40 to 8386.33 (mg/kg), 0.60 to 101.00 (mg/kg) Cu and 101. 20760.43 (mg/kg) Fe. The lower and upper quartiles of the data set range from 28.54 to 72.84 Pb, 37.52 to 121.67 P, 168 to 280.76 K, 93.57 to 253.36 Mg, 88.17 to 252.57 for Ca, 2.98 to 6.15 for Cu and 4134.27 to 7114.3 for Fe. The standard deviation values of Pb and soil nutrients were relatively high, while Cu was below 6.50. This is due to the increased heterogeneity of variables in the study area (Agyeman et al. 2021). The normality of the data based on the estimated skewness indicates that the data does not follow a normal distribution because skewness values exceed 1 to -1. Additionally, the kurtosis and skewness values in Table 1 indicate that the distribution of Pb and soil nutrients is not balanced in the right direction. The coefficient of variation (CV) of Pb indicated that the calculated CV of Pb (68.22%) in the study area was mainly from anthropogenic sources. According to Karimi Nezhad et al. (2015) for CVs greater than 20%, it implies that lead variability is likely to be primarily from anthropogenic sources.

Prediction applying the three diverse contexts

Prediction of Pb levels in forest soils was performed using three different methods, namely application of Pb in forest soils using terrain attributes with regression kriging

(Context 1), prediction of Pb in forest soils using coupled soil nutrients with regression kriging (Context 2) and prediction of Pb in forest soil using soil nutrients in combination with coupled terrain attributes to regression kriging (context 3). In context 1 (see Table 1), the prediction of the Pb content of forest soils suggests that QRF_RK (Quantile Regression Forest-Regression Kriging) 0.63, 25.71, 18.26, 12.79 and 0.79 for R2, RMSE, MAE, MdAE and RPIQ received respectively. With an R2 value of 0.56, RMSE of 27.46, MAE of 20.69, MdAE of 16.01 and RPIQ of 0.74 have reported that KNN_RK (k-Nearest Neighbors regression kriging) predicted the Pb content of forest soils. In contrast, the other modeling approaches yielded R2, RMSE, MAE, MdAE, and RPIQ values of 0.61, 26.25, 19.16, 15.05, and 0.99 for CUBIST_RK (cubist_regression kriging) and 0, 62, 26.25, 20.37, 15.10 and 0.93 for SGB_RK (Stochastic Gradient Boosting-Regression Kriging). The application of QRF_RK was the best approach in the prediction of Pb in forest soil (Table 2).

In context 2 (see Table 3), forest soil Pb prediction was modeled using soil nutrient and regression kriging models. The results indicate that the use of QRF_RK in combination with soil nutrients resulted in R2, RMSE, MAE, MdAE and RPIQ values of 0.74, 22.72, 15.72, 9.53 and 1. 78 respectively. Other modeling techniques, such as KNN_RK and CUBIST_RK, have yielded R2, RMSE, MAE, MdAE and RPIQ values of 0.69, 24.22, 17.53, 12.32 and 1.24 for KNN_RK and 0, 77, 21 0.81, 14.37, 9.67 and 1.66 for CUBIST_RK. The SGB_RK modeling approach, in combination with soil nutrients as a supplementary dataset, predicted forest soil Pb with the following results: 0.78, 21.30, 14.73, 10.52 and 1.91 for R2, RMSE, MAE, MdAE and RPIQ respectively. Cumulative evaluation of the results in context 2 of the modeling approaches indicated that SGB_RK in combination with soil nutrients gave the best results in Pb prediction in forest soil.

Table 2 Prediction of lead in forest using terrain attributes (Context 1)

	Terrain attributes				
	R^2	RMSE	MAE	MdAE	RPIQ
QRF_RK	0.63	25.71	18.26	12.79	0.79
KNN_RK	0.56	27.46	20.69	16.01	0.84
CUBIST_RK	0.61	26.25	19.16	15.05	0.99
SGB_RK	0.62	26.25	20.37	15.10	0.93

Stochastic Gradient Boosting-regression kriging (SGB-RK), cubist regression kriging (CUB_RK), quantile regression forest kriging and k nearest neighbour regression kriging (KNN_RK) and quantile regression forest (QRF_RK)

Table 3 Prediction of lead (Pb) in forest using soil nutrients (Context 2)

	Soil nutrients				
	R^2	RMSE	MAE	MdAE	RPIQ
QRF_RK	0.74	22.72	15.72	9.53	1.78
KNN_RK	0.69	24.22	17.53	12.32	1.24
CUBIST_RK	0.77	21.81	14.37	9.67	1.66
SGB_RK	0.78	21.30	14.73	10.52	1.91

Stochastic Gradient Boosting-regression kriging (SGB-RK), cubist regression kriging (CUB_RK), quantile regression forest kriging and k nearest neighbour regression kriging (KNN_RK) and quantile regression forest (QRF_RK)

Modeling results for predicting Pb in forest soil in context 3 (see Table 3) showed that QRF_RK and KNN_RK gave the following results: 0.77, 21.19, 15.56, 12.32, and 1.88 for R2, RMSE, MAE, MdAE, and RPIQ in QRF_RK and 0.65, 25.15, 18.20, 12.99, and 0.98 for R2, RMSE, MAE, MdAE, and RPIQ in KNN_RK. CUBIST_RK and SGB_RK, on the other hand, achieved the following R2, RMSE, MAE, MdAE, and RPIQ results: 0.82, 19.04, 12.74, 8.04, and 1.96 for CUBIST_RK; 0.77, 21.84, 15.10, 11.26, and 1.83 for SGB_RK. An overall evaluation of the modeling approach in context 3 of Pb prediction in forest soil shows that the best results are obtained using CUBIST_RK in combination with terrain attributes and soil nutrients.

Comparing contexts 3 and 1, the R2 values of QRF_RK, KNN_RK, CUBIST_RK, and SGB_RK in context 3 increased by 22.31%, 16.51%, 33.48%, and 24.71% compared to context 1, respectively. Compared with Context 1, the RPIQ values of QRF_RK, KNN_RK, CUBIST_RK, and SGB_RK in Context 3 increased by 139.49%, 17.18%, 97.55%, and 96.90%, respectively. The estimated errors of the modeling approaches QRF_RK, KNN_RK, CUBIST_RK, and SGB_RK for RMSE, MAE, and MdAE decreased by 17.57%, 14.74%, and 3.69% (QRF_RK), 8.41%, 33.74%, and 18.84% (KNN_RK), 8.41%, 33.74%, and 18.84%

(CUBIST_RK), and 16.80%, 25.88%, and 25.40% (SGB_RK) in context 3.

Comparing contexts 2 and 3, the following modeling methods QRF_RK and CUBIST_RK achieved an increase in R2 values by 4.43% and 6.67%, respectively, in context 3 in predicting Pb in forest soil. In contrast, the following modeling methods KNN_RK and SGB_RK increased R2 values by 6.47% and 1.46% in context 2 compared to context 3, respectively. More specifically, the estimated RPIQ values for the modeling methods in Contexts 2 and 3 show that the RPIQ values of QRF_RK and CUBIST_RK in Context 3 increased by 5.54% and 17.76%, respectively, compared to Context 2. Alternatively, the RPIQ values of KNN_RK and SGB_RK increased by 20.83% and 4.09% in Context 2 compared to Context 3. The quantified errors of the modeling approaches used in the prediction of Pb in forest soil revealed that RMSE, MAE, and MdAE realized an error margin reduction of 6.74%, 1.00%, and 29.26% for QRF_RK and 12.71%, 11.34%, and 16.90% for CUBIST_RK in context 3 compared to context 2. Similarly, the estimation errors (e.g., RMSE, MAE, and MdAE) of KNN_RK and SGB_RK in context 2 are reduced by 3.84%, 3.82%, and 5.48% for KNN_RK and 2.55, 2.50, and 7.02 for SGB_RK compared to context 3.

The spatial variability of Pb was assessed by applying the semivariance portraying the quantified indexes that were predicted using diverse regression kriging models in the prediction of Pb in the soil using the gaussian semivariogram model alongside with its nuggets effect. The higher ratio indicates that stochastic factors, such as litter layers from trees, farming measures, lumbering, forest plantations, woody organic residues from deep roots, and affiliated soil microbial and fauna populations, as well as other human activities, are primarily responsible for the spatial variability of Pb in forest soil. The lower ratio indicates that structural factors such as climatic conditions, parental material, elevation, soil properties, and other natural factors play an important role in spatial variability. According to Cambardella et al. (1994), a nugget sill ratio of 0.25 or less represents strong spatial variability, 0.25 to 0.75 represents moderate spatial variability and anything above 0.75 suggest a weak spatial variability. Based on the regression kriging approaches used in this study the nugget sill ration is below 0.25 showing a high spatial variability but weak spatial randomness (see Table 4, 5).

The overall evaluation of the prediction methods in contexts 1, 2 and 3 showed that the prediction of lead in forest soil using terrain attributes as an auxiliary data set gave satisfactory results, but with a correspondingly high error rate. However, using soil nutrients as an auxiliary data set to predict Pb in forest soils gave good and satisfactory results. Nevertheless, the use of terrain attributes and soil nutrients as auxiliary datasets for Pb prediction in forest

Table 4 Prediction of lead (Pb) in forest using a combination of soil nutrients and terrain attributes (Context 3)

	Terrain attributes + soil nutrients				
	R^2	RMSE	MAE	MdAE	RPIQ
QRF_RK	0.77	21.19	15.56	12.32	1.88
KNN_RK	0.65	25.15	18.20	12.99	0.98
CUBIST_RK	0.82	19.04	12.74	8.04	1.96
SGB_RK	0.77	21.84	15.10	11.26	1.83

Stochastic Gradient Boosting-regression kriging (SGB-RK), cubist regression kriging (CUB_RK), quantile regression forest kriging and k nearest neighbour regression kriging (KNN_RK) and quantile regression forest (QRF_RK)

soils increased the predictive output of all modeling methods except KNN_RK. Based on the three modeling environments, the best modeling approach for predicting forest soil Pb content is to use CUBIST_RK in combination with soil nutrients and terrain attributes, resulting in the best prediction performance with the lowest error. The modeling approach used in this study showed that predicting soil lead using soil nutrients as an auxiliary data set performed better than using terrain attributes. Therefore, it can be concluded that the combination of soil nutrients and terrain attributes has a positive effect on terrain attributes in predicting Pb in forest soils. Although Agyeman et al. (2022b) reported that geological topography is an influential factor in predicting soil PTE, and the results suggest that interactions between soil nutrients and PTE are also important in predicting soil PTE. Cubist_RK has been applied in various studies to predict soil properties and PTE. Pouladi et al. (2019) used Cubist_RK together with Random Forest_RK, and according to the output results, Cubist_RK performed better than Random Forest_RK. According to Ma et al. (2017), when residuals were incorporated into a cubist model using kriging, soil organic carbon prediction was optimized compared to other modeling methods used. When combined with regular kriging, cubist models can be as effective as random forests (Lamichhane et al. 2019) for predicting soil properties and PTE. Cubist_RK outperformed RF_RK (Random Forest Regression Kriging) and CIF_RK (Conditional Inference Forest Regression Kriging) in predicting antimony levels in

agricultural soils when data fusion was combined with terrain attributes (Agyeman et al. 2022b). The integration of soil nutrients and terrain attributes together with cubist_RK is potent in predicting PTE's in forest soils with small errors and high detection rates. According to Hengl et al. (2004), Umali et al. (2012), a Zhang et al. (2012) using RK in the learning algorithm combined with spatial interpolation provides better spatial interpolation results when predicting soil properties and PTE.

The combination of soil nutrients and topographic features gave good results in predicting forest soil Pb content. The effect of soil nutrient inclusions on terrain features in forest soil Pb prediction gave good results. It should be noted that the antagonistic and stimulatory effects of soil nutrient interactions probably explain the best results in context 3. This is consistent with Agyeman et al. (2023), who used spectral reflectance to predict agricultural soil zinc concentrations and soil nutrients in the same area. Geomorphic topography has a significant impact on the assessment of PTEs such as Pb in soil, and the relationship between bedrock, climatic conditions and geomorphic processes can cause the formation of complex soil matrices (Agyeman et al. 2022a, 2022b). Pb changes in soil layers are not uniform and closely related to hydroxides (especially Fe). Because the geochemistry of $Pb2+$ is similar to that of divalent alkaline earth metals, Pb can replace K and sometimes Ca in both minerals and on sorbent surfaces (Kabata 2010). According to Hettiarachchi et al. (2001), P compounds are very potent in reducing the bioavailability of Pb. P rocks are more effective than P fertilizer. Hashimoto and Sato (2007) studied the lead adsorption capacity of different waste materials and found that hydroxyapatite with low crystallinity was the most effective. Pb immobilization in the soil is strongly influenced by P and some minerals, especially pyrozoites (Kumpiene et al. 2008). Pb-Ca interactions are important in metabolism because Pb can mimic the physiological behavior of Ca and thus inhibit certain enzymes (Kabata 2010). Interactions between soil nutrients and topographic properties have enormous implications for the impact and effective characterization of spatial lead heterogeneity in soils depending on the context of pedogenesis and developmental evolution (Zeraatpisheh et al. 2020).

Table 5 Semivariogram model and spatial dependance results of lead (Pb) prediction in forest soil

Model	Model	Range (m)	Sill ($C_o + C$)	Nugget (C_o)	Nugget/Sill
KNN_RK	Gaussian	1222.11	4.9594	0.0049	0.001
Cubist_RK	Gaussian	1280.68	484.895	0.856	0.0018
SGB_RK	Gaussian	1280.68	162.067	0.696	0.0043
QRF_RK	Gaussian	848.273	17.4154	0.221	0.0127

Quantile regression forest (QRF_RK), Stochastic Gradient Boosting-regression kriging (SGB-RK), cubist regression kriging (CUB_RK), quantile regression forest kriging and k nearest neighbour regression kriging (KNN_RK)

Changes in PTEs such as PTEs in forest soil are strongly influenced by several soil processes, such as pedogenesis, parent material weathering, or the ability of PTEs to bind organic matter (Kabata-Pendias 2011) and their presence in the soil in different forms. Thus, these processes allow accurate identification of lead concentration attributes, spatial distribution, and factors affecting forest soils. However, the impact process is associated with various changes, such as the speed of different topographic gradients, which affect the spatial variability of Pb in soil and their interrelationship with different elements (Jiang et al. 2019b).

Spatial prediction of the optimal regression kriging approaches per context

Pb prediction plots based on 6 environmental covariates and 6 soil nutrients were used as independent variables to fit QRK-RK, SGB-RK, CUBIST-RK and KNN-RK. Figure 4 shows the top 3 models for predicting Pb in forest soils, showing their regression kriging, residuals, and spatial distribution plots of the modeling method using both contexts. The spatial distribution map of the regression kriging method is regulated by soil nutrients as an auxiliary

covariate, which exhibits relatively high spatial autocorrelation. The machine learning predicted (ML) maps and regression kriging predicted maps of the best methods (TASN_CUBIST_RK, SN_SGB_RK, and TA_QRF_RK) have similar spatial distribution patterns from northwest to southeast and southwest of the study area. The maps produced by regression Kriging and the machine learning part of regression Kriging show no significant differences between the predicted maps, but they differ in high and low values of the prediction range. This means that using topographic attributes and soil nutrients alone or in combination did not make a significant difference in the predicted maps in forest soils, regardless of whether machine learning algorithms or hybrid models such as regression kriging were used. All ML predictions and RK predictions of the optimal model produce hotspots in the northeast and southeast regions of the research area. On the other hand, the TASN_CUBIST_RK and SN_SGB_RK predictions and the RK map show hotspots in the northeastern enclave of the research area. The distribution of Pb concentration in the forest soil in the sampled area was low to moderate over a large area, and patches of elevated Pb concentration were observed in some areas of the study area. The residual images of the
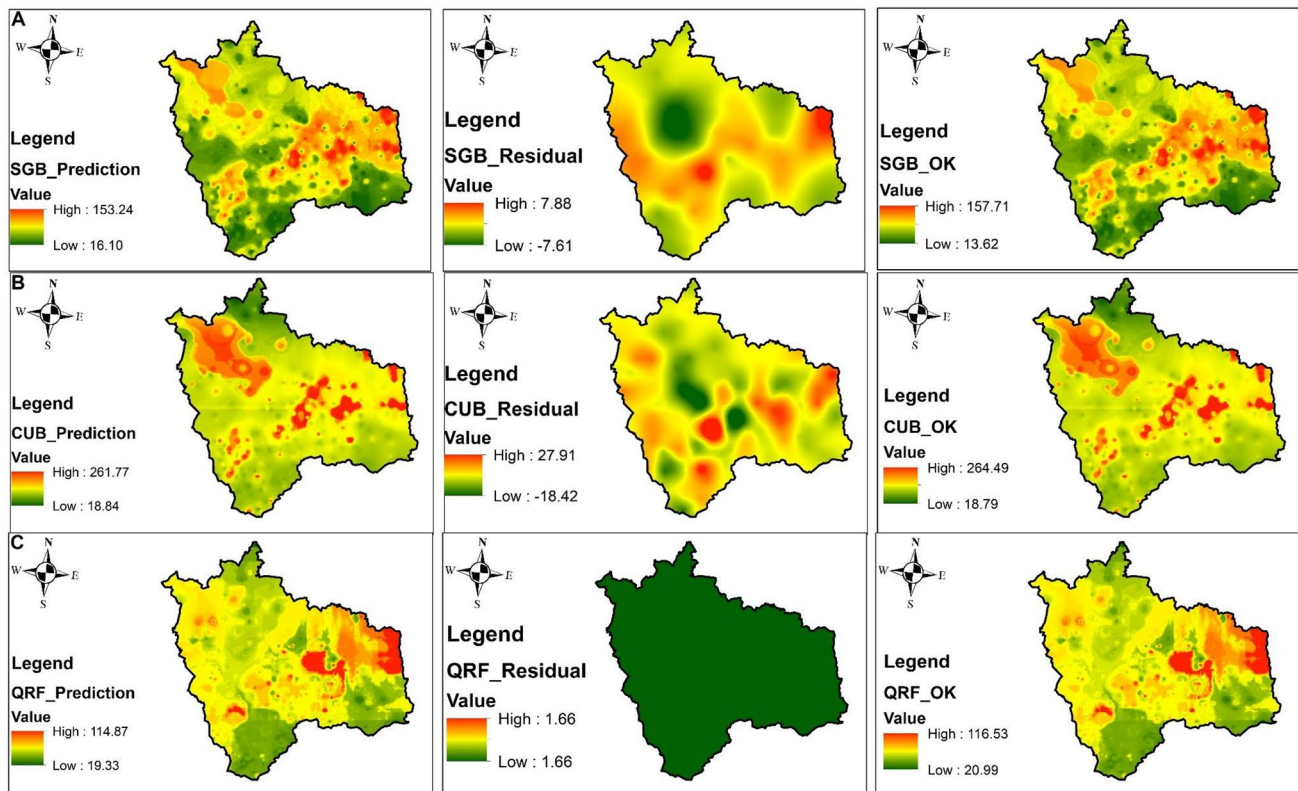
Fig. 2 Shows the predicted map of machine learning approaches, the residuals, and the regression kriging for the optimal modeling approaches in each context. Row **A** represents Stochastic Gradient Boosting-regression kriging, **B** cubist regression kriging and **C** quantile regression forest kriging

best maps show a mixture of low, medium, and high predictions, except for the QRF residuals, which show consistently low predictions across the study area. This may be because QRF was designed to combine the advantages of quantile regression (QR) and random forest (RF), thus splitting the residuals into low quantiles. Unlike ordinary regression, the effect of QRF regression assumes a line that best fits a given quantile of the estimated response variable (Meinshausen 2006). QR can be used to Investigate the expected relationships between variables is more useful when the affiliations between the variable means are very weak or non-existent (Brennan et al. 2015). The differences between the remaining maps produced with the TASN_CUBIST_RK, SN_SGB_RK and TA_QRF_RK maps (Fig. 2) may be due to overfitting caused by mapping artefacts. The use of additional datasets for modeling methods to map PTEs such as Pb in forest soils improves prediction accuracy. It is clear that the relationship between Pb and the auxiliary data is strong but using the TASN_CUBIST_RK and SN_SGB_RK maps provides a better control of the effects of map overfitting than TA_QRF_RK.

Uncertainty assessment of the modeling approaches based on each context

Figures 3, 4 and 5 show the estimated uncertainties for the modeling methods used in this study for each case. The uncertainty for each modeling method in each context in each graph is listed in all modeling methods in a columnal approach. In context 1, 5% confidence intervals (CIs) and 95% uncertainty spreads and mean predictions for Cubist_RK columns (A), KNN_RK columns (B), and SGB_RK columns (C) are plotted for the Northeast, Southeast, Center, Southwest, and Northwest territories. The modeling approach shows a mosaic distribution pattern of prevalence levels from low to high uncertainty (Cubist_RK column (A), KNN_RK column (B) and SGB_RK column (C). There are patches of high uncertainty in the central region of the map and in the eastern and western areas. The 95% CI of Cubist_RK shows a higher level of uncertainty than KNN_RK and SGB_RK. On the other hand, the QRF_RK modeling approach shows a low to moderate level of uncertainty in the lower limits, upper limits, and mean predictions. In context 2, the uncertainty distribution for all modeling methods
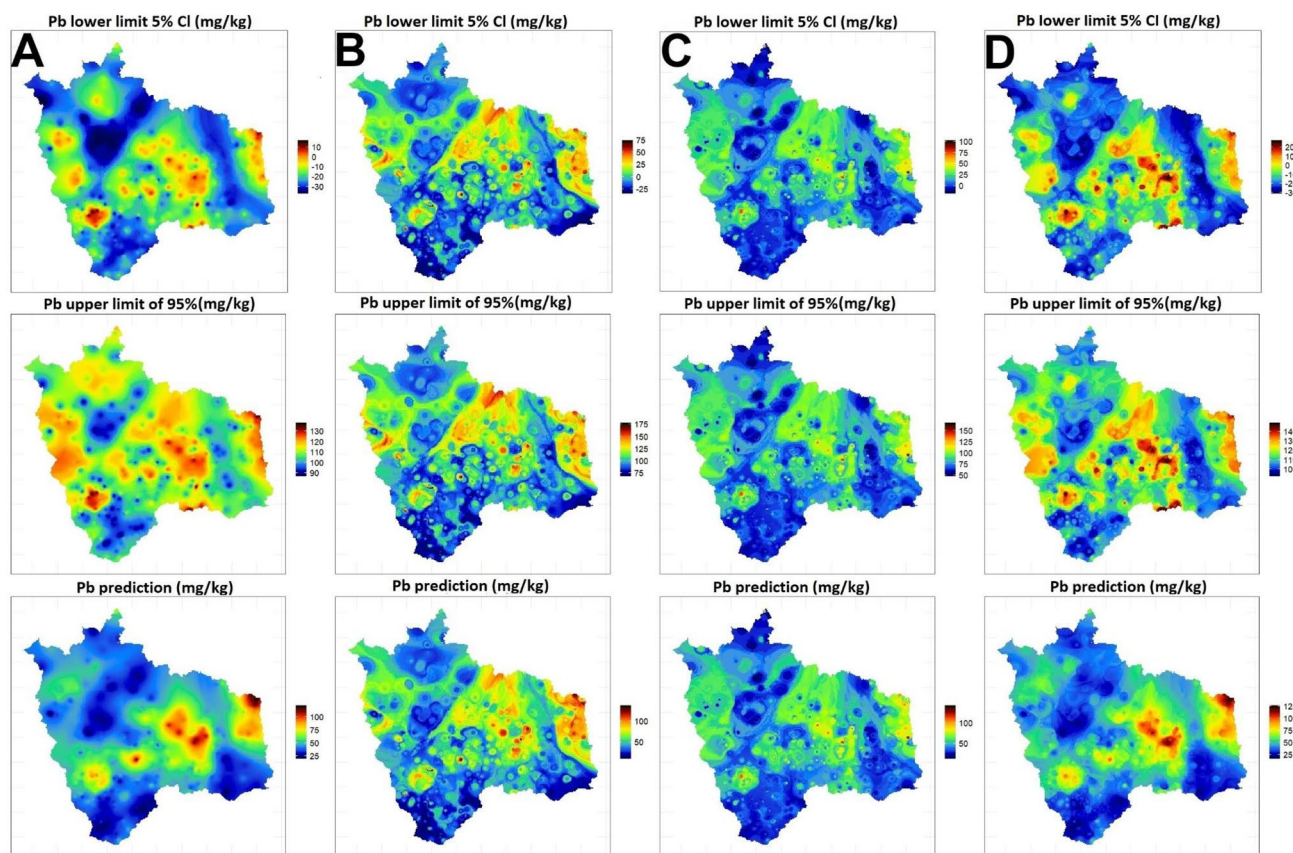
Fig. 3 Depicts an uncertainty prediction of Pb in forest soil using soil nutrients as an auxiliary dataset. For Cubist_RK column A, KNN_RK column B, QRF_RK column C, and SGB_RK column D, the 95% and 5% confidence interval predictions, as well as the mean prediction, are displayed in each column
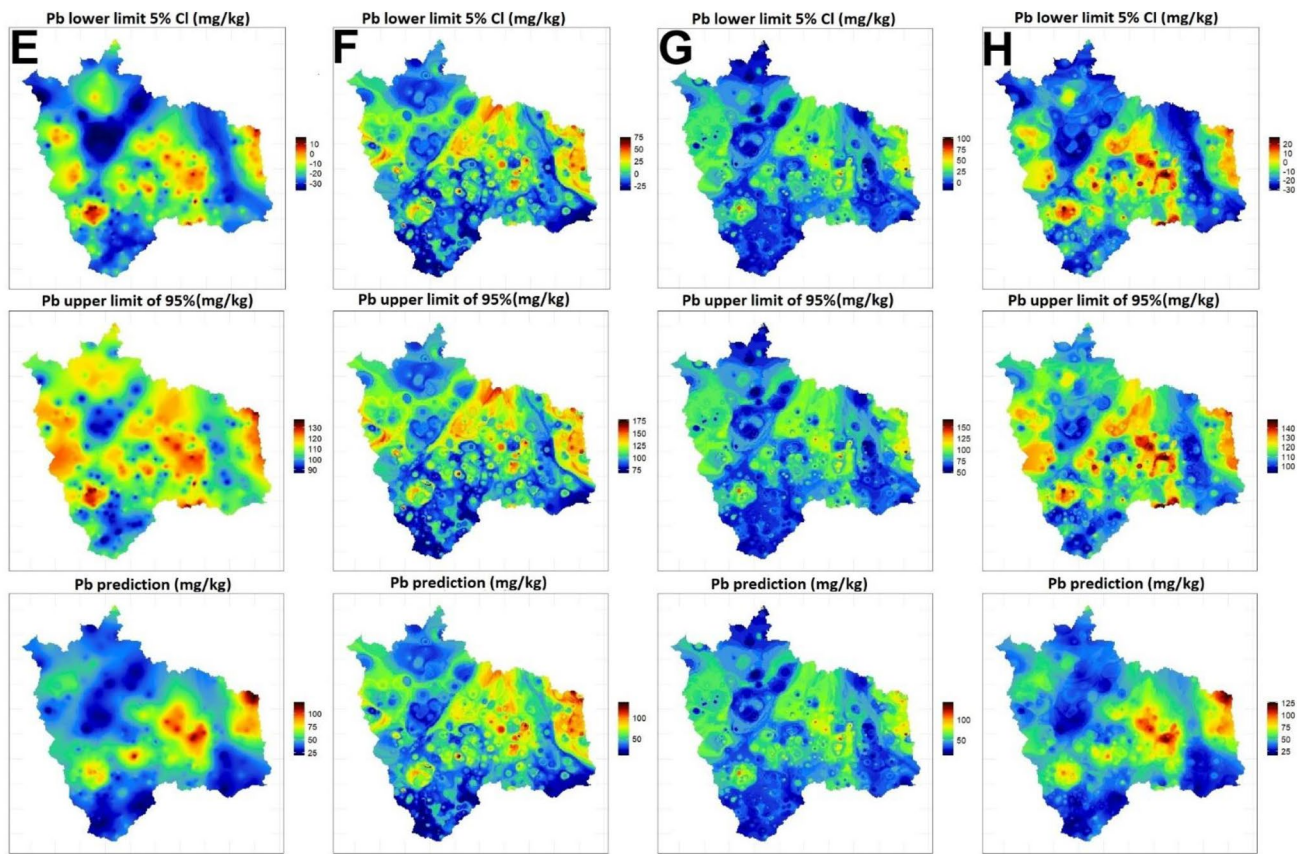
Fig. 4 Depicts an uncertainty prediction of Pb in forest soil using terrain attributes as an auxiliary dataset. For Cubist_RK column E, KNN_RK column F, QRF_RK column G, and SGB_RK column H, the 95% and 5% confidence interval predictions, as well as the mean prediction, are displayed in each column

is consistent with context 1. Applying uncertainty propagation to soil nutrients and topographic attributes did not reveal any significant differences in the level of uncertainty propagated using prediction intervals. This may be because soil nutrients and topographic features have geologic relationships with response variables. Long-term interactions between bedrock, climatic conditions, and geomorphological mechanisms lead to the formation of soil matrix composites (Agyeman et al. 2022b). In Context 3, all modeling methods have low to moderate uncertainty in the study area, with high patchiness in the center of the study area. The combination of soil nutrients and topographic features reduces the level of uncertainty in predicting soil lead levels. The relationship between soil nutrients, topographic properties and Pb affects the effective characterization of spatial heterogeneity in PTE, such as PTE in soil, depending on pedogenesis and evolutionary development, resulting in low uncertainty propagation in prediction intervals for all modeling methods. Figure 3 depicts an uncertainty prediction of Pb in forest soil using soil nutrients as an auxiliary dataset. For Cubist_RK column A, KNN_RK column B, QRF_RK column C, and SGB_RK column D, the 95% and 5% confidence interval

predictions, as well as the mean prediction, are displayed in each column.

Conclusion

This paper investigates the possibility of predicting lead (Pb) in forest soils by combining terrain attributes, soil nutrients and their combinations, and regression kriging in the Frydek Mistek district of the Czech Republic. The results showed that the prediction of Pb in forest soil using only terrain attributes and soil nutrients as additional data sets combined with regression kriging would produce satisfactory results. On the other hand, it performed well with the combination of soil nutrients and terrain attribute. Overall, the evaluation showed that using cubist_RK in combination with terrain attributes and soil nutrients provided the best prediction accuracy and the lowest error in Pb prediction in forest soils. Interactions between Pb and soil nutrients, as well as terrain attributes, can help to better identify sources of PTE pollution while improving predictive efficiency. Therefore, it is suggested that the use
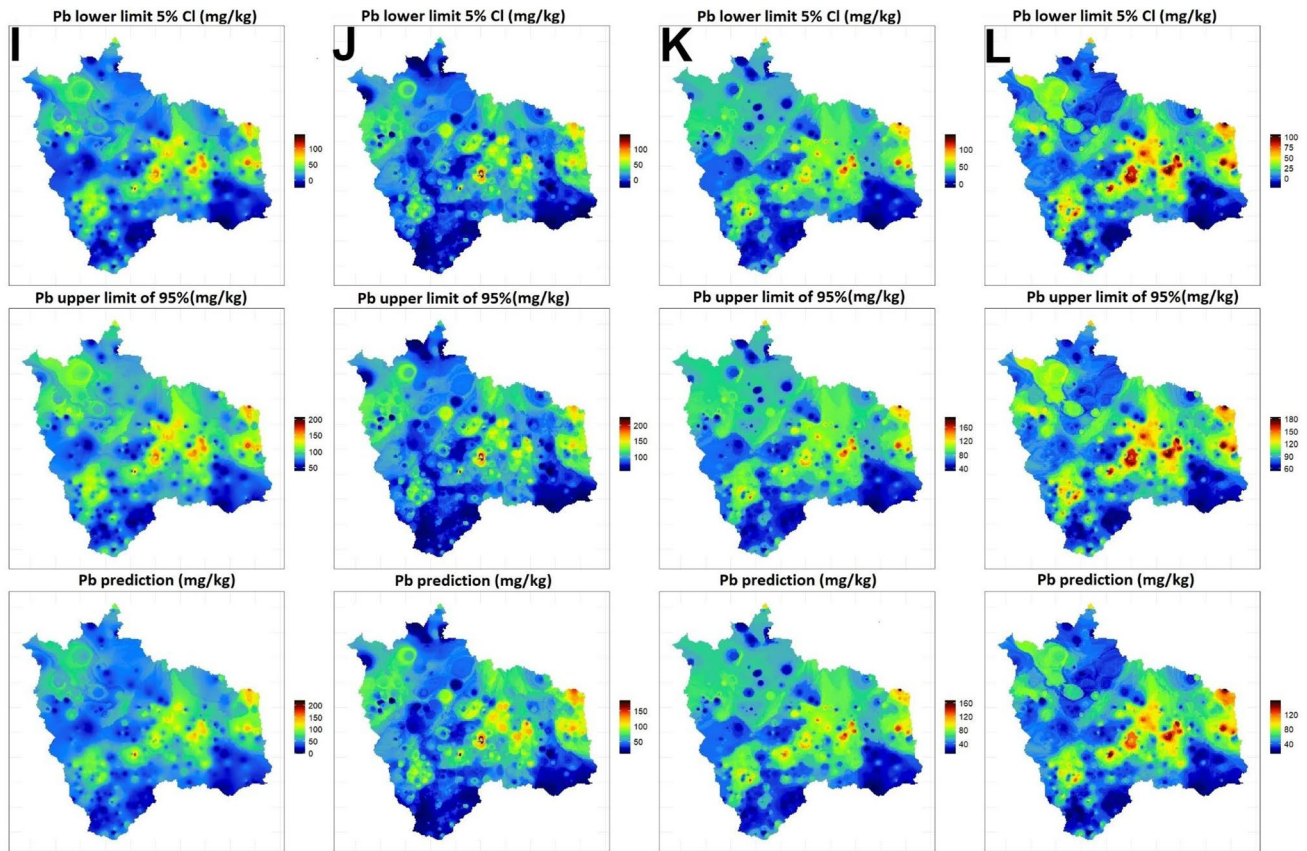
Fig. 5 Depicts an uncertainty prediction of Pb in forest soil using soil nutrients and terrain attributes as an auxiliary dataset. For Cubist_RK column I, KNN_RK column J, QRF_RK column K, and SGB_RK column L, the 95% and 5% confidence interval predictions, as well as the mean prediction, are displayed in each column

of soil chemical properties as a proxy, together with appropriate environmental covariates, will improve modeling efficiency for the predictions of lead (Pb) in forest soils. This finding will contribute to the existing knowledge on DSM using geostatistics and MLAs under forest soils.

Data availability The data that support the findings of this study are available from the corresponding author, [SKA], upon reasonable request.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that they have no known competing personal interests or relationships or conflict of interest that could have appeared to influence the scientific work in this manuscript.

References

Agyeman PC, Khosravi V, Kebonye NM, John K, Borůvka L, Vašát R (2022a) Using spectral indices and terrain attribute datasets and their combination in the prediction of cadmium content in agricultural soil. Comput Electron Agric 198:107077

Agyeman PC, Kingsley JOHN, Kebonye NM, Ofori S, Borůvka L, Vašát R, Kočárek M (2022b) Ecological risk source distribution, uncertainty analysis, and application of geographically weighted regression cokriging for prediction of potentially toxic elements in agricultural soils. Process Saf Environ Prot 164:729–746

Agyeman PC, Kebonye NM, Khosravi V, Kingsley J, Borůvka L, Vašát R, Boateng CM (2023) Optimal zinc level and uncertainty quantification in agricultural soils via visible near-infrared reflectance and soil chemical properties. J Environ Manage 326:116701

Alpaydin E (1997) Voting over multiple condensed nearest neighbors. Lazy Learn 1:115–132

Angelini ME, Kempen B, Heuvelink GBM, Temme AJ, Ransom MD (2020) Extrapolation of a structural equation model for digital soil mapping. Geoderma 367:114226

Arrouays D, Grundy MG, Hartemink AE, Hempel JW, Heuvelink GB, Hong SY, Lagacherie P, Lelyk G, McBratney AB, McKenzie NJ, dL Mendonca-Santos M (2014) GlobalSoilMap: toward a fine-resolution global grid of soil properties. Adv Agron 125:93–134

Arrouays D, Poggio L, Guerrero OAS, Mulder VL (2020) Digital soil mapping and GlobalSoilMap. Main advances and ways forward. Geoderma Reg 21:e00265

Behrens T, Schmidt K, Ramirez-Lopez L, Gallant J, Zhu AX, Scholten T (2014) Hyper-scale digital soil mapping and soil formation analysis. Geoderma 213:578–588

Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, Roger JM, McBratney A (2010) Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TRAC Trends Anal Chem 29(9):1073–1081

Bishop TFA, McBratney AB (2001) A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma 103(1–2):149–160

Borůvka L, Vašát R, Němeček K, Novotný R, Šrámek V, Vacek O, Pavlů L, Fadrhonsová V, Drábek O (2020) Application of regression-kriging and sequential gaussian simulation for the delineation of forest areas potentially suitable for liming in the Jizera Mountains region, Czech Republic. Geoderma Reg 21:e00286

Breiman L (2001) Random forests. Mach Learn 45:5–32

Brennan A, Cross PC, Creel S (2015) Managing more than the mean: using quantile regression to identify factors related to large elk groups. J Appl Ecol 52(6):1656–1664

Cambardella CA, Moorman TB, Novak JM, Parkin TB, Karlen DL, Turco RF, Konopka AE (1994) Field-scale variability of soil properties in central Iowa soils. Soil Sci Soc Am J 58(5):1501–1511

Dharumarajan S, Hegde R, Janani N, Singh SK (2019) The need for digital soil mapping in India. Geoderma Reg 16:e00204

Forkuor G, Hounkpatin OK, Welp G, Thiel M (2017) High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. PloS One 12(1):e0170478

Fouedjio F (2020) Exact conditioning of regression random forest for spatial prediction. Artif Intell Geosci 1:11–23

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 1:1189–1232

Friedman JH (2002) Stochastic gradient boosting. Comput Stat Data Anal 38(4):367–378

Hashimoto Y (2007) Citrate sorption and biodegradation in acid soils with implications for aluminum rhizotoxicity. Appl Geochem 22(12):2861–2871

Hengl T, Heuvelink GB, Stein A (2004) A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120(1–2):75–93

Hengl T, Heuvelink GB, Rossiter DG (2007) About regression-kriging: from equations to case studies. Comput Geosci 33(10):1301–1315

Hengl T, Heuvelink GB, Kempen B, Leenaars JG, Walsh MG, Shepherd KD, Sila A, MacMillan RA, Mendes de Jesus J, Tamene L, Tondoh JE (2015) Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PloS One 10(6):e0125814

Hengl T, Nussbaum M, Wright MN, Heuvelink GB, Gräler B (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6:e5518

Hengl T, Miller MA, Križan J, Shepherd KD, Sila A, Kilibarda M, Antonijević O, Glušica L, Dobermann A, Haefele SM, McGrath SP (2021) African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Sci Rep 11(1):6130

Hettiarachchi GM, Pierzynski GM, Ransom MD (2001) In situ stabilization of soil lead using phosphorus. J Environ Qual 30(4):1214–1221

Heuvelink GB, Webster R (2022) Spatial statistics and soil mapping: a blossoming partnership under pressure. Spat Stat 50:100639

Heuvelink GBM (2018) Uncertainty and uncertainty propagation in soil mapping and modelling. In: McBratney A, Minasny B, Stockmann U (eds) Pedometrics, Progress in soil science. Springer, Cham. https://doi.org/10.1007/978-3-319-63439-5_14

Jiang X, Zou B, Feng H, Tang J, Tu Y, Zhao X (2019) Spatial distribution mapping of hg contamination in subclass agricultural soils using GIS enhanced multiple linear regression. J Geochem Explor 196:1–7

John K, Afu SM, Isong IA, Aki EE, Kebonye NM, Ayito EO, Chapman PA, Eyong MO, Penížek V (2021) Mapping soil properties with soil-environmental covariates using geostatistics and multivariate statistics. Int J Environ Sci Technol 1–16

Kabata-Pendias A (2011) Trace elements in soils and plants, 4th edn. CRC Press/Taylor & Francis, Boca Raton

Kozák J (2010) Forest cover changes and their drivers in the Polish Carpathian Mountains since 1800. In: Reforesting landscapes: linking pattern and process, pp 253–273

Kumpiene J, Lagerkvist A, Maurice C (2008) Stabilization of as, Cr, Cu, Pb and Zn in soil using amendments–a review. Waste Manag 28(1):215–225

Lamichhane S, Kumar L, Wilson B (2019) Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: a review. Geoderma 352:395–413

Li L, Lu J, Wang S, Ma Y, Wei Q, Li X, Cong R, Ren T (2016) Methods for estimating leaf nitrogen concentration of winter oilseed Rape (Brassica napus L.) using in situ leaf spectroscopy. Ind Crops Prod 91:194–204

Liang-yan S, Li C (2009) A fast and scalable fuzzy-rough nearest neighbor algorithm. 2009 WRI Global Congress on Intelligent systems. IEEE 4:311–314

Ma Y, Minasny B, Wu C (2017) Mapping key soil properties to support agricultural production in Eastern China. Geoderma Reg 10:144–153

McBratney AB, Santos MM, Minasny B (2003) On digital soil mapping. Geoderma 117(1–2):3–52

Meinshausen N, Ridgeway G (2006) Quantile regression forests. J Mach Learn Res 7(6):1

Melo AT, Bartaula R, Hale I (2016) GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC Bioinform 17(1):1–15

Moreno-Seco F, Micó L, Oncina J (2003) A modification of the LAESA algorithm for approximated k-NN classification. Pattern Recognit Lett 24(1–3):47–53

Nezhad MTK, Tabatabaii SM, Gholami A (2015) Geochemical assessment of steel smelter-impacted urban soils, Ahvaz, Iran. J Geochem Explor 152:91–109

Nguyen H, Bui X-N, Tran Q-H, Mai N-L (2019) A new soft computing model for estimating and controlling blast-produced ground vibration based on hierarchical K-means clustering and cubist algorithms. Appl Soft Comput 77:376–386. https://doi.org/10.1016/j.asoc.2019.01.042

Nussbaum M, Spiess K, Baltensweiler A, Grob U, Keller A, Greiner L, Schaepman ME, Papritz A (2018) Evaluation of digital soil mapping approaches with large sets of environmental covariates. Soil 4(1):1–22

Odeh IO, McBratney AB, Chittleborough DJ (1995) Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67(3–4):215–226

Pouladi N, Møller AB, Tabatabai S, Greve MH (2019) Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. Geoderma 342:85–92

Quinlan JR (1992) November. Learning with continuous classes. In 5th Australian joint conference on artificial intelligence (Vol. 92, pp. 343–348)

Quinlan JR (2004) Data mining tools See5 and C5. 0. http://www.ruleq uest.com/see5-info. html

Rulequest (2016a) Data Mining with Cubist. https://www.rulequest. com/cubist-win.html. Accessed 26 Feb 2019

Rulequest (2016b) Data Mining with Cubist. https://www.rulequestc om/cubist-info.html RuleQuest Research Pty Ltd.,St. Ives, NSW, Australia

Sanchez PA, Ahamed S, Carré F, Hartemink AE, Hempel J, Huising J, Lagacherie P, McBratney AB, McKenzie NJ, Mendonça-Santos MDL, Minasny B (2009) Digital soil map of the world. Science 325(5941):680–681

Shi T, He L, Wang R, Li Z, Hu Z, Wu G (2023) Digital mapping of heavy metals in urban soils: a review and research challenges. Catena 228:107183. https://doi.org/10.1016/j.catena.2023.107183

Sistla SA, Schimel JP (2012) Stoichiometric flexibility as a regulator of carbon and nutrient cycling in terrestrial ecosystems under change. New Phytol 196(1):68–78

Spark W (2016) Average Weather in Frýdek-Místek, Czechia, Year-Round-Weather spark [WWW Document]. er-iFrýdek-Místek-Czech ia-Year-Round, URL https://weath erspa rk. com/y/83671/ Avera ge-Weath

Szatmári G, Pásztor L (2019) Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. Geoderma 337:1329–1340

Szatmári G, Pásztor L, Heuvelink GB (2021) Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. Geoderma 403:115356

Takoutsing B, GB (2022) Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors. Geoderma 428:116192

Takoutsing B, Weber JC, Rodríguez Martín JA, Shepherd K, Aynekulu E, Sila A (2018) An assessment of the variation of soil properties with landscape attributes in the highlands of Cameroon. Land Degrad Dev 29(8):2496–2505

Umali BP, Oliver, DP, Forrester S, Chittleborough DJ, Hutson JL, Kookana RS, Ostendorf B (2012) The effect of terrain and management on the spatial variability of soil properties in an apple orchard. Catena 93:38–48

Vacek Z, Cukor J, Linda, Vacek S, Šimůnek V, Brichta J, Gallo J, Prokůpková A (2020) Bark stripping, the crucial factor affecting stem rot development and timber production of Norway spruce forests in Central Europe. For Ecol Manag 474:118360

van der Westhuizen S, Heuvelink GB, Hofmeyr DP (2023) Multivariate random forest for digital soil mapping. Geoderma 431:116365

Vaysse K, Lagacherie P (2017) Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291:55–64

Veronesi F, Schillaci C (2019) Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. Ecol Indic 101:1032–1044

Vitousek PM, Porder S, Houlton BZ, Chadwick OA (2010) Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen–phosphorus interactions. Ecol Appl 20(1):5–15

Vrede T, Dobberfuhl DR, Kooijman SALM, Elser JJ (2004) Fundamental connections among organism C: N: P stoichiometry, macromolecular composition, and growth. Ecology 85(5):1217–1229

Wadoux AMC, Brus DJ, Heuvelink GB (2018) Accounting for non-stationary variance in geostatistical mapping of soil properties. Geoderma 324:138–147

Wadoux AMC, Minasny B, McBratney AB (2020) Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth-Sci Rev 210:103359

Wu Z, Chen Y, Yang Z, Liu Y, Zhu Y, Tong Z, An R (2023) Spatial distribution of lead concentration in peri-urban soil: Threshold and interaction effects of environmental variables. Geoderma 429:116193

Yuan ZY, Chen HY (2015) Negative effects of fertilization on plant nutrient resorption. Ecology 96(2):373–380

Zeraatpisheh M, Jafari A, Bodaghabadi MB, Ayoubi S, Taghizadeh-Mehrjardi R, Toomanian N, Kerry R, Xu M (2020) Conventional and digital soil mapping in Iran: Past, present, and future. Catena 188:104424

Zeraatpisheh M, Garosi Y, Owliaie HR, Ayoubi S, Taghizadeh-Mehrjardi R, Scholten T, Xu M (2022) Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. Catena 208:105723

Zhang S, Huang Y, Shen C, Ye H, Du Y (2012) Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. Geoderma 171:35–43

ORIGINAL PAPER

Spatial modelling and quantification of soil potentially toxic elements based on variability in sample size and land use along a toposequence at a district scale

S. K. Ahado[1] · C. Nwaogu[2,3]

Abstract
This study applied ordinary kriging (OK), geographically weighted regression (GWR), and positive matrix factorization to model soil Cu and Mn in the Frýdek Mistek district based on different sample sizes, topography, and land use. OK maps were validated using mean error, while the GWR maps used digital elevation model (DEM) as a covariate. The elements and their different sample sizes revealed high heterogeneous/variability. Cu and Mn showed similar strong, moderate, and weak spatial dependence between the various samples (ratios ranged from 0.00 to 0.99%). The OK interpolation revealed the highest PTE concentration levels (i.e. hotspots) in the north-eastern parts of the study district. The GWR coefficients for both Cu and Mn indicated a positive correlation between DEM and the PTEs towards the south-eastern parts. In addition to the sample sizes, land use and elevation to a large extent determined the distribution, variability, and concentrations of Cu and Mn. All the sample sizes showed the highest concentrations of the PTEs in the lowlands below 500 m where the industrial, commercial, and the arable activities dominated in contrast to the highlands (above 500 m) where forests were dominant. This study will benefit future researchers in selecting appropriate prediction models to enhance the achievement of accurate prediction of PTE contents and spatial distribution in soils. It will also support the land-use planners in identifying the land-use types that are associated with higher concentrations of the PTEs, thus proffering a sustainable solution.

Keywords Soil pollution · Spatial distribution · Potentially toxic elements · Cross-validation · DEM · Land use–land cover

Introduction

Potentially toxic elements in soils have constantly been a significant threat to the environment and the ecosystem due to their toxicity, multiple sources of origin, and their

✉ S. K. Ahado
ahados@af.czu.cz

1 Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamýcká 129, 16500 Prague, Czech Republic

2 Department of Environmental Management, School of Environmental Sciences, Federal University of Technology, Owerri, P.M.B. 1526, Owerri 460114, Nigeria

3 Department of Soil Science "Luiz de Queiroz" College of Agriculture, University of São Paulo, 11 Pádua Dias Avenue, Piracicaba, SP 13418-900, Brazil

non-biodegradable nature (Dai et al. 2018; Mahar et al. 2016; Mazur et al. 2015). PTEs in their enriched concentration levels have several negative effects on plant growth and reproduction as well as on microorganism activity. They disrupt essential biological processes such as nitrogen fixation, adenosine triphosphate development, soil enzyme activity, and eventually deter microbial biomass production (Ahmed and El-Arabi 2005; Buss et al. 2016). Furthermore, high PTE concentration levels have been shown to affect litter decomposition due to impaired biological functions (Berg et al. 1991; Cotrufo et al. 1995), and metal distribution is influenced by litter turnover, owing to the binding affinity between metals and humidified substances. PTE accumulation in soils has become a worldwide problem (Sabiha-Javid et al. 2009; Yang et al. 2013).

Rapid urbanization and industrialization have also accounted for the excessive accumulation of PTEs in the soils of Frýdek Mistek (Agyeman et al. 2021). Industrial activities are critical for economic development and urbanization (Antoci et al. 2018; Khademi et al. 2019; Li et al.

2013; Masto et al. 2017). However, the devastating short- and long-term effects of threats posed by these industrial chemicals, such as industrial influxes, which may eventually accumulate in the soils as PTEs, can never be underestimated and more especially its impact on the ecosystem and potential human health risks.

In most countries, unregulated mining and the uncontrolled release of mining waste cause environmental harm, and hazardous waste disposal activities continue to be practiced. The first step in determining the ecological effects of these pollutants in soils is to determine the concentration levels, bioavailability, and their spatial distribution patterns. PTE risk assessments have been widely published with a focus on contamination of soil and other land use including water (Ahmad et al. 2020; 2021a, b; Chabukdhara and Nema 2013; Fallatah et al. 2022; Zheng et al. 2010; Zhao et al. 2012). Also, there have been various studies on PTEs in the Czech Republic, focusing on either origin or distribution, and mostly done at post-mining sites, urban, mountains, forest soils, and/or in sediments (Borůvka et al. 2005; Gholizadeh et al. 2015; Lichnovský et al. 2017; Sáňka et al. 1995; Sysalová and Száková 2006; Weissmannová et al. 2015). The importance and novelty of this study could be explained by some reasons including: (1) the need to ameliorate the declining and degradation of the crop lands as a result of industrial pollutants, (2) the necessity to improve livestock forage quality by abating the extinction of the pastures species that are becoming endangered and endemic due to severe industrial emissions, (3) the ecological benefits of restoring the biodiversity and the potential ecosystem services of the area that has become threatened by long-term waste discharges from anthropogenic activities, (4) the innovative concept of modelling the potential influence of sample size, land use, and relief on the PTEs pollution level as well as on the spatial distribution and variability of soils. Therefore, these reasons make this work a rare study that has never been performed before in the study area (Frýdek Mistek district of the Czech Republic). Further, the use of positive matrix factorization (PMF), which is one of the latest software of the USA Environmental Protection Agency (US, EPA) used to model soil elemental sources, concentrations, and pollution level, makes this study a unique one in the study area. To further support and prove the novelty in this study is the application of CANOCO—a multivariate ordination analysis software used to measure and illustrate the relationships among the sample sizes, LULC, elevation and the PTEs.

In respect to the above-mentioned context, the study aimed at (1) quantifying and modelling the pollution levels of PTEs based on different sample sizes in the study area; (2) assessing the prediction accuracy of PTEs using OK and GWR methods; (3) identifying and mapping the spatial distribution patterns of Cu and Mn using different sample sizes;

and (4) examining the impacts of land use and relief as well as their nexus with the PTEs distribution in the soils. It is therefore hypothesized that there is (1) a positive relationship between sample size and prediction accuracy of the spatial distribution of PTEs in soils and (2) a strong association between LULC, elevation and PTE concentrations.

Materials and methods

Description of the study area

The research area is situated at latitude of 49° 41′ 0′ N and longitude of 18° 20′ 0′ E. It undulates at heights between 225 and 327 m above sea level. The study area is a cold temperate climate that receives heavy rainfall even during the dry months. The summer season in Frýdek Mistek is humid and partly wet, while the winter periods are very cold, dry, and windy. Though the temperature ranges from 24 to 75 °C throughout the year, it seldom falls below 8° or rising above 86° (Weather Spark 2016). The average annual precipitation is between 685 and 752 mm. The geomorphology is characterized by rugged terrain that is considered part of the Moravian-Silesian Beskydy with the outer Carpathian Mountain as the highest peak. The area survey of the district is estimated at 1208 km^2, with lower landscape found in the northern part (Fig. 1). A significant area for evaluating the distribution and related ecological impacts of PTEs is Trinec and Vítkovice. In and around the district is Ostrava which forms part of the area under investigation and is endowed with steel and metal industries. The soil properties are clearly distinguishable based on the colour, structure and carbonate content. The soil contains medium and fine texture, which comes from its parent materials, and is made up of colluvial, alluvial, or Aeolian deposits. Cambisols and stagnosols are the most common reference soil groups in the study region. Cambisols have 56.7% coverage in the Czech Republic, while stagnosols have 7.01% (Vacek et al. 2020).

About 39.38% of the land area is used for anthropogenic activities including industries, crop cultivation, pastures and animal husbandry, while 49.36% are forests (Fig. 2). The PTEs contamination in the region is thought to be caused by atmospheric deposition from human activities such as local steel mills, vehicular emissions, tire abrasion, and agriculture, which are dominant in the north, north-west and north-eastern parts of the area (Fig. 2).

Soil sampling and laboratory analysis

Legacy soil data and current soil maps for the Czech Republic were the primary sources of soil data for this study. The legacy data comprise a database of forest soil properties, which includes thousands of sampling locations from

Fig. 1 Sampling points, relief of the study area shown as digital elevation model (DEM), and location of the study area in the Czech Republic
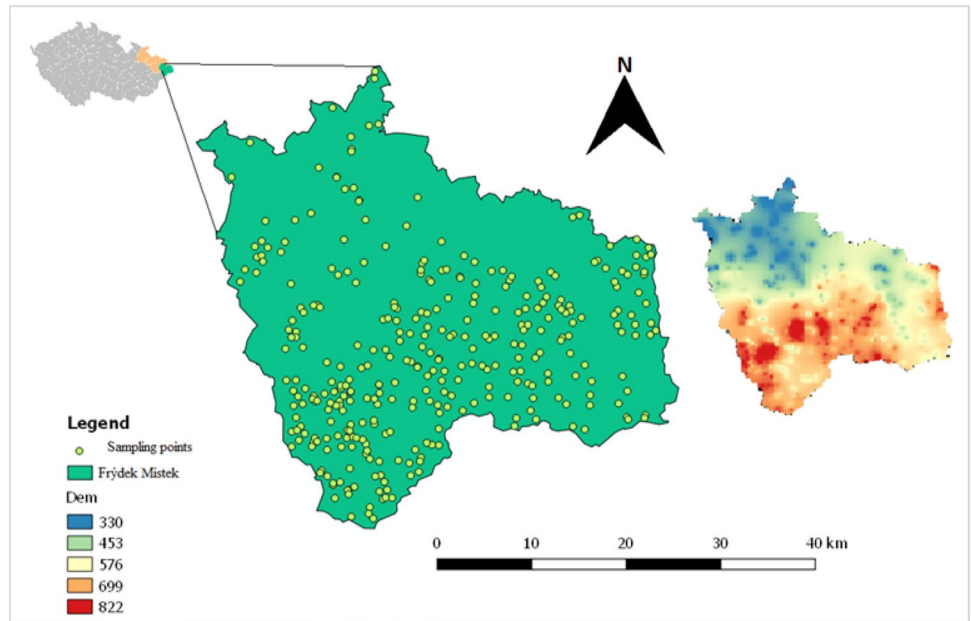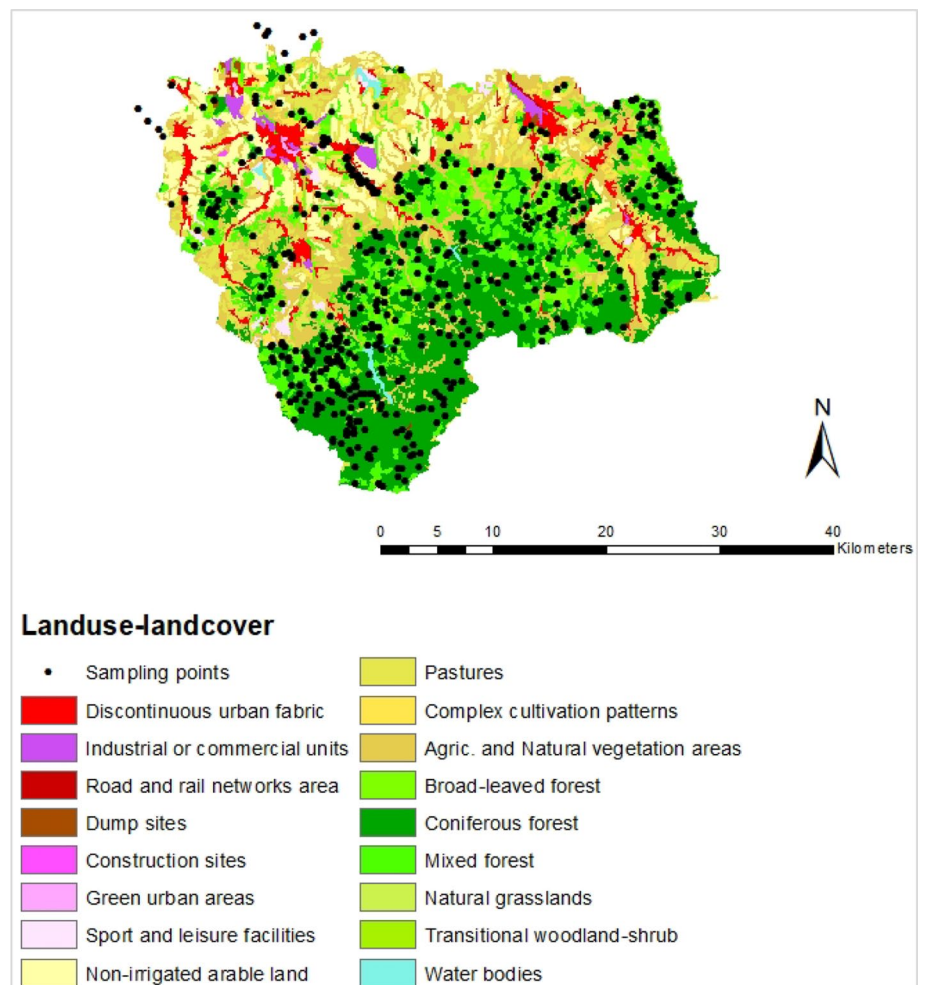

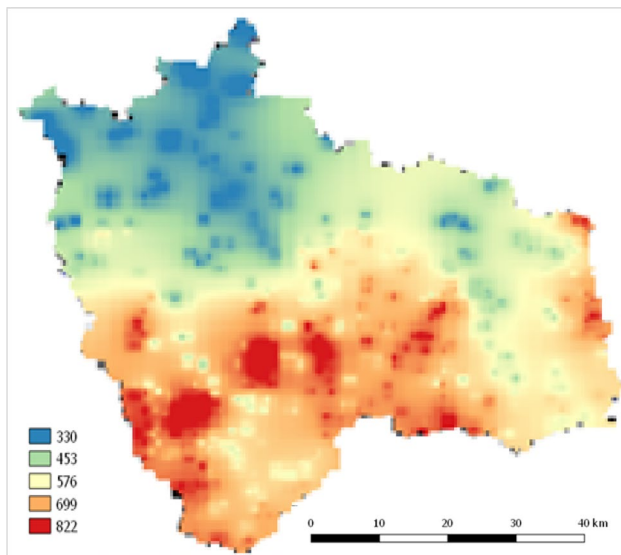**Fig. 2** Distribution of the sampling points across the land use–land cover

Fig. 3 The relief (in metres) shown as digital elevation model (DEM). This was categorized into two groups: lowland (below 500 m) and highland (above 500 m)

institutional samples gathered across the Czech Republic between 1998 and 2019. The data also cover environmental covariate of the sampled stands (Borůvka et al. 2018). A total of 336 mineral topsoil (0–30 cm) was used for this study after the data were transformed. The transformation is required because the soil data come from varying depths of the soil profile from different surveys. Environmental covariate and pedological data of the samples collected within the forests covered broadleaves, coniferous, mixed forests, and pollution risk forest areas.

Chemical analysis and instrument

The presence of elements such as Cu and Mn in the soil was extracted using the aqua regia standard method (ISO 11466:1995, 1995) to determine their pseudo-total content (Melo et al. 2016). For the quality control (QC) of the method, the standard addition technique was adopted. For example, the QC of the concentration determination was guaranteed using the SRM 2711 (Montana II soil) reference material (National Institute of Standards and Technology, Gaithersburg, Maryland, USA). The values achieved were consistent with the reference data. The recovery differences were generally $< 10\%$ ($n = 3$). The detection limits for the elements based on the applied method were Cu (0.015 mg L^{-1}) and Mn (0.05 mg L^{-1}).

Table 1 Description statistics of PTE concentrations and their spatial distribution in the soil

Sample sizes (mg kg^{-1})	Mean	Median	Mode	Mini	Maxi	Std dev	CV (%)	Kurtosis	Skewness
Cu ($n=336$)	5.05	4.19	3.30	0.60	43.65	43.65	72.40	38.40	4.46
Cu ($n=276$)	4.95	4.06	3.60	0.60	22.70	3.09	62.29	6.09	1.92
Cu ($n=216$)	4.80	4.04	3.30	0.60	20.36	2.86	59.42	3.98	1.53
Cu ($n=156$)	4.85	4.04	3.60	0.60	20.36	2.92	60.10	5.21	1.78
Cu ($n=96$)	4.72	3.93	3.60	0.60	12.60	2.47	51.96	0.78	1.04
Mn ($n=336$)	190.25	166.25	133.00	10.00	1190.00	131.77	69.16	12.57	2.37
Mn ($n=276$)	198.47	124.28	47.00	0.50	1332.22	216.78	109.03	6.03	2.15
Mn ($n=216$)	191.12	125.63	47.00	0.50	1332.22	210.72	110.00	7.05	2.30
Mn ($n=156$)	195.26	125.63	47.00	0.50	1332.22	209.60	107.00	6.59	2.16
Mn ($n=96$)	179.49	102.65	47.00	0.50	883.13	186.52	103.37	2.31	1.58
Mean values (mg kg^{-1}) for Cu*									
European mean values	17.30								
Worlds mean values	38.90								
Crati basin mean values	44.36								
US-EPA mean values	40.70								
Mean values (mg kg^{-1}) for Mn*									
European mean values	524								
Worlds mean values	488								
Crati basin mean values	1300								
US-EPA mean values	750								

*Sources: Kabata-Pendias and Szteke (2015), Guagliardi et al. (2012), Ahmad et al. (2021a, b), Waseem et al. (2014), Onyedikachi et al. (2018), US-EPA (2007a, 2007b)

Table 2 Semivariogram models and cross-validation results of Cu and Mn

Parameters	Models	Nuggets (C_o)	Sill ($C_o + C$)	Range	*N/S	**SAC
Cu ($n = 336$)	Exponential	1,093,754	1,423,081	5,418,205	0.77	Weak
Cu ($n = 276$)	Sphere	6,600,203	8,618,451	4,304,304	0.77	Weak
Cu ($n = 216$)	Exponential	6,002,543	7,741,647	13,895.89	0.78	Weak
Cu ($n = 156$)	Exponential	0.00	6,978,982	4,042,666	0.00	Strong
Cu ($n = 96$)	Linear	5.93	6.02	17,740.49	0.99	Weak
Mn ($n = 336$)	Linear	10,607.86	19,349.83	24,170.51	0.55	Moderate
Mn ($n = 276$)	Linear	3,438,621	54,785.09	43,779.77	62.77	Weak
Mn ($n = 216$)	Exponential	29,124.58	121,890.77	128,879.4	0.24	Strong
Mn ($n = 156$)	Linear	33,507.82	43,719	28,904.15	0.77	Weak
Mn ($n = 96$)	Linear	32,583.72	38,037.90	26,227.54	0.86	Weak

*Nugget/sill ratio

**Spatial autocorrelation

Software and spatial modelling process

Categorization into five sample sizes

The Rstudio software version (4.0.5) was used to further classify the soil samples into five different categories by reducing the total sample ($n = 336$) by 60 from the original data set. R software code was used to model the five (5) different sample sizes, $n = 336$, $n = 276$, $n = 216$, $n = 156$, and $n = 96$, respectively. Given the total sample size of $n = 336$, each of the subsequent subsamples was randomly selected by assigning a unique seed [e.g. set. seed (276)] to ensure reproducibility of results. The various sample sets were randomly generated together with the *x, y* coordinates, respective terrain height equivalent, and associated land use–land cover (LULC) as derived from the field. The sequential reduction of the number of models by 60 was based on the author's discretion as it could have been any other number. Basic statistical parameters such as mean, median, minimum, maximum, standard deviation, coefficient of variance, kurtosis, and skewness were determined. The OK interpolation technique was used to enhance the creation of the spatial distribution maps of the PTEs using the non-transformed values of Cu and Mn. The OK interpolation determines the weights based on the distances between measured and projected points and the semivariogram model, based on variogram theory and structural analysis and the premise of spatial correlation among regional variables (Cressie et al. 1990). In analysing the spatial similarity within the measured data points, a semivariogram was determined. Dragović et al. (2014) defined the empirical semivariogram as half the averaged squared difference between paired data

Table 3 Values and interpretation of spatial autocorrelation

Spatial autocorrelation (SAC)	Explanation
< 0.25	Strong autocorrelation
0.25–0.75	Moderate autocorrelation
> 0.75	Weak autocorrelation

Karami et al. (2009)

Table 4 Cross-validation results of ordinary kriging interpolation

Parameters	ME mg/kg	MAE	RMSE	MSDR
Cu ($n = 336$)	0.011	2.22	3.62	1.029
Cu ($n = 276$)	0.009	2.2	3.08	1.144
Cu ($n = 216$)	−0.003	2.04	2.74	1.13
Cu ($n = 156$)	−0.033	2.12	2.93	1.29
Cu ($n = 96$)	−0.003	1.93	2.48	1
Mn ($n = 336$)	0.132	94.54	129.42	1.39
Mn ($n = 276$)	1.383	157.24	214.48	1.23
Mn ($n = 216$)	−1.013	150.61	209.35	1.34
Mn ($n = 156$)	1.554	156.08	215.19	1.52
Mn ($n = 96$)	−0.058	146.88	190.39	1.04

values separated by a distance interval, which shows the spatial autocorrelation of samples and variation patterns on the entire spatial scale. This is given in Eq. (1) following Dragović et al. (2014) and Kumar et al (2023):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[z(xi) - z(x_i + h) \right]^2 \tag{1}$$

where $\gamma(h)$ is the semivariogram value; $z(x_i)$ and $z(x_i + h)$ represent the measured values of the same PTE at x_i and $x_i + h$, respectively; and $N(h)$ is the number of sample pairs that agree with the lag of h. Different h values will produce a series of $\gamma(h)$ values in the empirical semivariogram. In fitting the semivariogram cloud, the concise theoretical model should be chosen using the least square method fit the semivariogram. Spherical, exponential, Gaussian, linear and other theoretical models are commonly used, with nugget (C_0), range, and sill (C) being three critical parameters. When the sampling interval is 0, the nugget is the semivariogram value, which represents the calculated error and spatial variance at the minimum abstract scale. The scope of the spatial correlation is represented by the range. The maximal variance of samples is known as the sill, and when the semivariogram value is more than the sill, there is no spatial correlation between the models (Wang et al. 2014).

The ordinary kriging method of interpolation is presented in Eq. (2) according to Kumar et al (2023):

$$\hat{z}(x_o) = \sum_{i=1}^{n} \lambda_i z(x_i) \tag{2}$$

where $\hat{z}(x_o)$ is the predicted value of x_o, $z(x_i)$ is the measured x_i, value λ_i denotes the weight of samples x_i, and n represents the number of samples.

Sample sizes grouping under the different land use–land cover

By using the Rstudio software version (4.0.5) and ArcGIS version (10.7.1), the classified five soil sample sizes were grouped based on land use–land cover types. There are sixteen LULC types (excluding the waterbody) in the study area (Fig. 2). The LULC was classified into three distinct groups, namely industrial area, arable/pasture, and forestland area. All the LULC shown in green colours (broad-leaved forest, mixed forest, coniferous forest, and transitional wood-shrub) were grouped under the forestland area, while the arable/pasture includes non-irrigated arable land, pastures, natural grassland, complex cultivation area, agriculture, and natural vegetation. On the other hand, discontinuous urban fabric, industrial or commercial units, road and rail network, dump sites, construction sites, and urban green areas were grouped under the industrial area (see Fig. 2). This was done to measure the contributions of the LULC for the PTEs and the variability in sample sizes distribution. Here, the positive

matrix factorization (PMF, EPA version 5.0, Washington, DC, USA) was used for the estimation of source apportionment and contamination level of the PTEs, whereas the ordination model of CANOCO 5.0 was used to show the association among LULC, relief, sample size, and PTEs.

Sample sizes grouping under lowland (below 500 m) and highland (above 500 m)

The Rstudio software version (4.0.5) was used to further group the classified five soil samples sizes under lowland (denoting areas below 500 m) and highland soil areas (representing regions above 500 m). These topographical gradients are clearly visible in Fig. 3 where the lowlands are dominant in the northern part, while the highlands are found in the south. This was done to determine whether height above sea level influenced the pollution level, soil sample sizes distribution, and variability.

Cross-validation

Cross-validation was adopted to compare various interpolation methods and helped in the best interpolation parameter and model (Stone 1974). The concept behind leave-one-out cross-validation is that the significance of each sample is unknown and is estimated by the surrounding samples. The calculated and predicted values' mean error (ME), mean absolute error (MAE), root-mean-square error (RMSE), and mean-squared deviation ratio (MSDR) were chosen as evaluation indicators.

The variables separately include estimates of overall estimation bias and interpolation accuracy (Qu et al. 2017). The RMSE was used to evaluate the prediction accuracy. The smaller the RMSE value, the more accurate the prediction outcome will be. When the RMSEs are equal, the accuracy should be assessed using the smaller ME criterion, which leads to higher accuracy. The MSDR was also used to determine the degree of fitting of the theoretical variation function. The closer the MSDR value is to 1, the more accurate the fit variation function will be (Zhang et al. 2011). These validation methods as shown in Eqs. (3), (4), and (5) were applied following the works of Zhang et al. (2011), Sishah et al. (2023), and Mueller et al. (2023).

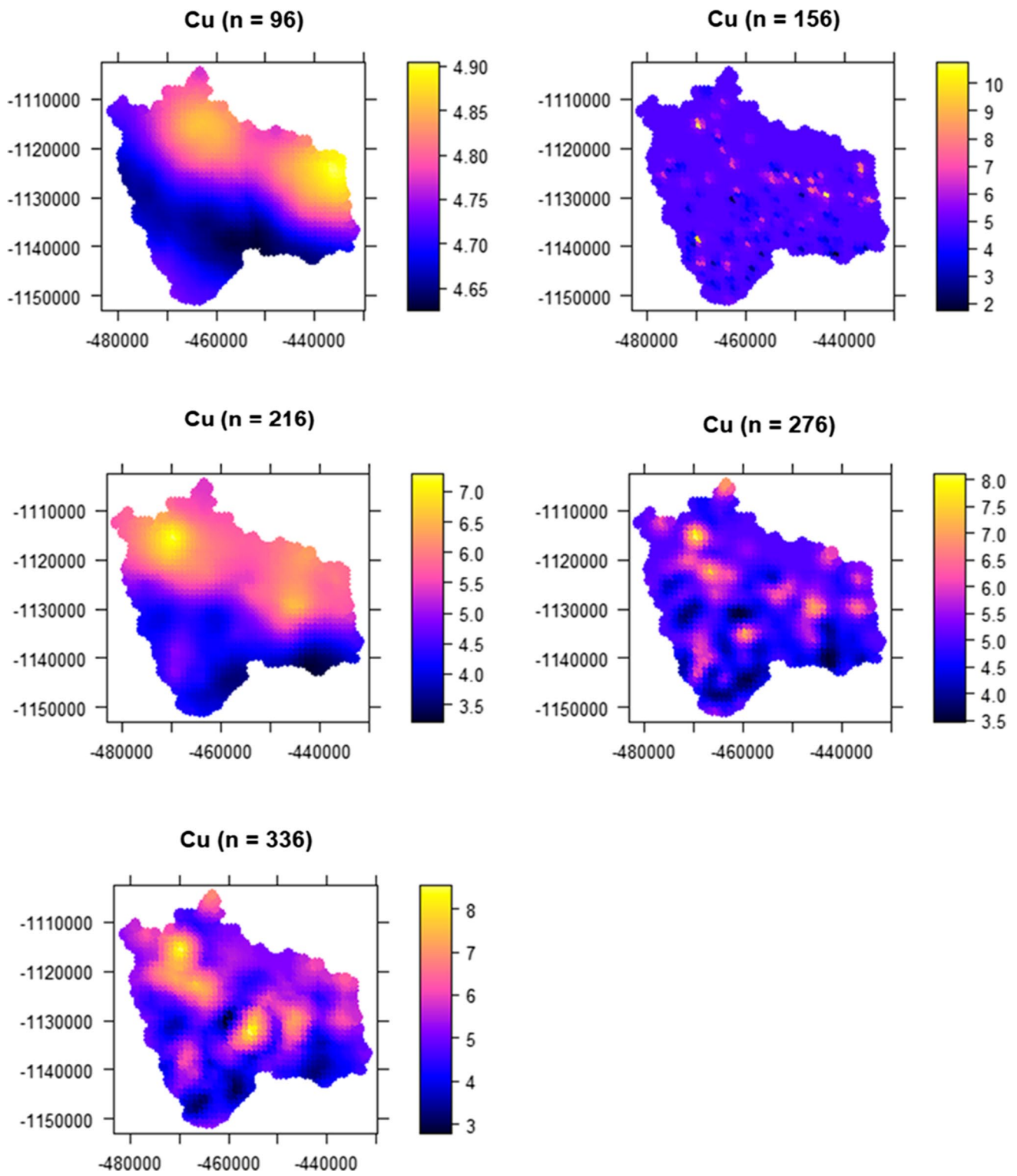$$ME = \frac{1}{n} \sum_{i=1}^{n} \left[Z(x_i) - Z^*(x_i) \right] \tag{3}$$

Fig. 4 Ordinary kriging interpolation and spatial distribution maps of Cu (mg kg^{-1}) with different sample sizes
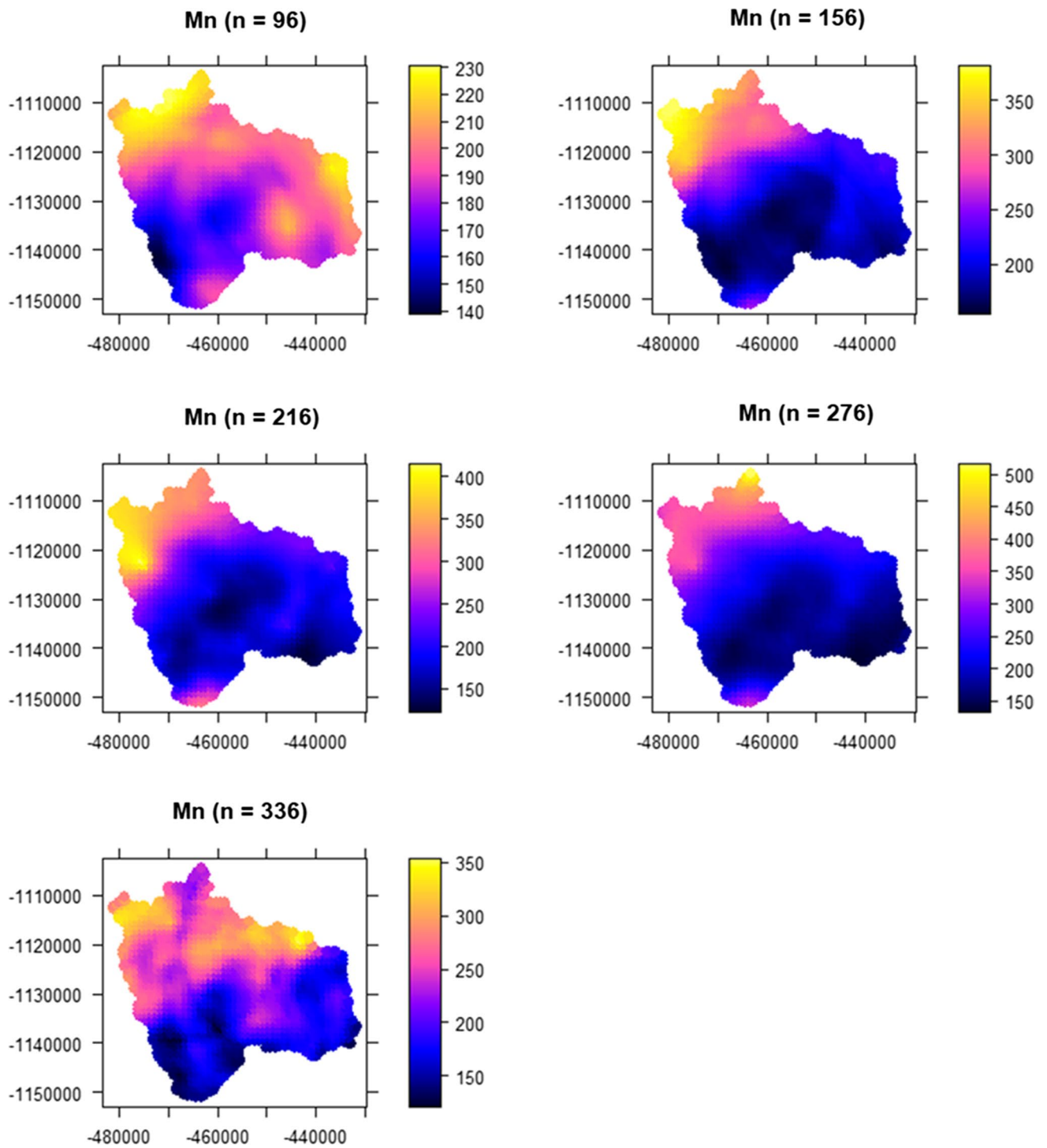
Fig. 5 Ordinary kriging interpolation and spatial distribution maps of Mn (mg kg^{-1}) with different sample sizes

$$\text{RMSE}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[Z(x_i) - Z^*(x_i)\right]^2} \qquad (4)$$

$$\text{MSDR} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left[(z(x_i) - \hat{z}(x_i))\right]^2}{\sigma_i^2}\right) \qquad (5)$$

where $z(x_i)$ is the measured value, $\hat{z}(x_i)$ are the predicted values, σ_i and n are the variance and the size of the validation set in samples, respectively.

Geographically weighted regression

Geographically weighted regression (GWR) model is the locally weighted least square method where the weight is the distance function between the approximate site's geospatial position and the other observation sites' geospatial locations. The GWR is a spatial statistical method that considers both spatial heterogeneity and autocorrelation and can be used to investigate local relationships among various possible spatial variations (Brown 2012; Hou et al. 2017). The GWR technique was used in this study to elucidate and visualize the relationship between PTEs' spatial distributions in surface soil and the DEM (Fig. 1). The relationship between the parameters and their geospatial location is characterized by the variation of the parameters with their geospatial location. The GWR model is stated as following the work of Li et al (2023) and Li and Wang (2023):

$$y_i = \beta_O(\mu_i, v_i) + \sum_{k=1}^{p}\beta_k(\mu_i, v_i)x_{ik} + \varepsilon_i \quad i = 1, 2, \ldots n \qquad (6)$$

where $y_i(\mu_i, v_i)$ is the coordinate of each sampling site (such as latitude and longitude), $\beta_k(\mu_i, v_i)$ is the regression parameter of each sample site, which is a function of the geographic location i, and ε_i represents the error at location i. $\varepsilon_i \sim N(O, \delta^2)$, $\text{cov}(\varepsilon_i\varepsilon_j) = 0(i = \neq j)$. In this model, it is critical to choose the best bandwidth. The geographically weighted regression analysis uses the AIC criteria to select the weight function bandwidth (Cao et al. 2017). Equation (7) is applied according to Lee et al. (2023) and Cao et al. (2017) that were adopted:

$$\text{AIC} = -2_n \text{In}L(\hat{\sigma}) + n\text{In}2 + n\left[\frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)}\right] \qquad (7)$$

where the trace tr(S) of the hat matrix S depends on the bandwidth b, and $\hat{\sigma}$ is the predicted maximum variance of the random error term. The bandwidth of the GWR function of the minimum AIC value of an individual sample is the sample's optimal bandwidth. The surface content of a single PTE is the dependent variable in GWR analysis, while the surface content of the other PTE and the covariate (DEM) of the sampling points are the independent variables.

Results and discussion

Descriptive statistics of PTE concentrations and their spatial distribution in the soil

The basic statistical characteristics of the studied PTEs, including Cu and Mn for the various sample sizes of soil, are described in Table 1. The coefficient of variance (CV) explains the degree of variation within PTE concentrations (Karami et al. 2009). A CV value of less than 20% represents low variability, and a CV that falls between 21 and 50% indicates a moderate variability. On the other hand, a CV ranging from 50 to 100% signifies high variability. On the other hand, when CV is greater than 100% (i.e. > 100%), it is described as extremely high heterogeneity. In this study, the CV of the PTEs Cu indicates a high variability, while the CV of Mn exhibited an extremely high heterogeneity pattern in variation (Table 1). In general, both elements and their different sample sizes revealed high heterogeneous/variability. The data for PTEs in all the samples had a clear positively skewed distribution: Cu ($n = 336$) having the highest skewness of 4.46 and Cu ($n = 96$) having the lowest skewness of 1.04. The skewness value for Mn ranged from 2.37 to 1.58.

The mean values obtained for both Cu and Mn were lower compared to the European mean values, worldwide mean value (Kabata-Pendias and Szteke 2015), and Crati basin values (Guagliardi et al. 2012). This may suggest that the soils had low pollution. Although this is consistent with the study, there is need to assess other aspects of biogeochemistry, fractions, and speciation of these elements.

Semivariogram models and cross-validation results of Cu and Mn maps

Semivariogram and the nugget-to-sill ratio were used to define spatial dependence and investigate spatial variability. Extreme spatial dependence was indicated by a ratio of less than 25%, and moderate spatial dependence was indicated by
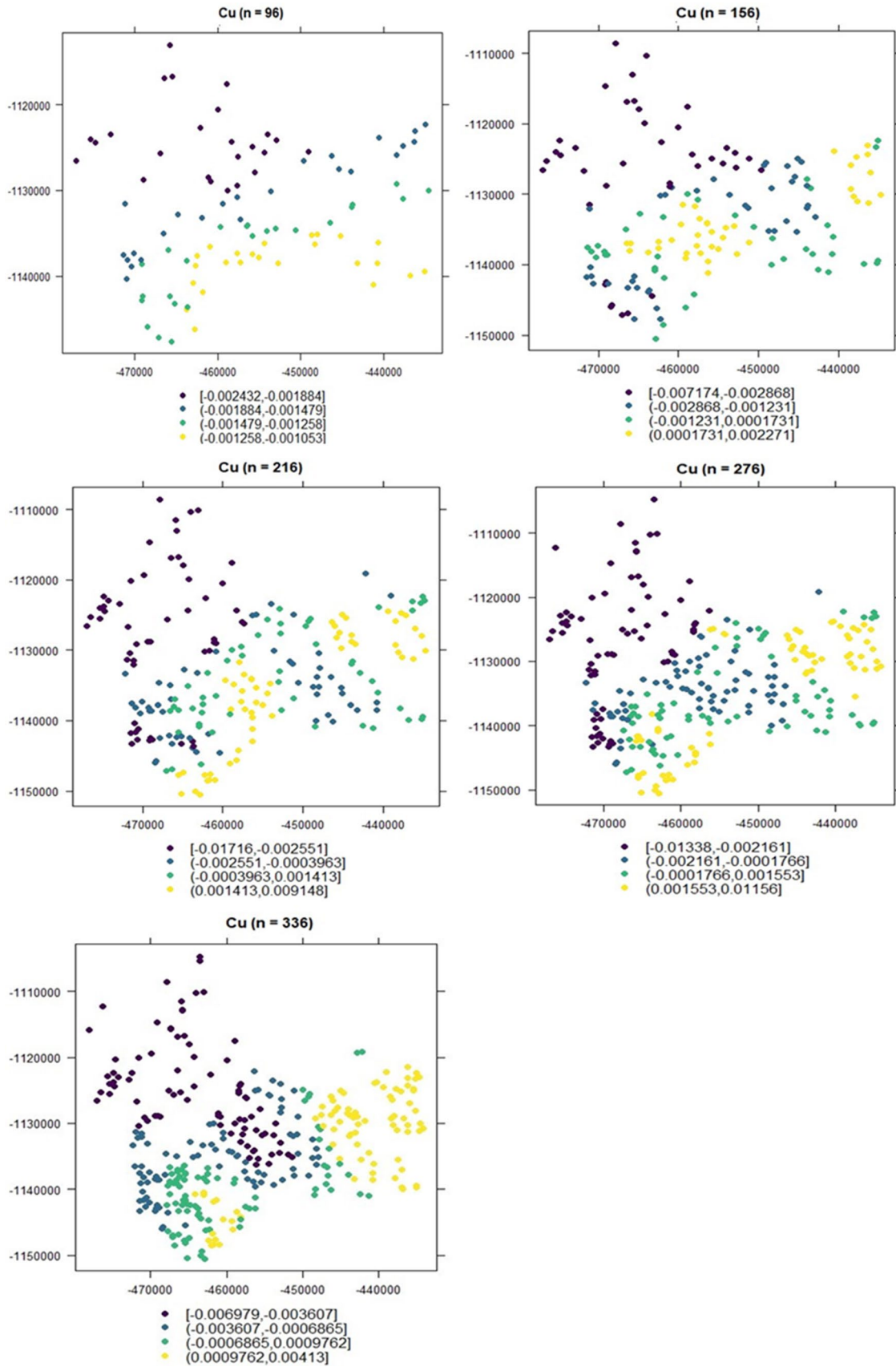
◄**Fig. 6** Geographically weighted regression (GWR) interpolation with digital elevation model (DEM) and point distribution maps of Cu (mg kg^{-1}) with different sample sizes

a ratio between 25 and 75%. Finally, a ratio of 75% showed a poor spatial relationship (Karami et al. 2009). Intrinsic factors are usually to blame for a strong spatial dependence of soil properties and PTEs, whereas external factors are to blame for weak spatial dependence (Cambardella et al. 1994; Wu et al. 2009). The total content of Cu and Mn in this study (Table 2) showed strong, and moderate-to-weak spatial dependence between the various samples (ratios ranged from 0.00 to 0.99%). This study showed strong autocorrelation (extreme spatial dependence), moderate (moderate spatial dependence), and weak autocorrelation (weak/poor spatial relationship) (Table 3). Strong spatial dependence is caused by intrinsic factors such as industrial production and soil practice management. At the same time, moderate spatial dependence might be explained by intrinsic and extrinsic factors including industrial production, agricultural practice, parent material, and topography (Hani and Karimineja 2010).

Cross-validation results of ordinary kriging interpolation

The components of the cross-validation test revealed that the interpolation of all soil variables for distribution maps was reasonable. The overall mean error was close to zero, except in the case of Mn ($n = 276$) and Mn ($n = 156$), and the mean-squared error was also smaller than the sample variance of all soil variables (Table 4). The standardized mean-squared error for Mn and Cu revealed that the interpolation was accurate. The RMSE values were slightly higher, while the MSDR values for all elements were significantly lower indicating the accuracy of the models. Combining ME, MAE, RMSE, and MSDR will replicate the characteristics of original samples, preserve their spatial variability and

mutability, ensure prediction accuracy, as well as enhance the model-fitting effect of various methods.

Ordinary kriging interpolation results of Cu and Mn

The spatial distribution maps of Cu and Mn were generated by the OK method (Figs. 4 and 5). All PTEs had a similar pattern of distributions except in the case of Cu ($n = 156$ and 276). The spatial distribution maps revealed similar regional patterns, with high PTE contents in the north, north-east and north-west. The northern parts of the study district had the highest concentration (hotspots) of all PTEs. This could be attributed to the facts that the socio-economic and industrial activities are concentrated towards the northern part of the study area; thus, toxic elements are regularly discharged (Wang et al. 2020; Guo et al. 2019). However, $n = 156$ and 276 are slightly distributed across the entire study area. The spatial dependency values (Table 3) and the structure of the PTEs pollution in the study area (Figs. 4 and 5) might be attributed to a typical soil-geochemical patterns of PTEs pollution formation by the disposal of human-made waste in the forest soils (Xu et al. 2016). The spatial distribution maps of Cu and Mn showed a similar geographic trend. The pattern of similarity is strongly supported by the mean values of Cu and Mn (Table 1).

Geographically weighted regression analysis

The remarkable fit and accuracy of the explanation variables indicated that they could adequately characterize the dependent variables. To further examine the relationship or effects of environmental covariates on the behaviour of PTEs, the DEM was used as an independent variable. A digital elevation model (DEM) is a three-dimensional (3D) representation of the surface of the terrain generated from elevation data. In the 1970s, the term DEM was coined to differentiate the most basic form of terrain relief modelling from more complex forms of digital surface representation.
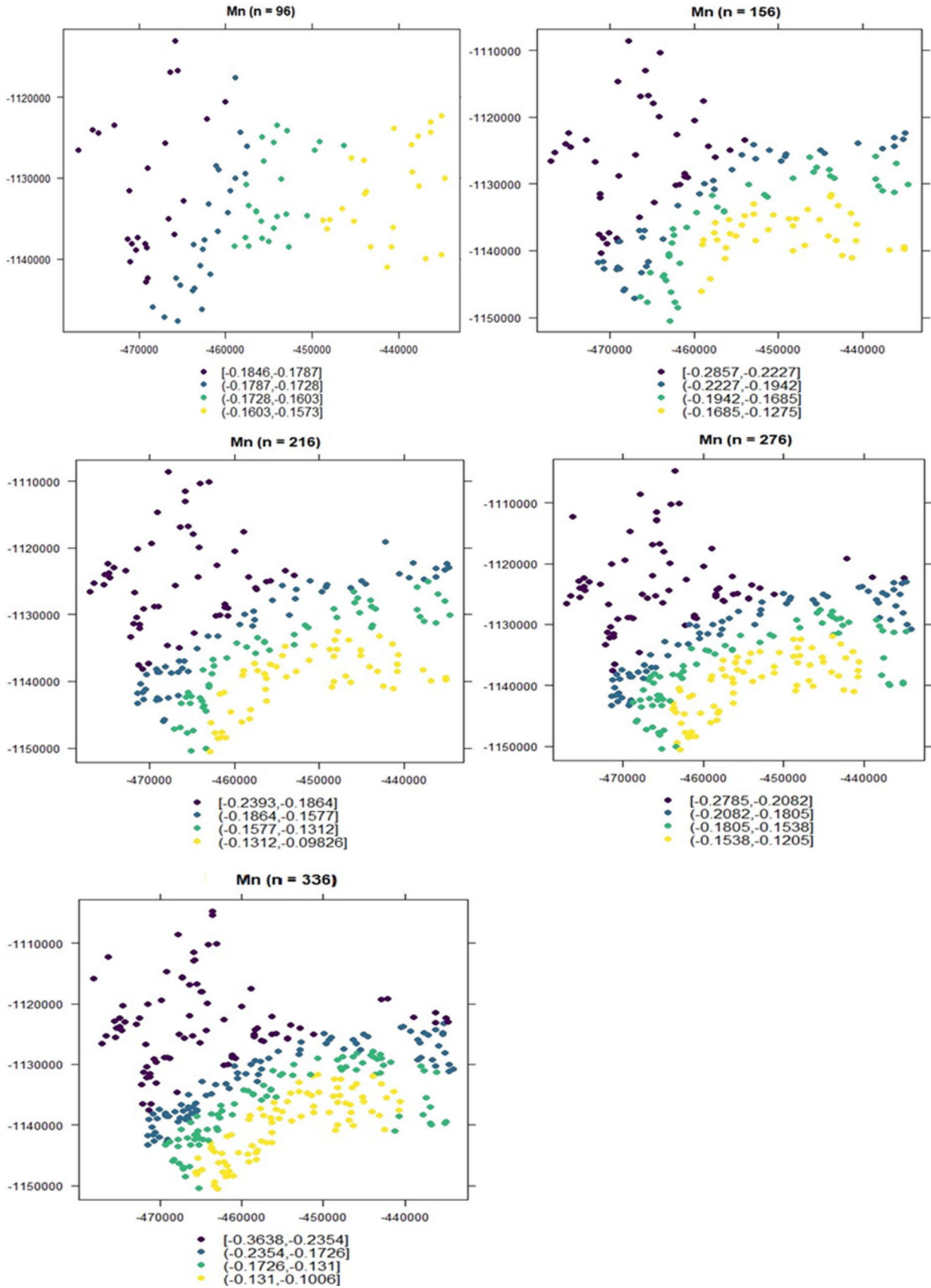
◀**Fig. 7** Geographically weighted regression (GWR) interpolation with digital elevation model (DEM) and point distribution maps of Mn (mg kg⁻¹) with different sample sizes

In the south-eastern part of the study area, DEM and Cu concentrations were positively associated for all sample sizes, while in the north-western part, the regression coefficient steadily decreased before turning negative (Fig. 6). The regression coefficients of Mn and DEM followed the same trend with concentration strongly correlated towards the south-eastern part of the study area (Fig. 7). This revealed that DEM affected the distribution of Cu and Mn. Generally, the study results are important because the prediction accuracy of PTEs is not dependent on total sample size, but other environmental variables such as elevation and land use are involved. In sum, the selection of appropriate prediction models will enhance the achievement of accurate prediction of PTEs spatial distribution in soils.

The multivariate ordination analysis of RDA demonstrated the association between the sample sizes and land use along the toposequence (Fig. 8). All the sample sizes showed the highest concentrations of the PTEs around the industrial/commercial and the arable land soils under the lowland below 500 m. This was also shown in the OK modelling analyses (Figs. 4 and 5) where the north had higher PTE concentrations than the southern part. The contents of Cu and Mn in all the sample sizes were relatively insignificant in the forest land area where the topography is high. This finding was consistent with the map showing the land use–land cover of the study area (Figs. 1 and 2). It could be attributed to the concentration of industrial, commercial, and arable activities in the lowland areas when compared to the highland areas. The result is consistent with the report by Wang et al (2020) who revealed that the type of LULC in an area contributes significantly to the distribution and redistribution of PTEs in the surface soil. Industrial, commercial, and agricultural sites have been known as important sources of PTEs (Guo et al. 2019).

For example, the enrichment of PTEs in agricultural soil was primarily because of the excessive use of fertilizers and the animal dungs (Xu et al. 2016). Other studies on the correlation between elevation and concentration of PTEs revealed that lowland soils have higher contents of the PTEs because of runoff from the highlands and slopes downwards (Qiao et al. 2017; Wang et al. 2020).

This study examined the sources of PTEs in the study area by applying the positive matrix factorization (PMF) model developed by the US-EPA. The PMF model was introduced as one of the most suitable and latest models with high capabilities for efficient PTE source evaluation (Chen et al. 2010a, b). The validity and reliability of the analysis are centred on minimum Q to model the residual matrix that influences a substantial number of variables. To achieve the most accurate result, the PMF model was run for at least 20 times, and the best outputs (which were Run 6, Run 7, and Run 20) were chosen following the software developer's guide (U.S. EPA. 2010). The PMF simulation generated six factors (Figs. 9, 10, and 11) and revealed the sources of the contributions based on each PTE (Norris et al. 2014). The result for the sample sizes showed that factor 1 was dominated by Cu in the sample size 96 with a factor loading of 77.8%, while factor 2 had Mn in sample size 216, which accounted for the highest element with factor loads of 83.4% (Fig. 9). The manganese content in a sample size of 276 accumulated factor 3 with factor loads of 84.5%. The model analysis revealed that factor 4 had the highest loads for the accumulation of the PTEs in the different sample sizes. These include Cu in 336 sample size (80.1%), Cu in 276 sample size (62.3%), Cu in 156 sample size (61.9%), Mn in 336 sample size (40.4%), Cu in 216 sample size (35.7%), and Cu in 96 sample size (22.3%). Factor 5 was dominated by Mn in the 156-sample size, whereas Mn in sample size 96 accrued factor 6 with factor loads of 78.9%.

The percentage contributions of Mn in each land-use types revealed an interesting information where none of the forest soils except broad-leaved forest had a significant factor load or contribution in Cu accumulations (Fig. 10). Green urban area showed the highest Cu accumulation in factor 1 with 70.1% contribution factor, while construction sites (75.3%) and transition woodland shrubs (73.2%) had the highest for factor 2. Factor 3 had discontinuous urban fabric (75.3%) and natural irrigated arable land (39.9%), whereas factor 4 had natural grassland as the land use with the highest soil Cu and accrued 85.1% factor loads.

In the cases of factor 5 and factor 6, pastures (83.6%) and broad-leaved forest (70.5%), respectively, accounted for the highest. Similar to the results from PMF model analysis for Cu, the percentage contributions of Mn in each land use types indicated that only the soil under the broad-leaved forest had a significant factor contribution in Mn accumulation (Fig. 11). In factor 1, green urban soil accounted for 83% factor loading for Mn, while in factor 2 natural grassland accrued the highest factor loads of 24.2% followed by broad-leaved forest (23.9%) and transitional woodland
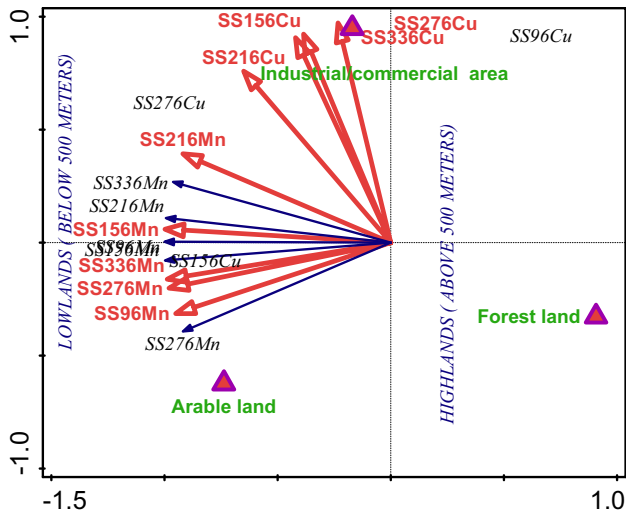
Fig. 8 Multivariate ordination analysis of RDA Showing the association between the sample sizes, land use, and DEM. Description of Abbreviations: SS96Cu, SS156Cu, SS216Cu, SS276Cu, and SS336Cu represent sample sizes of 96. 156, 216, 276, and 336, respectively, for copper, while SS96Mn, SS156Mn, SS216Mn, SS276Mn, and SS336Mn represent sample sizes of 96. 156, 216, 276, and 336, respectively, for manganese

shrubs (23.5%). Factor 3 had construction sites as the land use with the highest Mn accumulation with loading factor of 42.6%, whereas transitional woodland shrub had 40.4%.

Natural irrigated arable land (52.1%) accounted for the highest Mn in factor 4. In factor 5, discontinuous urban fabric had the highest accumulation of Mn with 61.0%. Factor 6 had industrial/commercial sites and pasture lands recording the highest factor loads of 77.5% and 70.1%, respectively. The highest factor loads were found in either the industrial or arable and pasture lands, while the forests recorded relatively low factor loads except in broad-leaved forest. The OK modelling results in Figs. 4 and 5 also reflected this pattern. This could be attributed to the high concentrations of the PTEs in the lowland area (below 500 m) where the anthropogenic activities are dominant (Guo et al. 2019; Nguyen et al. 2020; Vannini et al. 2021; Xiao et al. 2021; Xu et al. 2016). In addition, agents of erosion such as wind and water might carry loads of pollutants containing Cu and Mn from the highlands to the industrial, commercial, arable, and pasture areas at the lowlands, which consequently contaminated the soil (Xiao et al. 2021). The broad-leaved forest was the only forest type that recorded some high factor loads of the PTEs. This could be explained by the reason that the broad-leaved trees have the capacity to retain and absorb in their leaves pollutant carried by wind. These retained PTEs are later transferred into the soil through stem-flow or
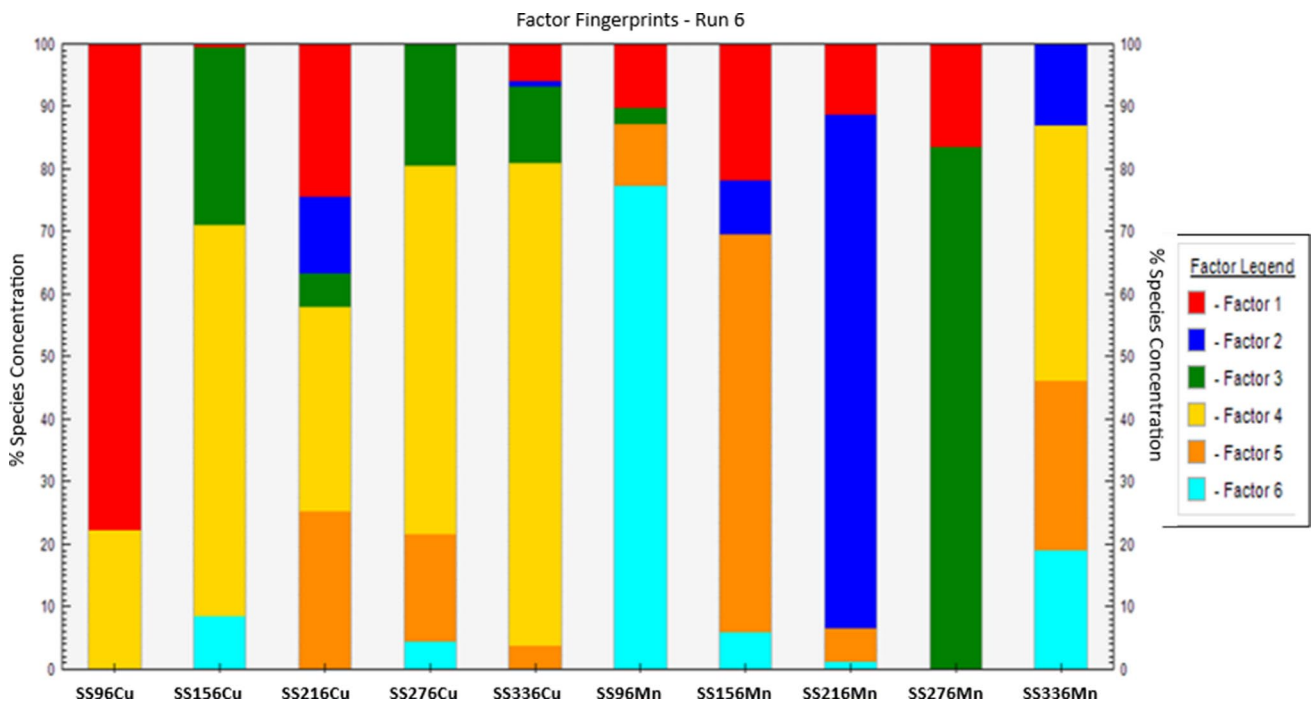

Fig. 9 Source fingerprint of different sample sizes from the PMF model analysis showing the percentage contributions of Cu and Mn. Description of Abbreviations: SS96Cu, SS156Cu, SS216Cu, SS276Cu, and SS336Cu represent sample sizes of 96. 156, 216, 276, and 336, respectively, for copper, while SS96Mn, SS156Mn, SS216Mn, SS276Mn, and SS336Mn represent sample sizes of 96, 156, 216, 276 and 336, respectively, for manganese
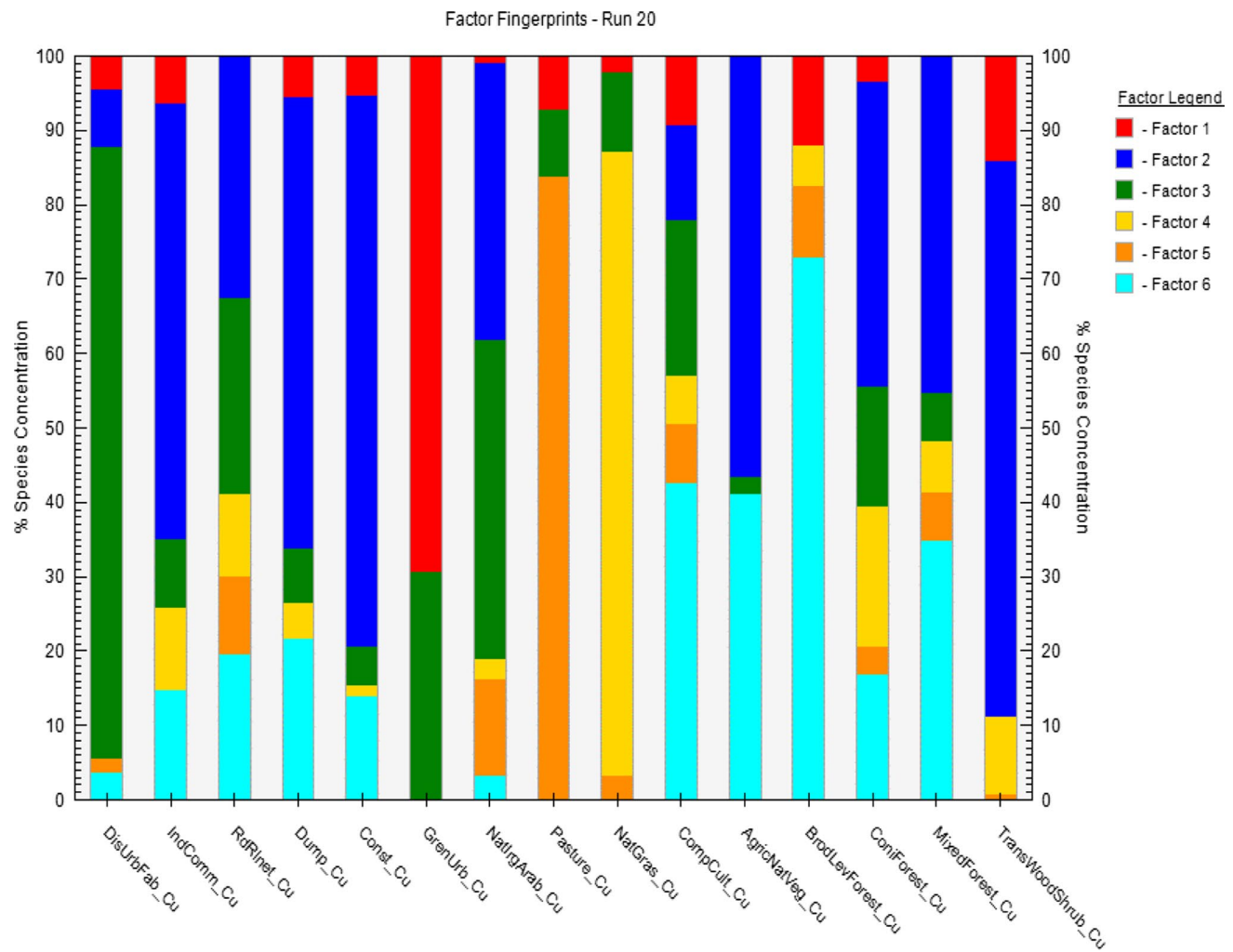
Fig. 10 Source fingerprint from the PMF model analysis showing the percentage contributions of Cu in each land use. Description of abbreviations; copper contents in: Discontinuous urban fabric = DisUrbFab_Cu, Industrial/commercial units = IndComm_Cu, Road&rail network area = RdRlnet_Cu, Dump sites = Dump_Cu, Construction sites = Const_Cu, Green urban area = GrenUrb_Cu, Natural Irrigated Arable area = NatIrgArab_Cu, Pastures = Pasture_Cu, Natural grasslands = NatGras_Cu, Complex cultivation area = CompCult_Cu, Agric & natural vegetation = AgricNatVeg_Cu, Broad-leaved forest = BrodLevForest_Cu, Coniferous forest = ConiForest_Cu, Mixed forest = MixedForest_Cu, Transition woodland shrubs = TransWoodShrub_Cu

biological translocation processes from the leaves. In addition, the broad-leaved trees have large buttress roots that could stop the removal and erosion of the PTEs in the soil when compared with the other forest tree types. Further, the amount of litter falls from the broad-leaved forest might have increased the concentration of CU and Mn in the soil (Solgi et al. 2018). This was also reported in some other studies where the soil PTE contents were highly increased in the broad-leaved forests than in the coniferous or tiny-leaved forest (Fu et al. 2010; Ma et al. 2015; Solgi et al. 2018; Zhou et al. 2018).

The correlation analysis for the five sample sizes and the PTEs revealed that manganese content in sample size 96 and copper content in sample size 216 had significant correlations across all the sample sizes and elements (Table 5).

Similarly, Mn content in sample size 336 significantly correlated with all the PTE contents in all the sample.

Figures (or values) in bold are significant at the 0.05 confidence level. Description of Abbreviations: SS96Cu, SS156Cu, SS216Cu, SS276Cu, and SS336Cu represent sample sizes of 96. 156, 216, 276, and 336, respectively, for copper, while SS96Mn, SS156Mn, SS216Mn, SS276Mn, and SS336Mn represent sample sizes of 96. 156, 216, 276, and 336, respectively, for manganese.

The influence of elevation on the PTEs in the sample sizes was remarkable as significant differences were recorded in the study (Table 6). The lowlands (≤ 500 m above sea level) showed significant p values for Cu and Mn concentrations in the different sample sizes except for SS156Cu and SS276Cu (Table 6). In the highland sites (> 500 m above sea level),
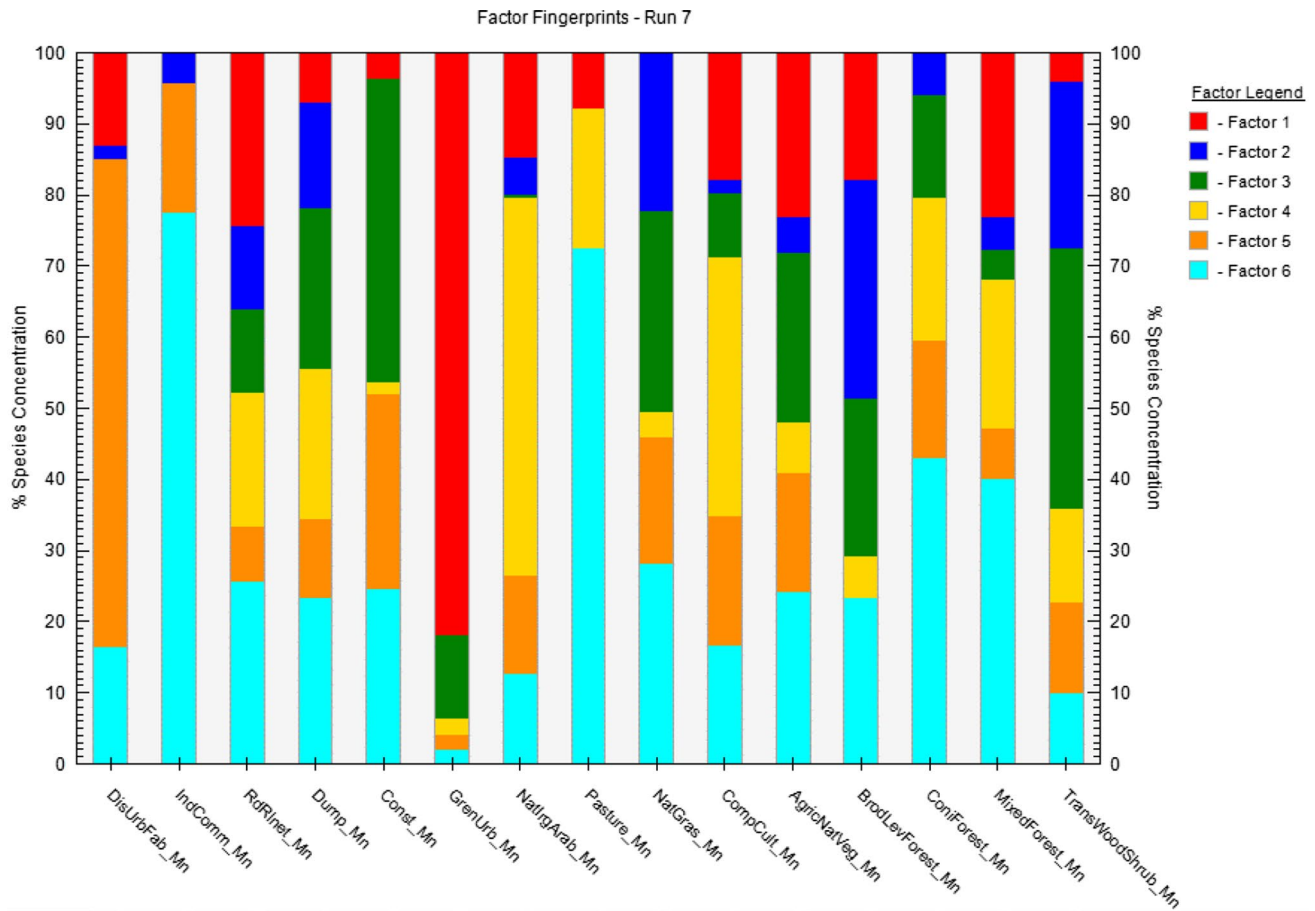
Fig. 11 Source fingerprint from the PMF model analysis showing the percentage contributions of Mn in each land use. Description of abbreviations; Manganese contents in: Discontinuous urban fabric = DisUrbFab_Mn, Industrial/commercial units = IndComm_Mn, Road&rail network area = RdRlnet_Mn, Dump sites = Dump_Mn, Construction sites = Const_Mn, Green urban area = GrenUrb_Mn, Natural Irrigated Arable area = NatIrgArab_Mn, Pastures = Pasture_Mn, Natural grasslands = NatGras_Mn, Complex cultivation area = CompCult_Mn, Agric & natural vegetation = AgricNatVeg_Mn, Broad-leaved forest = BrodLevForest_Mn, Coniferous forest = ConiForest_Mn, Mixed forest = MixedForest_Mn, Trans woodland shrubs = TransWoodShrub_Mn

Table 5 Summary of correlation analyses between the soils concentrations of Cu and Mn for the sample sizes

Parameter	SS96Cu	SS156Cu	SS216Cu	SS276Cu	SS336Cu	SS96Mn	SS156Mn	SS216Mn	SS276Mn	SS336Mn
SS96Cu	1.00									
SS156Cu	0.07	1.00								
SS216Cu	0.58	−0.15	1.00							
SS276Cu	0.00	0.56	0.58**	1.00						
SS336Cu	−0.04	0.09	0.61*	0.14	1.00					
SS96Mn	0.65*	0.71*	0.93*	0.54**	0.82*	1.00				
SS156Mn	0.47	0.00	0.81*	0.06	−0.37	0.48	1.00			
SS216Mn	0.32	−0.17	0.84*	−0.14	−0.31	0.42	0.68*	1.00		
SS276Mn	0.07	0.15	0.75*	0.00	−0.53	0.15	0.00	0.14	1.00	
SS336Mn	0.37*	0.01	0.65**	0.28*	−0.31	0.78*	0.70**	0.63*	0.53**	1.00

Description of Abbreviations: SS96Cu, SS156Cu, SS216Cu, SS276Cu, and SS336Cu represent sample sizes of 96,156, 216, 276, and 336, respectively, for copper, while SS96Mn, SS156Mn, SS216Mn, SS276Mn, and SS336Mn represent sample sizes of 96, 156, 216, 276, and 336, respectively, for manganese

*Correlation is significant at the 0.01

**Correlation is significant at the 0.05

Table 6 Summary of ANOVA for PTE concentrations for the different sample sizes in lowlands (≤ 500 m) and highlands (> 500 m)

Parameters	Elevation status (m)	F statistics	p value	Parameters	Elevation status (m)	F statistics	p value
SS96Cu	≤ 500	9.41	**< 0.001**	SS96Cu	> 500	0.58	0.73
SS156Cu	≤ 500	− 5.16	0.027	SS156Cu	> 500	16.71	0.08
SS216Cu	≤ 500	0.63	**< 0.001**	SS216Cu	> 500	23.55	0.51
SS276Cu	≤ 500	47.85	0.138	SS276Cu	> 500	0.64	0.31
SS336Cu	≤ 500	18.39	**< 0.001**	SS336Cu	> 500	− 11.38	0.02
SS96Mn	≤ 500	1.17	**< 0.001**	SS96Mn	> 500	− 4.92	0.55
SS156Mn	≤ 500	0.92	**< 0.001**	SS156Mn	> 500	54.70	0.89
SS216Mn	≤ 500	− 7.15	**< 0.001**	SS216Mn	> 500	3.59	0.00
SS276Mn	≤ 500	9.98	**< 0.001**	SS276Mn	> 500	7.13	0.07
SS336Mn	≤ 500	61.76	**< 0.001**	SS336Mn	> 500	− 6.88	0.03

The values in bold are statistically significant at 0.05 probability level

there were no significant differences in the observed outputs for the PTEs across the different sample sizes. The high concentration and significant differences seen among the lowland area might be related to the fact that industrial, commercial, and arable activities were concentrated in the lowland areas when compared with the highlands areas (Vannini et al. 2021). Agricultural activities could have applied reasonable quantities of fertilizers and animal manures, which might have increased the association between the PTEs and the sample sizes at the lowland soils (Nguyen et al. 2020; Xu et al. 2016).

Conclusion

The forest soils (0–30 cm) of the Frýdek mistek district of the Czech Republic were examined. Geostatistics approaches were used in this analysis to evaluate the spatial variability of PTEs in the soil and predict accuracy based on different sample sizes. Maps of Cu and Mn were developed using the GWR for the distribution. The models based on the obtained data formed strong spatial dependence patterns (nugget-to-sill ratio ranged from 0.00 to 0.24%), while Mn ($n = 336$) formed moderate spatial dependence patterns (nugget-to-sill ratio 0.55%); all other samples indicate weak spatial dependence pattern (nugget-to-sill ratio ranged from 0.77 to 62.77%). PTE concentrations were low for the studied elements in the district, indicating no pollution. The GWR coefficients for both Cu and Mn revealed a positive correlation between DEM towards the south-eastern part of the study area.

In addition to the sample sizes, land use and elevation to a large extent determined the distribution, variability and concentrations of Cu and Mn in the study area. For example, all the sample sizes showed the highest concentrations of the PTEs around the industrial, commercial, and the arable

land soils under the lowland below 500 m when compared with the highlands (above 500 m) where forests dominated.

It is important to emphasize that the concentrations of Cu and Mn in the soils of our study area were within the reference range recommended by FAO/WHO, EU, and/or US-EPA ,which are FAO/WHO (35–75 mg kg^{-1}), EU (50–140 mg kg^{-1}), and US-EPA (28–80 mg kg^{-1} average) for Cu, and FAO/WHO (20-10E04 mg kg^{-1}) and EU (524 mg kg^{-1} on average), and US-EPA (< 750 mg kg^{-1} average) for Mn, respectively. Although the rate of pollution per the mean values indicated non-polluted soils, there may be an intermittent need to assess the soils for control measures to be taken to curtail excessive accumulation and escalation to safeguard the well-being of the inhabitant and the ecosystem. Also, the results might support policy-developers in sustainable farming and forestry for the health of the ecosystem towards food security, forest safety, as well as animal and human welfare. The research also concluded that sample size does not necessarily influence the quantification of spatial variability and predictive accuracy of PTEs, rather other environmental factors such as covariates, land use, topography and the selection of appropriate models influencing the prediction and distribution of these elements in soils.

soil properties in the Czech Republic, supported by the Ministry of Agriculture of the Czech Republic.

Declarations

Conflict of interest The authors declared no conflict of interest in submitting, reviewing, and publishing this work.

References

Agyeman PC, Ahado SK, Kingsley J, Kebonye NM, Biney JKM, Borůvka L, Vašát R, Kocarek M (2021) Source apportionment, contamination levels, and spatial prediction of potentially toxic elements in selected soils of the Czech Republic. Environ Geochem Health 43(1):601–620

Ahmad K, Nazir MA, Qureshi AK, Hussain E, Najam T, Javed MS, Ashfaq M (2020) Engineering of Zirconium based metal–organic frameworks (Zr-MOFs) as efficient adsorbents. Mater Sci Eng B 262:114766

Ahmad K, Ashfaq M, Shah SSA, Hussain E, Naseem HA, Parveen S, Ayub A (2021a) Effect of metal atom in zeolitic imidazolate frameworks (ZIF-8 & 67) for removal of Pb^{2+} & Hg^{2+} from water. Food Chem Toxicol 149:112008

Ahmad W, Alharthy RD, Zubair M, Ahmed M, Hameed A, Rafique S (2021b) Toxic and heavy metals contamination assessment in soil and water to evaluate human health risk. Sci Rep 11(1):17006

Ahmed NK, El-Arabi AGM (2005) Natural radioactivity in farm soil and phosphate fertilizer and its environmental implications in Qena governorate, Upper Egypt. J Environ Radioact 84(1):51–64

Antoci A, Galeotti M, Sordi S (2018) Environmental pollution is an engine of industrialization. Commun Nonlinear Sci Numer Simul 58:262–272

Berg B, Ekbohm G, Söderström B, Staaf H (1991) Reduction of decomposition rates of Scots pine needle litter due to heavy-metal pollution. Water Air Soil Pollut 59(1):165–177

Borůvka L, Vacek O, Jehlička J (2005) Principal component analysis as a tool to indicate the origin of potentially toxic elements in soils. Geoderma 128(3–4):289–300

Borůvka L, Penížek V, Kozák J, Marhoul A M, Vacek O, Šrámek V, Žížala D, Poláková Š, Smatanová M, Fiala P (2018) Combining and harmonizing soil data from different sources: Problems and approaches. In: GlobalSoilMap-digital soil mapping from country to globe: proceedings of the global soil map 2017 conference, July 4–6, 2017, Moscow, Russia

Brown G (2012) Public participation GIS (PPGIS) for regional and environmental planning: Reflections on a decade of empirical research. J Urb Reg Inf Syst Assoc 24(2):7–18

Buss W, Graham MC, Shepherd JG, Mašek O (2016) Risks and benefits of marginal biomass-derived biochars for plant growth. Sci Total Environ 569:496–506

Cambardella CA, Moorman TB, Novak JM, Parkin TB, Karlen DL, Turco RF, Konopka AE (1994) Field-scale variability of soil properties in central Iowa soils. Soil Sci Soc Am J 58(5):1501–1511

Cao W, Li R, Chi X, Chen N, Chen J, Zhang H, Zhang F (2017) Island urbanization and its ecological consequences: a case study in the Zhoushan Island, East China. Ecol Indicat 76:1–14

Chabukdhara M, Nema AK (2013) Heavy metals assessment in urban soil around industrial clusters in Ghaziabad, India: probabilistic health risk approach. Ecotoxicol Environ Saf 87:57–64

Chen L, Lowenthal DH, Watson JG, Koracin D, Kumar N, Knipping EM, Wheeler N, Craig K, Reid S (2010a) Towards effective source apportionment using positive matrix factorization: experiments with simulated PM2.5 data. J Air Waste Manag Assoc 60:43–54

Chen X, Xia X, Zhao Y, Zhang P (2010b) Heavy metal concentrations in roadside soils and correlation with urban traffic in Beijing, China. J Hazard Mater 181:640–646

Cotrufo MF, De Santo AV, Alfani A, Bartoli G, De Cristofaro A (1995) Effects of urban heavy metal pollution on organic matter decomposition in *Quercus ilex* L. woods. Environ Pollut 89(1):81–87

Cressie N, Gotway CA, Grondona MO (1990) Spatial prediction from networks. Chemomet Intel Lab Syst 7(3):251–271

Dai L, Wang L, Li L, Liang T, Zhang Y, Ma C, Xing B (2018) Multivariate geostatistical analysis and source identification of heavy metals in the sediment of Poyang Lake in China. Sci Total Environ 621:1433–1444

Dragović R, Gajić B, Dragović S, Đorđević M, Đorđević M, Mihailović N, Onjia A (2014) Assessment of the impact of geographical factors on the spatial distribution of heavy metals in soils around the steel production facility in Smederevo (Serbia). J Clean Prod 84:550–562

Fallatah AM, Shah HUR, Ahmad K, Ashfaq M, Rauf A, Muneer M, Babras A (2022) Rational synthesis and characterization of highly water stable MOF@GO composite for efficient removal of mercury (Hg^{2+}) from water. Heliyon 8(10):e10936

Fu X, Feng X, Dong Z, Yin R, Wang J, Yang Z, Zhang H (2010) Atmospheric gaseous elemental mercury (GEM) concentrations and mercury depositions at a high-altitude mountain peak in south China. Atmos Chem Phys 10(5):2425–2437

Gholizadeh A, Borůvka L, Vašát R, Saberioon M, Klement A, Kratina J, Drábek O (2015) Estimation of potentially toxic element contamination in anthropogenic soils on a brown coal mining dumpsite with reflectance spectroscopy: a case study. PLoS ONE 10(2):e0117457

Guagliardi I, Cicchella D, De Rosa R (2012) A Geostatistical approach to assess concentration and spatial distribution of heavy metals in urban soils. Water Air Soil Pollut 223(9):5983–5998

Guo D, Fan Z, Lu S (2019) Changes in rhizosphere bacterial communities during remediation of heavy metal-accumulating plants around the Xikuangshan mine in southern China. Sci Rep 9:1947

Hani A, Karimineja MT (2010) Toxic metal distribution in soils of Kaveh Industrial city, Iran. World Appl Sci J 8(11):1333–1342

Hou Y, Lü Y, Chen W, Fu B (2017) Temporal variation and spatial scale dependency of ecosystem service interactions: a case study on the central Loess Plateau of China. Landsc Ecol 32(6):115991. https://doi.org/10.1016/j.envpol.2020.115991

Kabata-Pendias A, Szteke B (2015) Trace elements in abiotic and biotic environments. Taylor & Francis, Abingdon, p 468

Karami M, Afyuni M, Khoshgoftarmanesh AH, Papritz A, Schulin R (2009) Grain zinc, iron, and copper concentrations of wheat grown in central Iran and their relationships with soil and climate variables. J Agric Food Chem 57(22):10876–10882

Khademi H, Gabarrón M, Abbaspour A, Martínez-Martínez S, Faz A, Acosta JA (2019) Environmental impact assessment of industrial activities on heavy metals distribution in street dust and soil. Chemosphere 217:695–705

Kumar P, Rao B, Burman A, Kumar S, Samui P (2023) Spatial variation of permeability and consolidation behaviors of soil using ordinary kriging method. Groundw Sustain Dev 20:100856

Lee Y, Rojas-Perilla N, Runge M, Schmid T (2023) Variable selection using conditional AIC for linear mixed models with data-driven transformations. Stat Comput 33(1):27

Li M, Wang J (2023) The productivity effects of two-way FDI in China's logistics industry based on system GMM and GWR model. Ambient Intell Humaniz Comput 14(1):581–595

Li Z, Feng X, Li G, Bi X, Zhu J, Qin H, Dai Z, Liu J, Li Q, Sun G (2013) Distributions, sources, and pollution status of 17 trace

metal/metalloids in the street dust of a heavily industrialized city of central China. Environ Pollut 182:408–416

Li S, Song Y, Xu H, Li Y, Zhou S (2023) Spatial distribution characteristics and driving factors for traditional villages in areas of China based on GWR modeling and geodetector: a case study of the Awa mountain area. Sustainability 15(4):3443

Lichnovský J, Kupka J, Štěrbová V, Andráš P, Midula P (2017) Contamination of potentially toxic elements in streams and water sediments in abandoned Pb–Zn–Cu deposits (Hrubý Jeseník, Czech Republic). IOP Conf Ser Earth Environ Sci 92(1):012037

Ma M, Wang D, Du H, Sun T, Zhao Z, Wei S (2015) Atmospheric mercury deposition and its contribution of the regional atmospheric transport to mercury pollution at a national forest nature reserve, southwest China. Environ Sci Pollut Res 22(24):20007–20018

Mahar A, Wang P, Ali A, Awasthi MK, Lahori AH, Wang Q, Zhang Z (2016) Challenges and opportunities in the phytoremediation of heavy metals contaminated soils: a review. Ecotoxicol Environ Saf 126:111–121

Masto RE, George J, Rout TK, Ram LC (2017) Multi-element exposure risk from soil and dust in an industrial coal area. J Geochem Explor 176:100–107

Mazur Z, Radziemska M, Fronczyk J, Jeznach J (2015) Heavy metal accumulation in bioindicators of pollution in urban areas of northeastern Poland. Fresen Environ Bull 24(1):216–213

Melo LC, Puga AP, Coscione AR, Beesley L, Abreu CA, Camargo OA (2016) Sorption and desorption of cadmium and zinc in two tropical soils amended with sugarcane-straw-derived biochar. J Soils Sediment 16:226–234

Mueller U, Selia SRR, Tolosana-Delgado R (2023) Multivariate cross-validation and measures of accuracy and precision. Math Geosci 55:1–19

Nguyen TB, Do DD, Nguyen TX, Nguyen VN, Nguyen DT, Nguyen MH, Truong TT, Dong HP, Le AH, Bach O (2020) Seasonal, spatial variation, and pollution sources of heavy metals in the sediment of the Saigon River, Vietnam. Environ Pollut 256:113412

Norris G, Duvall R, Brown S, Bai S (2014) EPA positive matrix factorization (PMF) 5.0 fundamentals and user guide. U.S. Environmental Protection Agency, Office of Research and Development, Washington

Qiao P, Lei M, Guo G, Yang J, Zhou X, Chen T (2017) Quantitative analysis of the factors influencing soil heavy metal lateral migration in rainfalls based on geographical detector software: a case study in Huanjiang County, China. Sustain 9(7):1227

Onyedikachi UB, Belonwu DC, Wegwu MO (2018) Human health risk assessment of heavy metals in soils and commonly consumed food crops from quarry sites located at Isiagwu, Ebonyi State. Ovidius Univ Ann Chem 29(1):8–24

Qu X, Zhang J, Wang S (2017) On the stochastic fundamental diagram for freeway traffic: model development, analytical properties, validation, and extensive applications. Transp Res Part B Methodol 104:256–271

Sabiha-Javied S, Mehmood T, Chaudhry MM, Tufail M, Irfan N (2009) Heavy metal pollution from phosphate rock is used for the production of fertilizer in Pakistan. Microchem J 91(1):94–99

Sáňka M, Strnad M, Vondra J, Paterson E (1995) Sources of soil and plant contamination in an urban environment and possible assessment methods. Int J Environ Anal Chem 59(2–4):327–343

Sishah S, Abrahem T, Azene G, Dessalew A, Hundera H (2023) Downscaling and validating SMAP soil moisture using a machine learning algorithm over the Awash River basin, Ethiopia. PLoS ONE 18(1):e0279895

Solgi E, Keramaty M, Solgi M (2018) Biomonitoring of airborne Cu, Pb, and Zn in an urban area employing a broad-leaved and a conifer tree species. J Geochem Explor. https://doi.org/10.1016/j.gexplo.2019.106400

Spark W. Average weather in Frýdek-Místek, Czechia, year-round—weather spark, 2016.

Stone M (1974) Cross-validation and multinomial prediction. Biometrika 61(3):509–515

Sysalová J, Száková J (2006) Mobility assessment and validation of toxic elements in tunnel dust samples—subway and road using sequential chemical extraction and ICP-OES/GF AAS measurements. Environ Res 101(3):287–293

U.S. Environmental Protection Agency. Ecological soil screening levels for copper interim final. OSWER Directive 9285.7–68. 2007a. Washington, USA. https://www.epa.gov/sites/default/files/2015-09/documents/eco-ssl_copper.pdf.

U.S. Environmental Protection Agency. Ecological soil screening levels for manganese interim final. OSWER Directive 9285.7–68. 2007b. Washington, USA. https://www.epa.gov/sites/default/files/2015-09/documents/eco-ssl_manganese.pdf.

U.S. EPA. EPA positive matrix factorization (PMF) 3.0 model. Research Triangle Park: U.S. Environmental Protection Agency; 2010. Available online: http://www.epa.gov/heasd/products/pmf/pmf.html. Accessed 20 March 2022.

Vacek Z, Cukor J, Linda R, Vacek S, Šimůnek V, Brichta J, Prokůpková A (2020) Bark stripping, the crucial factor affecting stem rot development and timber production of Norway spruce forests in Central Europe. For Ecol Manag 474:118360

Vannini A, Tedesco R, Loppi S, Cecco VD, Martino LD, Nascimbene J, Dallo F, Barbante C (2021) Lichens as monitors of the atmospheric deposition of potentially toxic elements in high elevation Mediterranean ecosystems. Sci Total Environ 798:149369

Wang Y, Jodoin PM, Porikli F, Konrad J, Benezeth Y, Ishwar P. An expanded change detection benchmark dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2014, pp. 387–394.

Wang Z, Xiao J, Wang L, Liang T, Guo Q, Guan Y, Rinklebe J (2020) Elucidating the differentiation of soil heavy metals under different land uses with geographically weighted regression and self-organizing map. Environ Pollut 260:114065

Waseem A, Arshad J, Iqbal F, Sajjad A, Mehmood Z, Murtaza G (2014) Pollution status of Pakistan: a retrospective review on heavy metal contamination of water, soil, and vegetables. BioMed Res Int. https://doi.org/10.1155/2014/813206

Weissmannová HD, Pavlovský J, Chovanec P (2015) Heavy metal contaminations of urban soils in Ostrava, Czech Republic: assessment of metal pollution and using principal component analysis. Int J Environ Res 9(2):683–696

Wu W, Xie DT, Liu HB (2009) Spatial variability of soil heavy metals in the three gorges area: multivariate and geostatistical analyses. Environ Monitor Assess 157(1):63–71

Xiao J, Han X, Sun S, Wang L, Rinklebe J (2021) Heavy metals in different moss species in alpine ecosystems of Mountain Gongga, China: geochemical characteristics and controlling factors. Environ Pollut 272:115991

Xu X, Cao Z, Zhang Z, Li R, Hu B (2016) Spatial distribution and pollution assessment of heavy metals in the surface sediments of the Bohai and Yellow Seas. Mar Pollut Bull 110(1):596–602

Yang X, Shen Z, Zhang B, Yang J, Hong WX, Zhuang Z (2013) Silica nanoparticles capture atmospheric lead: implications in the

treatment of environmental heavy metal pollution. Chemosphere 90(2):653–656

Zhang S, Wang S, Liu N, Li N, Huang Y, Ye H (2011) Comparison of spatial prediction method for soil texture. Trans Chin Soc Agric Eng 27(1):332–339

Zhao H, Xia B, Fan C, Zhao P, Shen S (2012) Human health risk from soil heavy metal contamination under different land uses near Dabaoshan Mine, Southern China. Sci Total Environ 417:45–54

Zheng N, Liu J, Wang Q, Liang Z (2010) Health risk assessment of heavy metal exposure to street dust in the zinc smelting district, Northeast of China. Sci Total Environ 408(4):726–733

Zhou J, Wang Z, Zhang X (2018) Deposition and fate of mercury in litterfall, litter, and soil in coniferous and broad-leaved forests. J Geophys Res Biogeosci 123:2590–2603

Source apportionment, contamination levels, and spatial prediction of potentially toxic elements in selected soils of the Czech Republic

Prince Chapman Agyeman · Samuel Kudjo Ahado · John Kingsley · Ndiye Michael Kebonye · James Kobina Mensah Biney · Luboš Borůvka · Radim Vasat · Martin Kocarek

Abstract The sustenance of humans and livestock depends on the protection of the soil. Consequently, the pollution of the soil with potentially toxic elements (PTEs) is of great concern to humanity. The objective of this study is to investigate the source apportionment, concentration levels and spatial distribution of PTEs in selected soils in Frýdek-Místek District of the Czech Republic. The total number of soil samples was 70 (topsoil 49 and 21 subsoils) and was analysed using a portable XRF machine. Contamination factor and the pollution index load were used for the assessment and interpreting the pollution and distribution of PTEs in the soils. The inverse distance weighting was used for the spatial evaluation of the PTEs. The results of the analysis showed that the area is composed of low-to-high pollution site. PTEs displayed spatial variation patterns. The average PTE concentration decreases in this $Fe > Ti > Ba > Zr > Rb > Sr > Cr > Y > Cu > Ni > Th$ order for the topsoil and also decreases in this $Fe > Ti > Zr > Ba > Rb > Sr > Cr > Y > Cu > Ni >$ and Th order for the subsoil. These PTEs Cr, Ni, Cu, Rb, Y, Zr, Ba, Th, and Fe were far above the baseline European average value and the World average value level, respectively. The source apportionment showed the dominance of Cr, Ni, Rb, Ti, Th, Zr, Cu, Fe in the topsoil, while the subsoil was dominated by all the PTEs (factor 1 to 6) except Ba. The study concludes that indiscriminate human activities have an enormous effect on soil pollution.

Introduction

Potential toxic elements (PTEs) are a generic name that includes the phrase '*heavy metals*', '*trace elements*'*, and* '*toxic elements*' with a weight density more than or less than 5 g cm^{-3} (Fu et al. 2008; Ali et al. 2013; Fang et al. 2016; Anwakang 2018). PTEs are commonly present in almost all environmental matrices, including but not limited to soil, plants, and water. Similarly, PTEs originate naturally from rocks and mineral ores (Alloway 2013), from normal geological processes such as rock weathering and ore formation, and also from anthropogenic effects

P. C. Agyeman (✉) · S. K. Ahado · J. Kingsley · N. M. Kebonye · J. K. M. Biney · L. Borůvka · R. Vasat · M. Kocarek
Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamycka 129, 16500 Prague, Czech Republic
e-mail: agyeman@af.czu.cz

associated with urbanization, manufacturing activities, agricultural practices, mining, and extraction of natural resources (e.g. coal) (Alyazichi et al. 2017; Jones et al. 2019). As a result, their soil enrichment is always unclear until research has been conducted to assess the probability of elevated levels of toxicity. However, a lot of research work has been done on PTEs due to their potential health threats within society.

Generally, the rising rate of anthropocentric activities such as agriculture, industrialization, mining, and urbanization as a means of ensuring human survival has profoundly affected our soils. Anthropocentric activities can be traced back to the pre-medieval era, whereby most of the impacts exerted on soils were attributed to human actions than to natural sources (Brevik 2005). According to Anatolaki and Tsitouridou (2007) and Xu et al. (2014), industrialized and urbanized areas harbour pollutants released from a diverse range of sources (e.g. vehicular emissions, fugitive sources residential heating, industrial and natural activities). As a result, soils have been a sink for pollutants, although they play a central role in ensuring ecosystem balance as well as the basis for food production (Alyazichi et al. 2016).

According to Rehman et al. (2008), pollutants such as PTEs are discharged into the environment, often in the form of waste products from both point and non-point sources pollution. Although these PTEs may sometimes be released at low concentrations, their continued accumulation, non-biodegradable nature, and persistence may ultimately pose a high risk to the environment, and the potential for bioconcentration or biomagnification in plants and animal tissues (Ikem et al. 2003; Alyazichi et al. 2015). Khan et al. (2015) reported that an increased concentration of PTEs in the soil environment could harm the health of both plants and animals because of bioaccumulating effects. However, the World Environmental Agency bodies (e.g. EPA) have described PTEs as problematic in the environment (e.g. Ghana's mining site) and have a significant influence on soil quality (Sterckeman et al. 2000; Ordóñez et al. 2003; Alyazichi et al. 2015; Jones et al. 2019).

Many environmental factors (e.g. parent material, climatic factors, biotic factors, and anthropogenic activities) contribute spatially to soil variability (Zhu et al. 2001; McBratney et al. 2003; Minasny and McBratney 2016). These factors contribute to the assessment and determination of soil quality and promote the mobility, speciation, and distribution of PTEs (Ogundiran and Osibanjo 2009; Jones et al. 2019). Some natural factors, such as climate, slope, elevation, and parent material, are not the only factors that have an impact on development and soil chemical formation, but human factors, such as pollution source, vehicle emissions, agriculture, domestic waste, and industry, are eminent in soil pollution. (Yu et al. 2014). The anthropogenic effect is the key contributor to the spatial dispersion of PTEs rather than the natural accumulations (Poonam et al. 2014), and the excessive accumulation rate in the environment has been reported to be alarming (Dunea et al. 2016). PTEs accumulation in the soils has become a global problem (Sabiha-Javied et al. 2009; Yang et al. 2013) as it poses the greatest threat to human health (Chen et al. 2015; Alyazichi et al. 2015; Jones et al. 2019). According to Song et al. (2009), the inhibition of PTEs in soil is one of the feasible functions of studies. On the other hand, the spatial prediction of PTEs provides an avenue for identifying the distribution, concentration, occurrence, and probably the source of contamination of PTEs.

However, there is still knowledge gap in the pollution assessment processes (i.e. Why the exposure to these risk elements differs from the environment) as well as the advantages of positive matrix factorization (PMF) in source identification of these PTEs in Central European district (e.g. Frýdek-Místek) in agricultural soil actively engaged for crop production. Besides, the study location has been actively engaged in both industrial and farming activities, hence the need to understand the chemical composition of the prevalent pollutants and the spatial distribution over the area. Specifically, the study is aimed at investigating the source apportionment, contamination levels, and spatial prediction of potentially toxic elements in some selected soils in the Czech Republic.

Materials and methods

The study area

Frýdek-Místek is a district of the Moravian-Silesian Region, located in the foothills of the Moravian-Silesian Region in the Czech Republic, Europe. The district was formerly made up of two separate cities,

Silesian Frýdek and Moravian Místek. Both municipalities were merged in the year 1943, and the name has been used since 1955. The area under study is situated within the geographical coordinates Latitude 49° 41′ 0″ North and Longitude 18° 20′ 0″ East at an altitude ranging between 225 and 327 m above sea level, characterized by a cold temperate climate and a high amount of rainfall even in dry months (Fig. 1). In Frýdek-Místek, the summers are hot and partially cloudy, and the winters are cold, dry, windy, and mainly cloudy (Weather Spark 2016). Over the course of the year, temperatures usually range from 24 to 75° F and are rarely below 8 or above 86° F, while the average annual precipitation ranges between 685 and 752 mm (Weather Spark 2016). The area survey of the district is estimated at 1208 km^2 with 39.38% of the land size designated for agricultural activities and 49.36% for forest lands. The PTE pollution in the area is assumed to originate from atmospheric deposition emitted from the steel industry nearby, vehicular emission, abrasion from tires, and agricultural activities (e.g. pesticide and insecticide applications). Meanwhile, the dominant reference soil groups in the study area are cambisols and stagnosols. Cambisol dominates the Czech Republic with a percentage coverage of 56.7 and stagnosols 7.01% (Vacek et al. 2020).

Soil sampling and analysis

A grid sampling design was adopted for the soil investigation in this study. The soil samples were collected at intervals of 2 km using a handheld GPS system (Leica Zeno 5 GPS), and at every 4 km, the samples were collected for both topsoil (0 to 20 cm) and subsoil (21 to 40 cm), respectively. The total number (n) of samples obtained for both topsoil (ts) and subsoil (ss) is $n = 49$ and $n = 21$, respectively, across seven cities (Havirov, Terlicko, Trinec, Bystrica, Jablunkov, Mosty u Jablunkova, and Hrcava). They were taken into plastic bags and appropriately labelled and transported to the laboratory. The samples were air-dried, crushed by a mechanical device, and then sieved (< 2 mm) to obtain a pulverized sample. These samples were analysed using a portable X-ray fluorescence spectrometer. Each sample was measured in triplicates.

Quality assurance and quality control (QA/QC)

The quality assurance and control process, the standard reference material for a portable device (i.e. XRF 2711a NIST, the National Institute of Standards and Technology), was used in the analysis to ensure quality compliance. The reference material was occasionally measured alongside with soil samples to ensure that the analysis remained accurate until completion. The detection limits for the PTEs tested were < 10 mg/kg (Ni), < 10 mg/kg (Cu), < 5 mg/kg (Sr), < 20 mg/kg (Ba), < 5 mg/kg (Ti), < 10 mg/kg (Fe), < 10 mg/kg (Cr), < 5 mg/kg (Y), < 5 mg/kg (Zr), < 5 mg/kg (Th) and < 5 mg/kg (Rb). The PTEs recovery percentage was 82.3(Ni),89.9(Cu), 86.4(Sr), 88.1(Ba), 84.7(Ti), 87.9(Fe),81.2(Cr), 96.2(Y), 92.5(Zr), 100.9(Th), and 98.7(Rb).

PTEs contamination analysis

The PTEs pollution level of the study area was assessed through various contamination assessment indices, including the contamination factor (CF) and the pollution index (PLI).

a. Contamination factor

CF is defined as the ratio of the metal content of the sample to the background value of the same metal. It is given by,

$$CF = \frac{(C_m)\text{Sample}}{(C_m)\text{Background}} \tag{1}$$

where the C_m sample is the concentration of metal analysed from sampled soil and C_m Background is the geochemical background concentration. The categories used for interpretation of CF values are as follows: < 1 (No Pollution), 1–3 (Moderate Pollu­tion), 3–6 (Considerable Pollution), and > 6 (Very High Pollution). The baseline values used were the World average value (WAV) (Alina Kabata-Pendias 2011).

b. Pollution load index

This measure of estimation was proposed by Tomlinson et al. (1980) and is used to detect pollution that
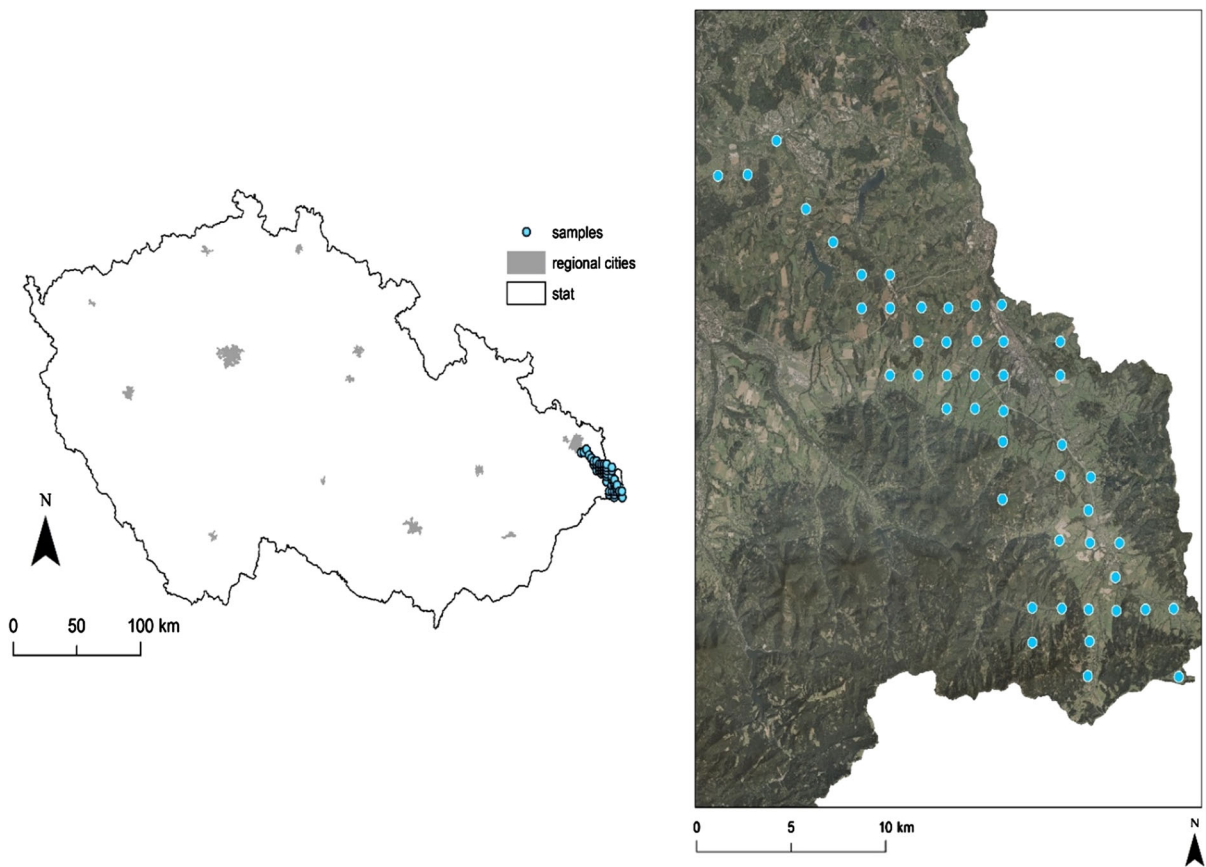
Fig. 1 Study area and sampling location map

permits a comparison of pollution levels amongst sites and at different times. The PLI was calculated as a result of the concentration factor of each PTE with regard to the background value in the soil where CF is the contamination factor mentioned above (Eq. 1) and n is the number of metals studied. PLI < 1 refers to the optimal soil quality, and PLI = 1 shows that only the baseline levels of contaminants are present, while PLI > 1 suggests the degradation of the quality of the site by Tomlinson et al. (1980).

The pollution load index (PLI) equation is given by

$$\text{PLI} = (\text{CF}_1 x \text{CF}_2 x \text{CF}_3 x . . . x \text{CF}_n)^{\frac{1}{n}} \tag{2}$$

Source apportionment

Positive matrix factorization (PMF), EPA–PMF v5.0 (U.S. EPA 2014), was employed to estimate the contribution of PTEs sources to contamination in the area. This mathematical technique is a receptor model used in calculating the contribution of the sources to samples built on the composition or fingerprints of the sources. PMF model apportions the collaborations of elements in soil composition by solving chemical mass balance such as:

$$C_{ij} = \sum_{k=1}^{p} G_{ik} + F_{kj} + E_{ij} \tag{3}$$

where C_{ij} represents the content of PTEs j in soil sample i, p is the number of factors (i.e. pollution sources), G_{ik} denotes contribution of factor k to soil sample i, F_{kj} is the content of PTEs j in factor k, and E_{ij} signifies the residual. Assume the matrix of PTEs contents in the soil samples; PMF can derive the factor contribution matrix G as well as factor profile matrix F by minimizing the objective function Q, which is presented as:

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{\varepsilon_{ij}}{u_{ij}} \right)^2 \qquad (4)$$

where m is the number of PTEs investigated, n denotes the number of soil samples, and U_{ij} signifies the uncertainty of PTEs j in soil sample i. U_{ij} is determined based on the PTEs content (C_{ij}), the relative standard deviation (σ), and the method detection limit (C_{MDL}). Therefore, it implies that the PTEs content is above C_{MDL} value; U_{ij} is computed as:

$$U_{ij} = \sqrt{\left(\sigma \times C_{ij} \right)^2 + C_{MDL}^2} \qquad (5)$$

The model recommends that the data below the detection limit would be substituted with the value of $C_{MDL}/2$, i.e. data that do not occur in this study, and the associated uncertainty is calculated as:

$$U_{ij} = 5/6 \, C_{MDL} \qquad (6)$$

Besides the constraint of no significant negative contribution (Gik), the utmost optimal factors were derived using the multilinear engine algorithm in PMF. It is vital to note that the minimum Q can be global or local. Consequently, multiple attempts using diverse starting points were carried out to reach the global minimum Q and reliable solution.

The uncertainty and the bias of the results obtained were further estimated using two error estimation methods encompassed in the PMF model: classical bootstrap (BS) and displacement of factor elements (DISP) (Brown et al. 2015). The uncertainty of solutions can arise generally based on three causes: (1) errors coming from the data set randomly, which are based on measurement procedures; (2) ambiguity resulting from rotation and the fact that multiple PMF solutions can have the similar or very close values of object function Q; and (3) modelling errors triggered by the simplification of the real system. Bootstrap was performed to curb random errors, whereas displacement was engaged to explore data errors and rotational ambiguity. The bootstrapping allowed us to create a new data set that randomly selects subsamples from the original observation with replacement. This set of data generated was then fitted by the PMF model to derive various factors and the concentration contribution. The fingerprint and the factor contribution of each element in the soil were assessed based on the results generated by the software from repeating the resampling and model fitting procedure. The

displacement analysis is repeatedly displaced to each potentially toxic element in the factor profile far enough from its fitted values and then fits the PMF model. The object function Q would then increase by a predetermined maximum value changed Q_{max}. The interpretation is based on the adjustment of the upper or lower interval of the displaced element. This analysis was conducted using the positive matrix factorization model (version 5.0) developed by the United States Environmental Protection Agency (USEPA).

Data analysis

The statistical analyses were performed using Excel (maximum and minimum number, average value), PMF EPA 5.0 for estimation of source apportionment, and RStudio for mapping as well as estimation of the Pearson correlation matrix. Inverse distance weighting (IDW) interpolation was used in computing the correlation and the similarities between neighbours with a proportional distance between them. This interpolation technique enabled us to create the spatial distribution maps of PTEs of the study area under investigation.

Results and discussion

PTEs concentration in soil

The descriptive statistics for the PTEs Fe, Ti, Ba, Zr, Rb, Sr, Cr, Y, Cu, and Ni in the topsoil and the subsoil from the study area listed in Table 1 display the average, maximum, minimum, and the percentage coefficient of variability. According to Karimi et al. (2015), the coefficient of variance (CV) indicates the degree of heterogeneity within PTE concentrations. If the CV ≤ 20% indicates low variability, 21% ≤ CV ≤ 50% is considered to be moderate variability, and 50% ≤ CV ≤ 100% indicates high variability, and if the CV is above 100%, it is considered to be extremely high variability. The coefficient of variation (CV%) of the PTEs in the current study area of the topsoil decreases in this order Ni < Fe < Cu < Y < Zr < Cr < Sr < Th < Ba < Rb < Ti accruing 63.47%, 52.25%, 43.68%, 31.59%, 30.53%, 22.96%, 18.45%, 18.01%, 17.86%, 15.98%, and 11.07%, respectively. The CV for the subsoil also decreases in this order Ni,

Table 1 Average concentration of PTEs from the study area and background level of toxic elements

	Cr mg/kg	Ni	Cu	Rb	Sr	Y	Zr	Ba	Th	Ti	Fe
Topsoil average	62.10	16.30	21.30	93.70	79.40	28.40	478.00	490.20	11.50	4664.40	23,380.59
Coef. var.% (topsoil)	22.96	63.47	43.68	15.98	18.45	31.59	30.53	17.86	18.01	11.07	52.25
Minimum (topsoil)	44.00	0.00	8.87	70.47	49.67	13.70	29.43	378.33	4.97	3133.67	11,788.00
Maximum (topsoil)	111.70	43.00	58.33	161.50	127.70	78.47	673.00	912.67	15.30	5560.00	75,987.67
Subsoil average (subsoil)	59.80	17.50	18.80	97.50	81.40	27.60	514.00	483.30	11.10	4863.00	21,154.02
Coef. var.% (subsoil)	20.50	55.35	30.13	21.55	18.37	17.33	26.11	16.99	18.42	13.32	24.41
Minimum (subsoil)	38.00	0.00	10.07	75.67	54.80	13.87	256.70	386.67	4.90	3356.67	16,160.00
Maximum (subsoil)	88.67	45.00	33.53	179.00	119.30	35.67	674.70	787.00	14.30	6724.00	37,026.67
*European average value	94.80	37.00	17.30	87.00	130.00	23.00	2.10	400.00	9.20	6070.00	–
*World average value	59.50	29.00	38.90	68.00	175.00	23.00	267.00	460.00	9.20	7038.00	–
**Crati Basin	90.54	34.67	44.36	104.66	234.25	24.99	209.36	603.05	–	0.73	5.47

*Kabata-Pendias (2011 pp. 41 and 42) **Guagliardi et al. (2012)

Cu, Zr, Fe, Rb, Cr, Th, Sr, Y, Ba, and Ti accounting for 55.35%, 30.13%, 26.11%, 24.41%, 21.55%, 20.50%, 18.42%, 18.37%, 17.33%, 16.99%, and 13.32%, respectively. The CV results suggest that there is a moderate variability between Cr, Cu, Y, and Zr in the topsoil and the subsoil (Cr, Cu, Rb, Zr, and Fe) (see Table 1). However, both (topsoil and subsoil) show more homogeneous variability. Ni showed high variability in both the topsoil and subsoil, whereas Fe showed high variability in the topsoil only. The spatial distribution of the non-homogeneity of Ni and Fe predicts the existence of a local source of enrichment. Comparing the findings from this present studies with the PTEs from Crati Basin (Guagliardi et al.2012), it is apparent that the Crati Basin is highly enriched in Cr, Ni, Cu, Rb, Sr, and Ba compared to the current studies. However, the current study is also enriched in Y, Zr, Ti, and Fe than in the Crati Basin.

The concentrations of the PTEs in the soils of the study area are shown in Tables 2 and 3, respectively. The average concentration of the PTEs from the highest to the lowest in the topsoil is as follows Fe > Ti > Ba > Zr > Rb > Sr > Cr > Y>Cu > Ni > Th and in the subsoil it follows in this order Fe > Ti > Zr > Ba > Rb > Sr > Cr > Y > Cu > Ni > and Th. Table 1 reveals that the average concentration of the PTEs (Cr, Cu, Rb, Y, Zr, Ba, Th) in the current study (topsoil) surpassed the World average value (WAV) and European average value (EAV) threshold limits. On the other hand, the subsoil

average values presented in Table 1 revealed that Cr, Rb, Y, Zr, Ba, and Th concentrations also exceeded the WAV and EAV limit. Out of the 49 topsoil samples analysed, some of the concentrations were far below the international standard limit (WAV), and the majority were far above the allowable limits. Chromium concentration measured (topsoil) indicated that 22 out of the 49 samples exceeded the WAV soil limits representing 44.9% of the total number of samples collected. Other PTEs such as Ni, Cu, Rb, Y, Zr, Ba, Th, and Fe (topsoil) also exceeded the WAV (8, 2, 49, 41, 49, 29, 43, and 3) representing 16.3%, 4.1%, 100%, 83.7%, 100%, 59.2%, 87.8%, and 6.1% of the total number of soil sample collected, respectively.

The concentration level of the PTEs revealed that several locations displayed high, moderate, and slightly contaminated levels (Tables 2 and 3) compared with the international limits (EAV and WAV). This may be attributed to diverse factors such as steel factory and agriculture activities (Zhao et al. 2007; Jones et al. 2019). The high concentration of Cr and Cu in the soil is consistent with the results obtained by Bi et al. (2006) from zinc smelting areas in Hezhang county, China. On the other hand, the concentration of Ba and Rb in the soil may be attributed to pedogenic factors and phosphate fertilizer application that accounted for the moderate increase compared to the international limits (Alina Kabata-Pendias and Mukherjee 2007). The PTEs average value concentrations of Cr, Cu, Y, Ba, Th, and Fe in the topsoil are

Table 2 PTEs concentration level in the topsoil

Sample	Cr mg/kg	Ni	Cu	Rb	Sr	Y	Zr	Ba	Th	Ti	Fe
1	67.33	5.67	16.23	83.07	81.27	28.90	601.67	443.33	11.37	4568.33	16,496.33
2	59.33	11.67	20.37	88.20	86.47	32.80	673.00	423.33	12.53	4470.00	16,877.00
3	51.67	15.00	19.20	82.10	80.00	27.90	647.33	451.67	9.73	4493.33	15,878.67
4	63.33	15.33	18.60	77.47	65.53	25.20	457.33	430.00	10.63	4638.33	16,011.67
5	57.67	12.67	19.93	92.50	81.43	31.90	638.00	475.00	12.93	4908.00	18,480.33
6	53.00	29.67	31.33	94.00	90.13	32.70	553.00	489.00	13.53	4682.67	22,481.67
7	54.00	12.00	17.53	86.97	82.37	30.60	615.33	417.67	13.27	4873.67	16,785.33
8	52.33	15.00	18.63	87.33	85.07	30.00	653.33	464.33	12.77	4790.67	17,920.33
9	57.33	22.33	19.00	98.33	90.07	32.20	507.67	509.67	13.30	4981.00	22,770.00
10	50.00	18.33	19.20	91.90	76.03	28.30	622.33	479.00	12.87	4944.67	22,881.00
11	56.67	15.33	18.10	92.87	79.13	28.90	592.33	485.00	13.30	5046.00	22,721.67
12	64.33	13.33	20.03	87.97	74.27	28.60	620.00	478.33	12.43	4760.00	19,550.00
13	83.33	15.00	29.33	89.33	70.10	27.80	555.67	490.33	13.63	4760.00	19,550.00
14	71.67	6.67	16.77	88.73	70.53	24.20	449.00	458.67	10.73	4863.33	33,403.67
15	111.70	17.00	14.67	111.10	65.20	26.20	204.67	912.67	9.67	5523.67	17,139.00
16	44.67	14.67	16.83	80.73	73.13	20.80	426.33	399.67	10.73	4734.00	20,128.67
17	88.33	36.33	18.87	109.90	113.80	26.30	304.00	554.00	8.33	3376.33	19,547.67
18	48.33	10.33	25.27	71.40	49.67	13.70	240.67	378.33	4.97	5345.33	28,212.00
19	63.67	17.00	29.67	70.47	73.10	21.50	490.67	518.67	11.87	3133.67	22,850.33
20	67.00	17.67	16.33	107.00	77.47	31.90	564.33	482.33	13.27	4290.33	50,875.33
21	45.33	0.00	12.80	95.63	58.43	18.10	355.33	434.67	6.73	4990.00	17,691.67
22	52.00	12.00	16.20	89.43	66.30	24.40	483.00	452.33	9.50	4665.00	21,907.00
23	49.00	7.67	15.60	85.57	67.53	27.70	497.33	458.00	9.53	4591.00	17,557.67
24	48.00	7.33	14.60	88.33	62.37	25.10	397.33	434.33	9.97	4172.33	20,900.33
25	58.00	21.67	19.23	93.47	74.73	27.20	502.00	474.33	12.37	4807.67	21,967.00
26	87.67	40.67	34.00	96.73	108.80	32.70	336.67	542.00	11.73	4437.00	34,046.00
27	60.33	17.33	21.33	91.10	77.43	29.30	623.00	480.67	13.40	5094.67	18,672.33
28	66.33	3.00	11.30	84.37	75.90	26.70	541.67	450.33	11.40	4720.33	16,726.33
29	74.33	9.00	16.90	87.13	83.30	29.90	637.33	433.67	13.53	4880.33	18,688.67
30	55.67	3.33	22.07	87.43	70.83	14.10	248.00	434.00	7.63	3162.33	21,782.67
31	88.00	34.67	26.77	111.10	108.40	35.30	388.67	562.00	11.97	4479.67	28,346.67
32	68.33	17.33	24.10	88.40	75.77	19.20	277.00	494.67	8.90	4801.00	21,439.33
33	54.33	9.00	58.33	91.33	71.80	29.10	572.67	432.00	9.07	4636.67	18,188.67
34	57.00	14.67	18.00	89.70	82.50	27.70	608.67	481.33	11.90	4849.67	19,277.00
35	55.67	6.33	16.83	88.40	60.10	26.20	497.00	483.67	11.60	5287.67	18,017.33
36	51.67	30.33	28.53	92.73	89.10	33.10	561.33	493.67	12.77	4917.33	22,360.67
37	61.33	11.00	19.10	121.40	73.40	30.10	559.67	548.67	13.90	5560.00	22,630.33
38	56.67	5.33	8.87	95.87	82.57	28.30	515.67	466.00	12.27	5195.00	16,335.67
39	63.67	20.33	19.27	87.50	92.97	36.20	655.67	491.67	15.30	4605.33	19,943.33
40	60.33	22.33	23.43	91.47	97.10	37.00	581.33	528.67	14.50	4887.67	25,840.33
41	81.00	39.33	57.00	93.87	95.13	31.40	478.33	667.00	11.57	5104.67	71,486.00
42	56.67	12.33	18.20	81.37	63.63	20.20	403.67	383.00	11.10	3768.00	12,473.00
43	51.33	23.00	20.40	129.80	74.27	24.90	308.33	493.00	10.80	4093.67	23,378.33
44	45.67	10.33	15.80	85.37	66.60	25.50	587.67	431.67	13.00	4598.33	18,348.67

Table 2 continued

Sample	Cr mg/kg	Ni	Cu	Rb	Sr	Y	Zr	Ba	Th	Ti	Fe
45	49.78	11.11	15.42	104.30	68.14	23.50	435.44	443.00	10.56	4957.67	18,846.33
46	91.00	43.00	30.67	109.40	127.70	33.90	270.33	534.67	9.57	4941.33	30,332.67
47	44.00	0.00	12.43	84.87	83.20	18.70	367.67	445.33	9.70	4656.00	11,788.00
48	74.00	29.67	23.77	161.50	89.50	26.60	273.00	704.67	13.03	3779.67	75,987.67
49	72.00	13.33	14.60	90.13	78.47	78.47	29.43	599.67	11.90	4734.00	20,128.67
Average values	62.14	16.25	21.25	93.65	79.40	28.39	477.75	490.20	11.45	4664.44	23,380.59

Table 3 PTEs concentration level in the subsoil

Sample	Cr mg/kg	Ni	Cu	Rb	Sr	Y	Zr	Ba	Th	Ti	Fe
1.00	62.33	14.33	26.30	75.67	54.80	13.87	256.70	386.67	4.90	3356.67	24,675.33
2.00	38.00	15.67	17.93	82.30	73.77	23.37	400.00	412.33	9.60	3671.67	19,120.00
3.00	56.33	15.67	18.43	107.80	78.63	33.50	572.00	473.00	12.10	4986.33	20,088.00
4.00	77.67	36.33	17.77	119.10	119.30	24.83	312.70	566.00	11.80	5473.67	28,377.67
5.00	56.33	18.33	18.53	95.47	90.67	32.07	506.70	504.00	9.90	4939.67	23,300.33
6.00	72.67	0.00	11.67	86.33	75.37	26.63	564.00	446.67	11.90	4620.67	17,263.00
7.00	45.67	17.00	10.97	99.50	84.10	29.60	548.30	451.33	12.90	4424.33	19,527.33
8.00	53.67	8.33	14.97	90.13	63.17	27.87	543.30	475.00	10.00	4903.00	19,261.00
9.00	50.00	16.00	15.27	86.73	81.60	28.83	672.70	473.00	13.40	5505.67	17,635.33
10.00	53.00	14.33	33.53	93.77	72.30	28.27	599.30	449.33	12.60	4636.67	18,188.67
11.00	60.00	13.67	19.03	79.90	82.70	27.83	636.70	447.67	11.00	4540.00	16,160.00
12.00	66.33	14.33	16.17	87.13	82.23	28.73	653.70	450.33	11.50	4698.67	17,486.67
13.00	50.33	19.00	20.60	91.60	79.10	30.63	653.30	446.67	14.00	5016.33	17,441.33
14.00	69.67	34.00	25.33	179.00	88.63	27.10	265.70	787.00	14.30	6724.00	37,026.67
15.00	50.00	17.33	16.93	86.30	69.80	24.63	586.30	434.67	10.50	4867.67	18,222.33
16.00	63.00	17.67	21.23	91.30	85.47	32.17	674.70	503.33	11.50	4888.33	18,470.00
17.00	47.67	5.67	11.03	87.93	72.97	23.23	495.00	455.67	9.83	4950.67	17,999.33
18.00	80.00	28.00	23.33	102.70	96.57	35.67	557.70	555.67	10.90	4789.67	23,150.67
19.00	88.67	45.00	24.77	113.90	114.70	32.50	288.70	554.67	11.20	5441.33	30,187.33
20.00	61.67	17.33	20.37	92.73	79.93	28.53	604.00	472.67	11.30	4985.33	23,937.00
21.00	52.33	–	10.07	97.90	63.57	20.13	410.30	404.33	7.87	4703.33	16,716.33
Average value	59.78	17.52	18.77	97.48	81.40	27.62	514.37	483.33	11.11	4863.03	21,154.02

higher than the PTEs in the subsoil; similarly, the samples average value concentrations of the PTEs (Ni, Rb, Sr, Zr, and Ti) in the subsoil are also higher than those same PTEs in the topsoil (Tables 2 and 3). The enrichment of the topsoil may be due to anthropogenic

activities such as atmospheric deposition, steel industry, and vehicular emission as well as agriculture (Jones et al. 2019). In contrast, the enrichment of the subsoil elements may be due to geogenic, pedogenic, and leaching. Oliva and Espinosa (2007) provided

evidence that collaborates with this paper's findings that topsoil enrichment is due to anthropogenic activities. In another vein, Li et al. (2017) agree with the results established in this study that potentially subsoil toxic metal enrichment is due to leaching, geogenic, and pedogenic influence.

Contamination factor and pollution load index for topsoil and subsoil

The results obtained indicate that the contamination factors (CF) values of most of the PTEs such as Zr, Rb, Th, Y, Ba, Cr, Fe, Ti, Ni, Cu, and Sr range from low-to-moderate contamination degree (0.00 to 3.00, see Table S2). These may be due to the impact of distinct external sources like industrial activities, agricultural activities, runoffs, and other anthropogenic inputs.

Yttrium was the only PTE that displayed severe contamination at a sampled point 49 (see Table S2). The CF for chromium at 49 observation points (for topsoil) indicated that 21 sampled locations were moderately contaminated. Out of 49 topsoil-sampled points, 9 observation points showed moderate contamination for Ni, 2 for Cu, Rb 49, Y 40, Zr 43, Ba 28, Th 43, and Fe 3 as displayed in the boxplot (Fig. 2). In the subsoil, 21 locations were sampled, and the PTEs

CF estimated showed moderate contamination in the following order Rb (21), Ba (19), Th (19), Ti (19), Zr (19), Cr(11, Ni (4), and Fe (1) as displayed in the boxplot (Fig. 3).

The pollution load index (PLI) was found to be generally very slightly contaminated to severe contamination (see Table S2 and Fig. 4). The results obtained here may be primarily attributed to the atmospheric deposition of PTEs due to the steel factory proximate to the pollution source and, likewise, the intensive agricultural activities in the area. These also confirm that there is a gradual rise of the PTEs concentration in the study area, especially around the observation point 41 (see Table S1) with PLI = of 0.656 mg/kg.

Comparison of baseline reference values to measured values

Figure S1 indicates that the measured average values are higher than that of the European baseline line average values. However, three out of the 49 sampled points were below the European baseline average value (see Table 1). Conversely, quite a number of the PTEs concentration values fell below the WAV baseline and a lot more above it (see Table 1). The
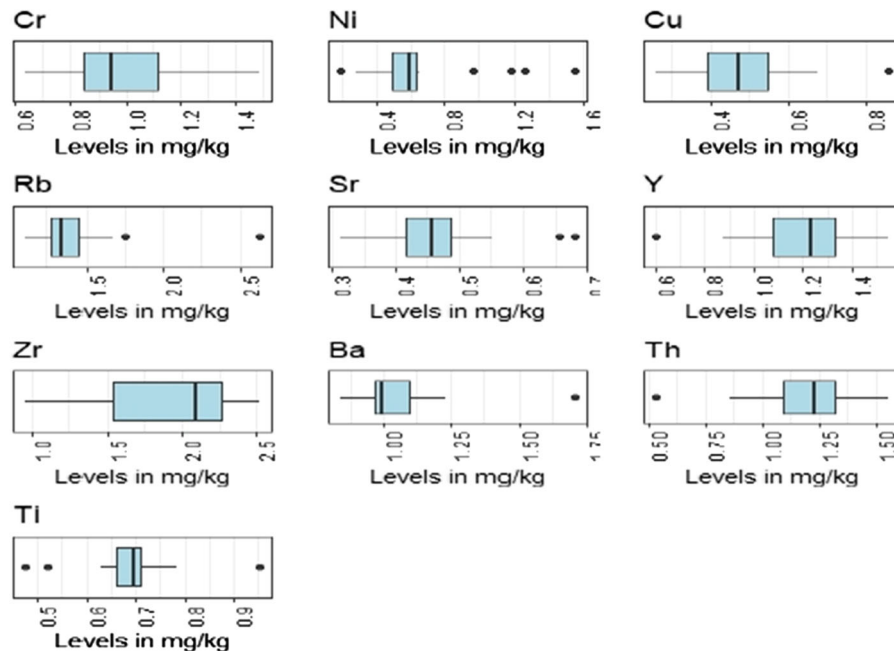
Fig. 2 A boxplot for the concentration factor of metals significance levels of topsoil
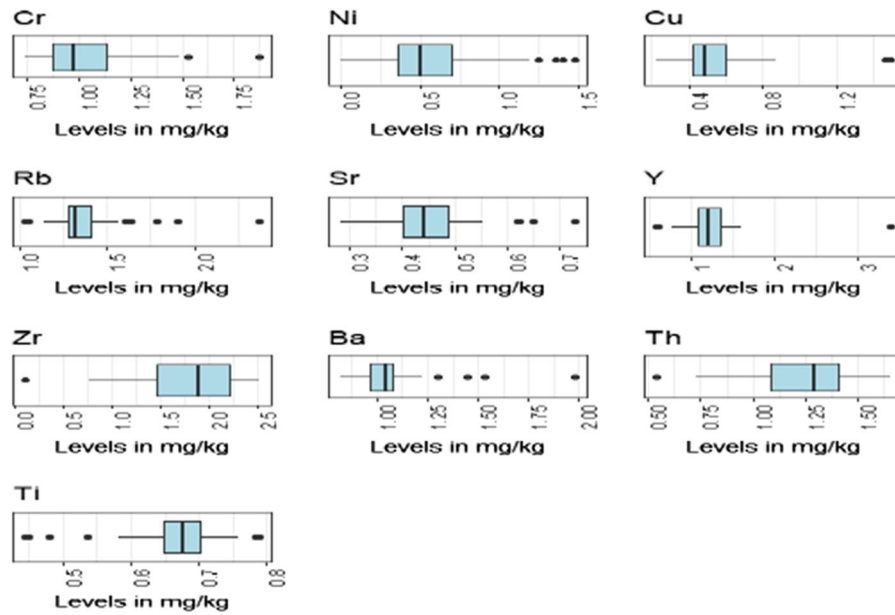
Fig. 3 A boxplot for the concentration factor of metals significance levels of subsoil
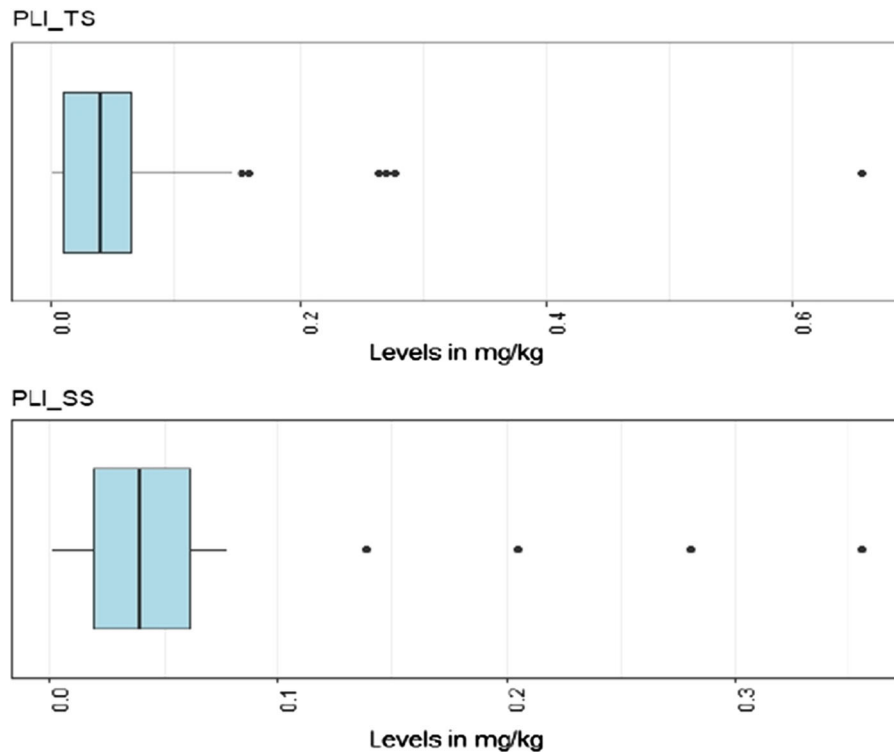

Fig. 4 A boxplot for the concentration of metals significance level for pollution load index in the topsoil (PLI_TS) and the subsoil (PLI_SS)

barium concentration in the soil was relatively high, and five measured values fell within the range 600 mg/ kg to 920 mg/kg. More so, most of the concentration of Ba in the studied agricultural soil was higher than

WAV and EAV (see Fig. S1), and this is coherent with other studies conducted by Zhang et al. (2018).

Figure S2 shows a quantifiable distribution of average chromium concentration values in the topsoil and the subsoil, which indicated that some measured Cr concentrations are slightly higher than the WAV baseline but lower than the EAV baseline (see Table 1). The average value concentration of chromium for topsoil is 62.14 mg/kg and for subsoil, it is 59.78 mg/kg. Juxtaposing the EAV baseline limits to the maximum concentration value of Cr in the topsoil (111.70 mg/kg, see Table S2), it is evident that the Cr maximum concentration is higher than the EAV baseline limit. Equally, some of the concentration values for Cr at some sampled location exceeded the EAV baseline.

Similarly, Fig. S3 presents a quantitative distribution of Cu in the topsoil and subsoil, compared with the EAV (see Table 1); some of the Cu concentration values were found to be higher than the EAV baseline threshold but lower than the WAV, respectively. Only 2 soil samples of the topsoil (58.3 mg/kg, 57 mg/kg) exceeded the WAV baseline.

Figure S4 shows the distribution of nickel (Ni) in the topsoil and the subsoil. Most of the concentration values of soil samples collected were below the WAV and EAV baseline (see Table 1). Only a subsoil sample and 3 topsoil samples collected exceeded both the EAV and WAV baseline (Fig. S4).

The displayed Fig. S5 illustrates the quantitative variation of Rubidium (Rb) in the top and the subsoil, which is in sharp contrast with the WAV and EAV baselines (see Table 1). Most of the concentration of Rb taken from the analysed soil samples was higher than WAV and EAV baselines. None of the Rb concentration values fell below the WAV baseline, but quite a few soils sampled dropped below the EAV baseline. A sample of the topsoil and subsoil recorded a very high concentration value (161.5 mg/kg, 179 mg/kg, respectively).

Figure S6 shows the concentration of strontium of the topsoil and subsoil from the study area. The strontium concentration level fell below both European and World average values of the World and European soil allowable limits (see Table 1). Strontium level in the soil of the study area was below the measurable limit of the European and the World soil threshold.

Thorium (Th) concentration level displayed in Fig. S7 shows that the concentration of Th from the study area mostly exceeded the EAV and WAV baseline (see Table 1). The thorium concentration level in the study area appears to be very high in both the topsoil and subsoil.

Figure S8 shows the concentration of titanium of the soils (topsoil and subsoil) from the study area, and it revealed that the concentration level of titanium was below both WAV and EAV baselines (see Table 1). Titanium concentration level, according to the EAV threshold, places the concentration in this present study under the acceptable limit, likewise the WAV baseline limit. Only a sample of the subsoil exceeded the EAV but below the WAV baseline.

Yttrium (Y) concentration level shown in Fig. S9 indicates that the quantitative distribution of Th from the study area mostly exceeded the EAV and WAV baseline. Yttrium concentrations level in the study area appears to be high in both topsoil and subsoil. One of the samples of the topsoil recorded a very high concentration level (78.47 mg/kg).

Figure S10 shows zirconium (Zr) concentration level and the quantitative distribution Zr from the study area. Most of the Zr concentrations level exceeded the EAV and WAV baseline (see Table 1). The zirconium concentration level in the study area appeared to be high in both topsoil and subsoil. Only four soil samples dropped below the WAV baseline, but the whole soil samples concentration level of Zr from this current study exceeded the EAV baseline.

Spatial distribution of PTEs

The spatial distribution of PTEs in the topsoil and the subsoil was evaluated with inverse distance weighting interpolation. This approach was adopted to estimate the distribution of enrichment and to detect hot spot areas of the PTEs concentration in the soil (see Figs. 5, 6, 7, 8) (Burgos et al. 2008). Figure 5 indicates the enrichment of Cu in the subsoil than the topsoil. The spatial distribution of Cu in subsoil indicates more spatial variability of Cu in the central region of the map spreading down the map. Enrichment and the variability of Cu in the peri-urban area, which is mainly agricultural land, and sites along with the steel industry, are consistent with Rauch and Pacyna (2009), and Liang et al. (2017) arguments that humans
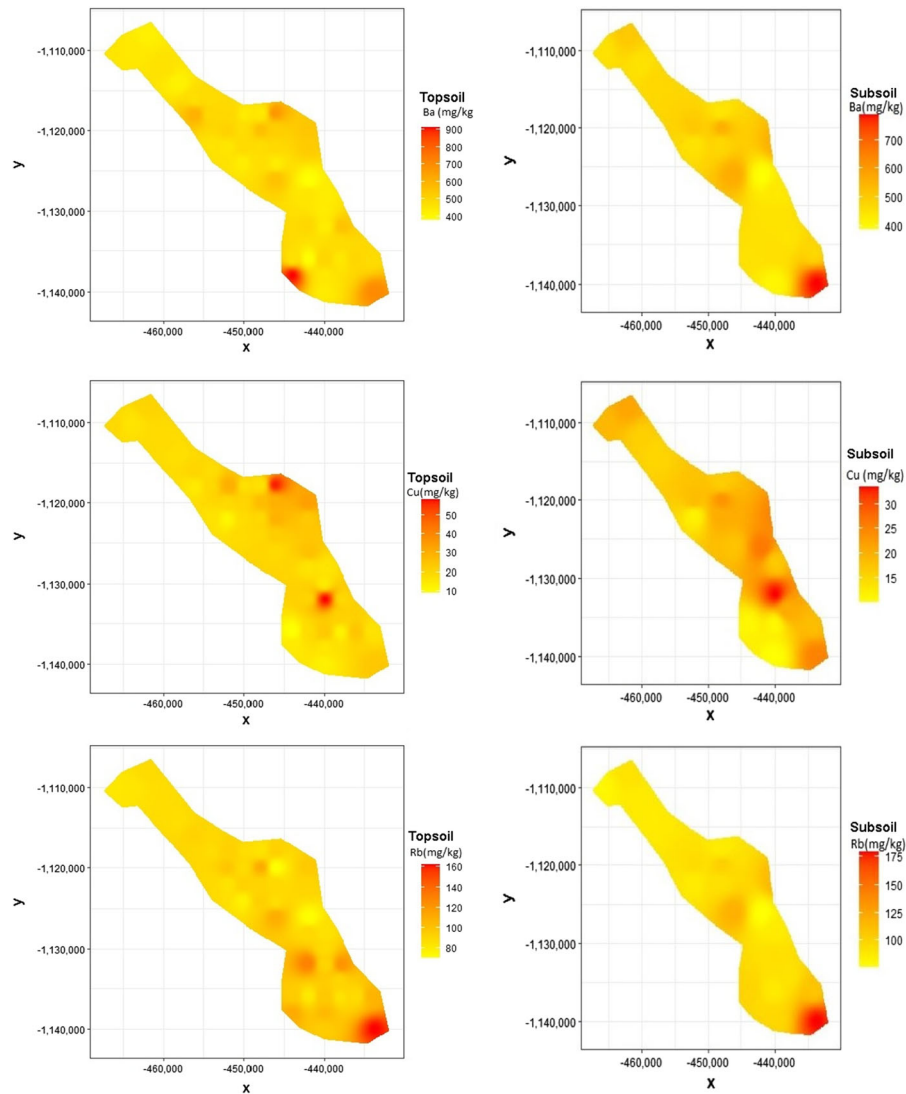
Fig. 5 Spatial distribution of Ba, Cu, and Rb (topsoil and subsoil)

activities such as sewage irrigation, fertilizer, and pesticide applications hugely impact Cu (see Fig. 5).

The variability of Rb and Ba concentrations in the soils was observed to be similar (see Fig. 5). The distribution of Ba confirms that the agricultural activity in the study area may have induced a high concentration of Ba in soil (Do Amaral Sobrinho et al. 2019). However, Ba and Rd have hot spots, but Rb concentrations may be due to both geological and climatic conditions (Kabata-Pendias and Pendias 1984). Similarly, Cr, Ni, and Sr showed regular spatial distributions over the study area (Fig. 6). Comparatively, the accumulation of Cr in the subsoil is more

than the topsoil. These may be attributed to both leaching and geogenic effects (Ferreira-Baptista and De Miguel 2005). For Cr, the central region of the map shows the hot spot in the subsoil and the southward direction indicates a relative hot spot of Cr in the topsoil. Furthermore, the spatial distribution pattern of Th, Ti, and Zr revealed similar severe distribution (Fig. 7). They are evenly distributed and suggest that Ti and Th concentrations in the soil are primarily from soil weathering. It was observed that the soils were predominantly cambisols with a leached B horizon generally developed from sedimentary rocks. This study's finding agrees with similar studies by
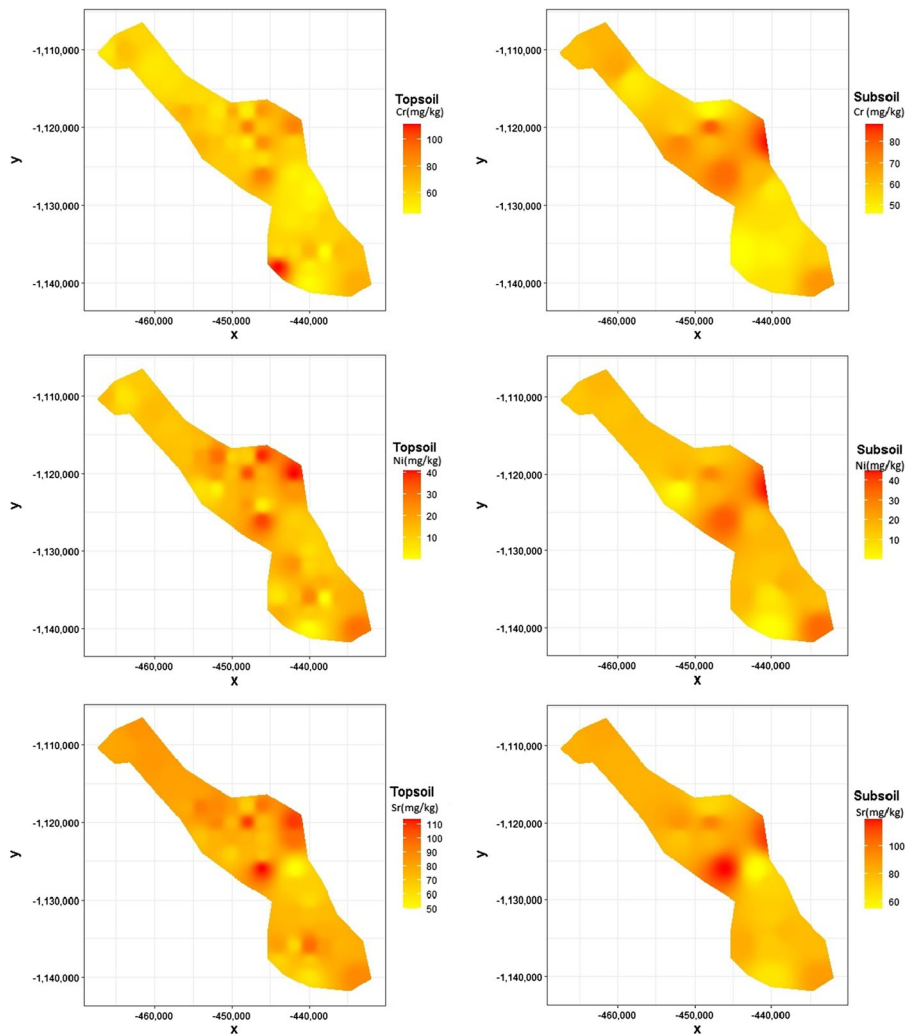
Fig. 6 Spatial distribution of Cr, Ni, and Sr (topsoil and subsoil)

Sartandel et al. (2009). Zirconium showed high concentration both in the topsoil and in the subsoil, respectively. The enrichment of Zr was high compared to the EAV benchmark level and, in most cases, exceeded the WAV benchmark level in quite a several sampled areas. Even though Zr is used as an immobile element (Bain et al. 1994), it shows a variant propensity as well.

The distribution of Y (Fig. 8) showed a marginal distribution with high enrichment in the topsoil. Distribution of Y concentrations in the soil may be attributed to the natural background of the prevailing mineral rocks. The distribution of Fe is mild, in the topsoil with a patchy hot spot, whereas in the subsoil, it shows a proportional distribution at the central region

and a hot spot at the southward direction (Fig. 8). In general, the spatial distribution pattern of the concentration levels of the PTEs indicates that their distribution is mostly heterogeneous with reference to their diverse sources.

Pearson correlation matrix

Figure SF11 (topsoil) and Fig. SF12 (subsoil) indicate the correlation matrix between the investigated PTEs, showing the relationship that exists between them.

The relationship of the PTEs in the topsoil (SF 11) revealed that there is a strong positive correlation between Ni and Sr, likewise, Cr and Ba with $r = 0.76$ and 0.73, respectively. Zr and Th showed a strong
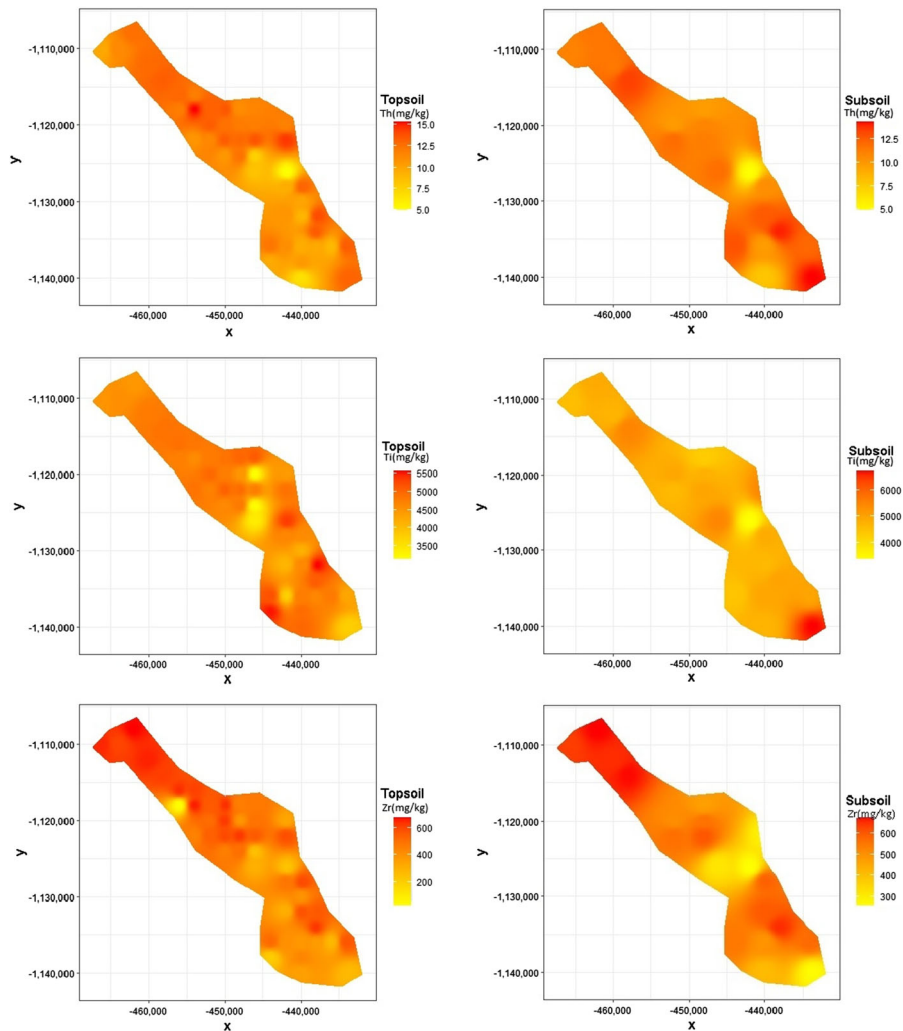
Fig. 7 Spatial distribution of Th, Ti and Zr (topsoil and subsoil)

correlation amongst each other, and similarly, Rb and Ba, Cr and Ni, Ni and Fe, Ni and Cu as well as Rb and Fe all revealed a stronger positive correlation amongst each other. PTEs such as Ni and Sr, Cr and Ba, Zr and Th, Rb and Ba, Cr and Cu, Ni and Fe as well as Rb have shown strong positive correlations between themselves and are likely to have the same or closely related roots, whereas those with weak or negative associations are likely to have different origins. The highest correlation ($r = 0.76$) was observed for Ni and Sr, whereas the lowest was found between several PTEs such as Fe and Th ($r = 0.11$). The results suggest that Ni and Sr, Cr and Ba, Zr and Th, Rb and Ba, Cr and Cu, Ni and Fe probably shared familiar sources,

while the sources of Fe and Th as well as the others are poorly correlated.

The correlation matrix for the PTEs in the subsoil is presented in Fig. SF12. Rb and Ba, Rb and Fe, Ba and Ti, Ba and Fe, as well as Ni and Sr, showed a strong positive correlation between each other. The weak correlation was observed between Th and Fe. The elements in the subsoil showing the high relationship are probably sharing the same source and vice versa for the weak and the negatively correlated PTEs. The highest correlation in the subsoil has $r = 0.93$, which is between Rb and Ba. Besides the correlation in the subsoil, the correlation matrix in the subsoil appears to be stronger than the correlation in the topsoil (Fig. S12
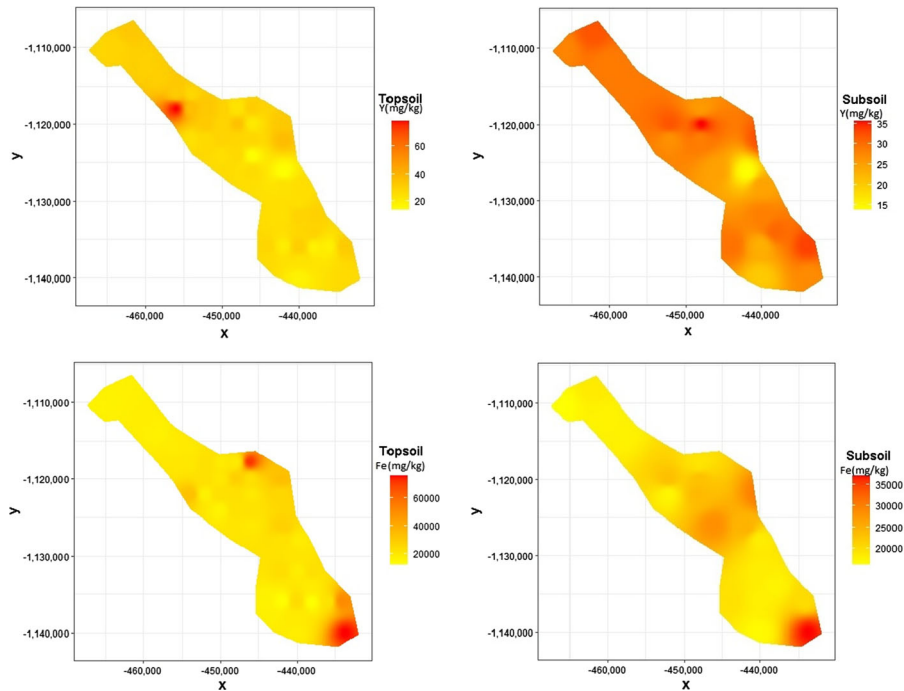
Fig. 8 Spatial distribution of Y and Fe (topsoil and subsoil)

and Fig. S13). This probably suggests that these PTEs are more closely related in the subsoil than the topsoil.

Source apportionment by PMF model

The study area is currently polluted by the activities of the steel industry, productive agriculture that covers 39.38% of the total land area (1208 km^2), vehicular emission, biological waste (animal excreta), and domestic waste. The earlier discussed PMF model provides a better way of identifying the source of PTEs in the selected soils. The rationality of the analysis is premised on minimum Q to control the residual matrix that determines a reasonable number of factors. The PMF analysis in this paper discharged six factors and run for 20 runs. Figures 9 and 10 display the source contribution per every PTEs.

Factor 1 was characterized by Cr and Ni with a factor loading of 27.8% and 35.1% in the topsoil, whereas in the subsoil, Rb, Th, and Ti accrues factor loads of 36.4%, 31.7%, and 31.7%, respectively. The concentration of Cr and Ni in the topsoil may be attributed predominantly to agriculture activities and geogenic source. Moreover, a study conducted by Manta et al. (2002) argues that Cr and Ni in the topsoil

in Italy were primarily inherited from geogenic sources. Besides that, other studies carried out by Mamat et al. (2014) in Yanqi Basin soils agree with the Manta et al. (2002) assertion, which supports this current study.

The study area has been traditionally associated with agricultural activities, and Saha et al. (2011) opined that the topsoil enrichment of Cr is due to agricultural materials such as slaked lime-NPK. In addition, Zhang et al. (2016) conducted comprehensive research gathering 464 papers on the concentration of Cr in China, and the findings indicated that the higher concentration of Cr in agricultural soil (Cr agricultural soil 78.94 mg/kg) was well above baseline (57.30 mg/kg) levels. Even though the geogenic source cannot be undermined, the anthropogenic pollution of Cr is eminent and increasing daily. The Cr average concentration (62.14 mg/kg) in this present study falls below the Europeans average soil concentration baseline value (94.80 mg/kg). On the other hand, it exceeds the World average (59.5 mg/kg) and the upper continental crust Cr concentration (35.0 mg/kg). It is quite interesting to note that during an acute water shortage, farming communities may create an avenue and resort to using raw sewage to
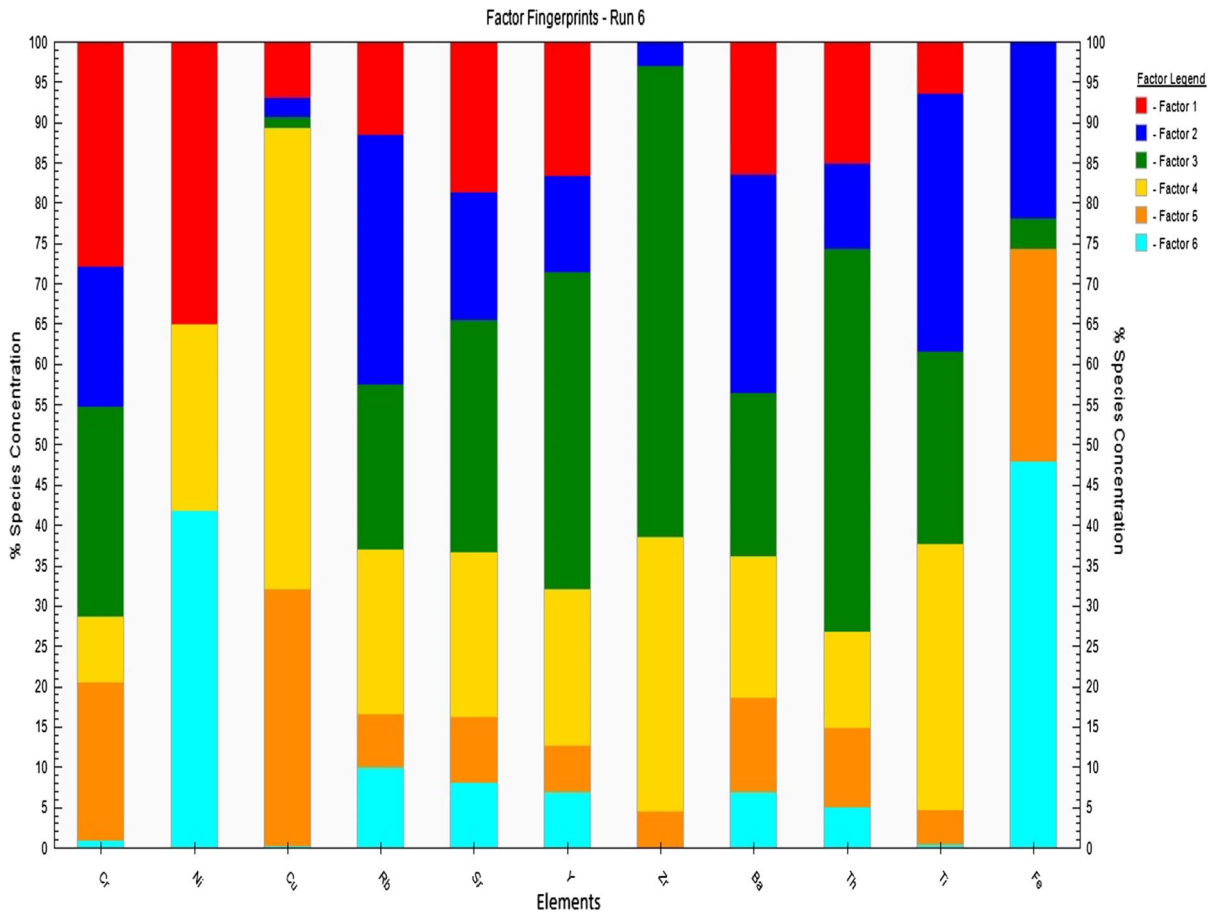
Fig. 9 Source fingerprint of the topsoil from PMF model analysis showing percentage contribution of PTEs

irrigate and fertilize cropland, which is another way to upsurge the concentration of Cr in the soil as reported Lente et al. (2012).

Agri-anthropogenic-related source forms part of the soil pollution in this area of study due to the large presence of farming that covers 39.8% of the total land area. Nevertheless, the steel industry in Frýdek-Místek is also another hub that deposits Ni and Cr into the soil. This is consistent with exposition made elsewhere by Chen et al. (2016) that non-ferrous metal and steel in the Gansu province in China have provided evidence on a few studies showing that Cr and Ni emanate from industrial activities like steel industry, mining, and metal processing factories.

Titanium, thorium, and rubidium, on the other hand, dominated the subsoil accruing 31.7%, 31.7%, and 36.4% factor loadings, respectively. Zirconium and titanium presence in agricultural soil primarily is traceable to the parent material, which is affirmed by

research carried out by Taboada et al. (2006). Notwithstanding, other research conducted by Hamby and Tynybekov (2002) as well as Négrel et al. (2018) proclaims that thorium and rubidium which exist in agricultural and shoreline areas are of natural occurrence such as weathering.

Factor 2 was controlled by Rb and Ti in the topsoil, whereas Cr and Fe are the major dominating elements in the subsoil. Both PTEs in the soil accounted for the following factor loadings 31.1%, 32%, 39.2%, and 41.5%, respectively. Fe and Cr form a great relationship with $r = 0.59$, which presupposes that they probably share a common source (that is, steel industry). Even though Fe is abundant, its content in the subsoil general comes from the natural occurrence; nonetheless, according to Luo et al. (2015), its content in the soil can be altered significantly by human activities such as the steel industry.
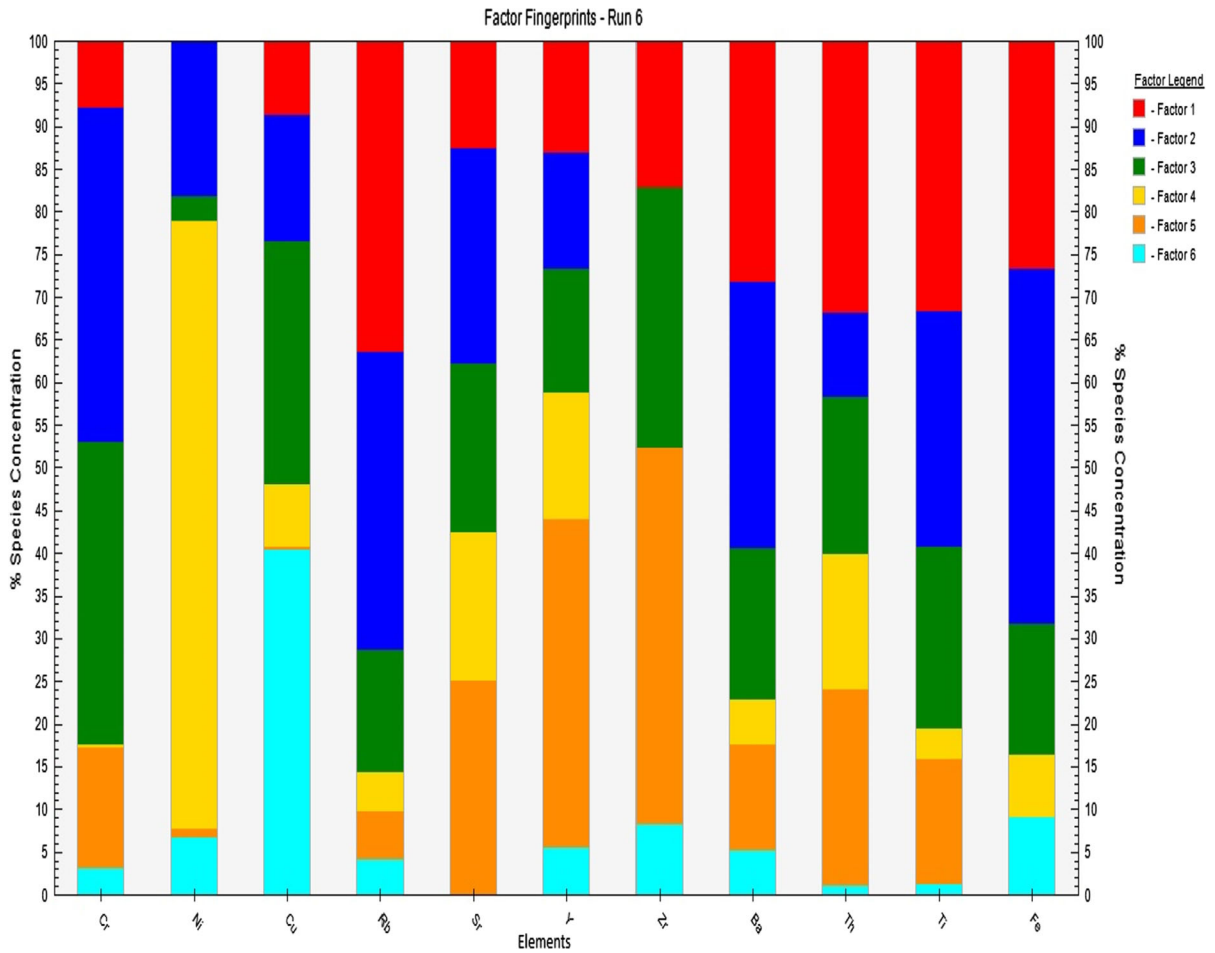
Fig. 10 Source fingerprint of the subsoil from PMF model analysis showing percentage contribution of PTEs

The 3rd factor in the loading in the topsoil and subsoil with high percentage factor contributions is Th (47.5%), Zr (58.4) topsoil and Cr (35.4%), and Zr (30.4%) subsoil. Some of these element sources come from a multiplicity of sources, especially Cr (Jones et al. 2019). The content of Zr that dominates the topsoil and subsoil in centrally controlled by the parent materials that is consistent with the research conducted by Taboada et al. (2006). Thorium occurrence is not entirely different; its high concentration in the soil is of a natural process.

The 4th factor load consists of Cu (57.3%) and Zr (34.1%) from the topsoil and Ni (71.1%) and Sr (17.3%) from the subsoil. Nickel shows a very high percentage in the subsoil; its high percentage is as a result of the diversity of sources such as industrial and geogenic source accounting for its high concentration

(Jones et al. 2019). Copper dominated the topsoil with more than half-percentage in this group. It sources primarily associated with livestock (Holm 1990; Cheng et al. 2014; Liang et al. 2017) and serve as an antibacterial agent to the gut (Rosen and Roberts 1996).

Factor 5 accounts for Cu and Fe dominating the topsoil, and Y as well as Zr constituted the highest amount in the subsoil. Cu accrued 31.8% and Fe 26.4%, whereas Y accounted for 38.5% as well as Zr 44.2%, respectively. The relatively high source of Y is due to their natural occurrence (Luo et al. 2015).

The last factor (6th) produced Ni and Fe as the dominant elements in the topsoil, whereas Cu as well as Fe also had high factor loadings. Fe accrued 47.9% in the topsoil and garnered 9.1% in the subsoil as well

as Cu and Ni, accounting for 40.5% and 41.8%, respectively.

Conclusion

This present study brings the attention into data concerning potentially toxic elements, concentration levels in different soil levels, source apportionment, and the spatial distribution of the toxic elements in agricultural exposed land. The identified PTEs such as Cr, Ni, Cu, Rb, Y, Zr, Ba, Th, and Fe exceeded the World average value (WAV) and European average value (EAV) soil limits from low concentration to a high concentration level for the topsoil and also the subsoil. Furthermore, the area is characterized by a mild to hot spot areas by these PTEs.

The PTEs' average value concentration in the topsoil was higher than the elements in the subsoil. Consequently, the average value concentration of the following PTEs Ni, Rb, Sr, Zr, and Ti in the subsoil was higher than those same elements in the topsoil. Furthermore, most of the soil samples fell within a moderately contaminated level and some at the brink of dropping in the significantly enriched level. Despite the various contamination assessment indices estimated and spatial distribution of PTEs analysis, agricultural production, industrial activities, vehicular emission, atmospheric deposition, and geogenic sources were identified as the sources of PTEs accumulation in the studied soil.

The source apportionment showed the dominance of Cr, Ni, Rb, Ti, Th, Zr, Cu, and Fe in the topsoil, while in the subsoil, all the PTEs at the factor loadings dominated from factors 1 to 6 apart from Ba. Therefore, it is essential to highlight that there should be periodic control as well as a monitoring process to put structures in place to discourage the release of these PTEs into soils in other to safeguard the health of humans, flora, and fauna. It is incumbent to articulate from the finding ascertain that the soil health in the study area is under threat and therefore calling for attention and actions for the reduction and the prevention of PTEs contamination, which serve as an imminent threat to the environment.

Hence, it is highly recommended, in the future, further research to evaluate the ecological and health risk of potentially toxic elements in these soils be evaluated.

Compliance with ethical standards

Conflict of interests The authors declare that they have no known competing personal interests or relationships that could have appeared to influence the scientific work in this manuscript.

References

Ali, H., Khan, E., & Sajad, M. A. (2013). Phytoremediation of heavy metals—Concepts and applications. *Chemosphere, 91*(7), 869–881.

Alloway, B. J. (2013). *Sources of heavy metals and metalloids in soils* (pp. 11–50). Dordrecht: Springer.

Alyazichi, Y. M., Jones, B. G., & McLean, E. (2015). Source identification and assessment of sediment contamination of trace metals in Kogarah Bay, NSW, Australia. *Environmental Monitoring and Assessment, 187*(2), 20.

Alyazichi, Y. M., Jones, B. G., & McLean, E. (2016). Lead isotope fingerprinting used as a tracer of lead pollution in marine sediments from Botany Bay and Port Hacking estuaries, southern Sydney, Australia. *Regional Studies in Marine Science, 7,* 136–141.

Alyazichi, Y. M., Jones, B. G., McLean, E., Pease, J., & Brown, H. (2017). Geochemical assessment of trace element pollution in surface sediments from the Georges River, Southern Sydney, Australia. *Archives of Environmental Contamination and Toxicology, 72*(2), 247–259.

Anatolaki, C., & Tsitouridou, R. (2007). Atmospheric deposition of nitrogen, sulfur and chloride in Thessaloniki, Greece. *Atmospheric Research, 85*(3–4), 413–428.

Anwakang, J. (2019.) Evaluation of concentration of heavy metals in surface and ground water in akpabuyo, southeastern nigeria, wa.*WORLD Journal of advance healthcare research*.

Bain, D. C., Mellor, A., Wilson, M. J., & Duthie, D. M. L. (1994). Chemical and mineralogical weathering rates and processes in an upland granitic till catchment in Scotland. *Water, Air, and Soil pollution, 73*(1), 11–27.

Bi, X., Feng, X., Yang, Y., Qiu, G., Li, G., Li, F., et al. (2006). Environmental contamination of heavy metals from zinc smelting areas in Hezhang County, western Guizhou, China. *Environment International, 32*(7), 883–890.

Brevik, E. (2005). A brief history of soil science. In *Land cover and land use, encyclopedia of life support* (Vol. 28, Issue 2).

Brown, S. G., Eberly, S., Paatero, P., & Norris, G. A. (2015). Methods for estimating uncertainty in PMF solutions: Examples with ambient air and water quality data and

guidance on reporting PMF results. *Science of the Total Environment*, *518*, 626–635.

Burgos, P., Madejón, E., Pérez-de-Mora, A., & Cabrera, F. (2008). Horizontal and vertical variability of soil properties in a trace element contaminated area. *International Journal of Applied Earth Observation and Geoinformation, 10*(1), 11–25.

Chen, H., Teng, Y., Lu, S., Wang, Y., & Wang, J. (2015). Contamination features and health risk of soil heavy metals in China. *Science of the Total Environment, 512–513*, 143–153.

Chen, T., Chang, Q., Liu, J., Clevers, J. G. P. W., & Kooistra, L. (2016). Identification of soil heavy metal sources and improvement in spatial mapping based on soil spectral information: A case study in northwest China. *Science of the Total Environment, 565,* 155–164.

Cheng, Q., Guo, Y., Wang, W., & Hao, S. (2014). Spatial variation of soil quality and pollution assessment of heavy metals in cultivated soils of Henan Province, China. *Chemical Speciation and Bioavailability, 26*(3), 184–190.

Do Amaral Sobrinho, N. M. B., Ceddia, M. B., Zonta, E., De Souza, C. D. C. B., & Lima, E. S. A. (2019). Barium and lead levels in sites for disposal of oil well waste. *Revista Caatinga, 32*(4), 1060–1068.

Dunea, D., Iordache, S., Radulescu, C., Pohoata, A., & Dulama, I. D. (2016). A multidimensional approach to the influence of wind on the variations of particulate matter and associated heavy metals in Ploiesti city, Romania. *Romanian Journal of Physics, 61*(7–8), 1354–1368.

EPA, E. (2014). Positive matrix factorization (PMF) 5.0-fundamentals and user guide. Washington: US Environmental Protection Agency.

Fang, H., Huang, L., Wang, J., He, G., & Reible, D. (2016). Environmental assessment of heavy metal transport and transformation in the Hangzhou Bay, China. *Journal of Hazardous Materials, 302,* 447–457.

Ferreira-Baptista, L., & De Miguel, E. (2005). Geochemistry and risk assessment of street dust in Luanda, Angola: A tropical urban environment. *Atmospheric Environment, 39*(25), 4501–4512.

Fu, J., Zhou, Q., Liu, J., Liu, W., Wang, T., Zhang, Q., et al. (2008). High levels of heavy metals in rice (*Oryza sativa* L.) from a typical E-waste recycling area in southeast China and its potential risk to human health. *Chemosphere, 71*(7), 1269–1275.

Guagliardi, I., Cicchella, D., & De Rosa, R. (2012). A geostatistical approach to assess concentration and spatial distribution of heavy metals in urban soils. *Water, Air, and Soil pollution, 223*(9), 5983–5998.

Hamby, D. M., & Tynybekov, A. K. (2002). Uranium, thorium, and potassium in soils along the shore of lake Issyk-Kyol in the Kyrghyz Republic. *Environmental Monitoring and Assessment, 73*(2), 101–108.

Holm, A. (1990). *E. coli* associated diarrhoea in weaner pigs: Zinc oxide added to the feed a preservative measure. Cabdirect.Org.

Ikem, A., Egiebor, N. O., & Nyavor, K. (2003). Trace elements in water, fish and sediment from Tuskegee Lake, Southeastern USA. *Water, Air, and Soil pollution, 149*(1–4), 51–75.

Jones, B. G., Alyazichi, Y. M., Low, C., Goodfellow, A., Chenhall, B. E., & Morrison, R. J. (2019). Distribution and sources of trace element pollutants in the sediments of the industrialized Port Kembla Harbour, New South Wales, Australia. *Environmental Earth Sciences, 78*(12), 357.

Kabata-Pendias, A. (2011). Trace elements in soils and plants. In *Trace Elements in Soils and Pl*ants, Fourth Edition.

Kabata-Pendias, A, & Pendias, H. (1984). Trace elements in soil and plants. http://www.sidalc.net/cgibin/wxis.exe/?IsisScript=UACHBC.xis&method=post&formato=2&cantidad=1&expresion=mfn=080164.

Khan, A., Khan, S., Khan, M. A., Qamar, Z., & Waqas, M. (2015). The uptake and bioaccumulation of heavy metals by food plants, their effects on plants nutrients, and associated health risk: A review. *Environmental Science and Pollution Research, 22*(18), 13772–13799.

Lente, I., Keraita, B., Drechsel, P., Ofosu-Anim, J., & Brimah, A. K. (2012). Risk assessment of heavy-metal contamination on vegetables grown in long-term wastewater irrigated urban farming sites in Accra, Ghana. *Water Quality, Exposure and Health, 4*(4), 179–186.

Li, J. S., Beiyuan, J., Tsang, D. C. W., Wang, L., Poon, C. S., Li, X. D., et al. (2017). Arsenic-containing soil from geogenic source in Hong Kong: Leaching characteristics and stabilization/solidification. *Chemosphere, 182,* 31–39.

Liang, J., Feng, C., Zeng, G., Gao, X., Zhong, M., Li, X., et al. (2017). Spatial distribution and source identification of heavy metals in surface soils in a typical coal mine city, Lianyuan, China. *Environmental Pollution, 225,* 681–690.

Luo, X. S., Xue, Y., Wang, Y. L., Cang, L., Xu, B., & Ding, J. (2015). Source identification and apportionment of heavy metals in urban soil profiles. *Chemosphere, 127,* 152–157.

Mamat, Z., Yimit, H., Ji, R. Z. A., & Eziz, M. (2014). Source identification and hazardous risk delineation of heavy metal contamination in Yanqi basin, northwest China. *Science of the Total Environment, 493,* 1098–1111.

Manta, D. S., Angelone, M., Bellanca, A., Neri, R., & Sprovieri, M. (2002). Heavy metals in urban soils: A case study from the city of Palermo (Sicily), Italy. *Science of the Total Environment, 300*(1–3), 229–243.

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*(1–2), 3–52.

Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma, 264,* 301–311.

Kabata-Pendias, Alina, & Mukherjee, A. B. (2007). Trace elements from soil to human. In *Trace elements from soil to human*.

Négrel, P., Ladenberger, A., Reimann, C., Birke, M., & Sadeghi, M. (2018). Distribution of Rb, Ga and Cs in agricultural land soils at European continental scale (GEMAS): Implications for weathering conditions and provenance. *Chemical Geology, 479,* 188–203.

Nezhad, M. T. K., Tabatabaii, S. M., & Gholami, A. (2015). Geochemical assessment of steel smelter-impacted urban soils, Ahvaz, Iran. *Journal of Geochemical Exploration*, *152*, 91–109.

Ogundiran, M. B., & Osibanjo, O. (2009). Mobility and speciation of heavy metals in soils impacted by hazardous waste. *Chemical Speciation and Bioavailability, 21*(2), 59–69.

Oliva, S. R., & Espinosa, A. J. F. (2007). Monitoring of heavy metals in topsoils, atmospheric particles and plant leaves to identify possible contamination sources. *Microchemical Journal, 86*(1), 131–139.

Ordóñez, A., Loredo, J., De Miguel, E., & Charlesworth, S. (2003). Distribution of heavy metals in the street dusts and soils of an industrial city in Northern Spain. *Archives of Environmental Contamination and Toxicology, 44*(2), 160–170.

Poonam, R. B., Sharma, R., Handa, N., Kaur, H., Kaur, R., Sirhindi, G., & Thukral, A. K. (2014). Prospects of field crops for phytoremediation of contaminants. In *Emerging technologies and management of crop stress tolerance* (Vol. 2, pp. 449–470). Elsevier Inc.

Rauch, J. N., & Pacyna, J. M. (2009). Earth's global Ag, Al, Cr, Cu, Fe, Ni, Pb, and Zn cycles. *Global Biogeochemical Cycles, 23*(2), GB2001.

Rehman, W., Zeb, A., Noor, N., & Nawaz, M. (2008). Heavy metal pollution assessment in various industries of Pakistan. *Environmental Geology, 55*(2), 353–358.

Rosen, G. D., & Roberts, P. A. (1996). *Comprehensive survey of the response of growing pigs to supplementary copper in feed*.

Sabiha-Javied, S., Mehmood, T., Chaudhry, M. M., Tufail, M., & Irfan, N. (2009). Heavy metal pollution from phosphate rock used for the production of fertilizer in Pakistan. *Microchemical Journal, 91*(1), 94–99.

Saha, R., Nandi, R., & Saha, B. (2011). Sources and toxicity of hexavalent chromium. *Journal of Coordination Chemistry, 64*(10), 1782–1806.

Sartandel, S. J., Jha, S. K., Bara, S. V., Tripathi, R. M., & Puranik, V. D. (2009). Spatial distribution of uranium and thorium in the surface soil around proposed uranium mining site at Lambapur and its vertical profile in the Nagarjuna Sagar Dam. *Journal of Environmental Radioactivity, 100*(10), 831–834.

Song, B., Lei, M., Chen, T., Zheng, Y., Xie, Y., Li, X., et al. (2009). Assessing the health risk of heavy metals in vegetables to the general population in Beijing, China. *Journal of Environmental Sciences, 21*(12), 1702–1709.

Sterckeman, T., Douay, F., Proix, N., & Fourrier, H. (2000). Vertical distribution of Cd, Pb and Zn in soils near smelters in the North of France. *Environmental Pollution, 107*(3), 377–389.

Taboada, T., Cortizas, A. M., García, C., & García-Rodeja, E. (2006). Particle-size fractionation of titanium and zirconium during weathering and pedogenesis of granitic rocks in NW Spain. *Geoderma, 131*(1–2), 218–236.

Tomlinson, D.L., Wilson, J.G., Harris, C.R., Jeffrey, D.W., 1980. Problems in the assessment of heavy-metal levels in estuaries and the formation of a pollution index. *Helgoländer Meeresuntersuchungen*, 33, 566–575.

Vacek, O., Vašát, R., & Borůvka, L. (2020). Quantifying the pedodiversity-elevation relations. *Geoderma, 373,* 114441.

Weather Spark. (2016). Average weather in Frýdek-Místek, Czechia, year round—Weather spark. https://weatherspark.com/y/83671/Average-Weather-in-Frýdek-Místek-Czechia-Year-Round.

Xu, X., Zhao, Y., Zhao, X., Wang, Y., & Deng, W. (2014). Sources of heavy metal pollution in agricultural soils of a rapidly industrializing area in the Yangtze Delta of China. *Ecotoxicology and Environmental Safety, 108,* 161–167.

Yang, X., Shen, Z., Zhang, B., Yang, J., Hong, W. X., Zhuang, Z., et al. (2013). Silica nanoparticles capture atmospheric lead: Implications in the treatment of environmental heavy metal pollution. *Chemosphere, 90*(2), 653–656.

Yu, H., Ni, S. J., He, Z. W., Zhang, C. J., Nan, X., Kong, B., et al. (2014). Analysis of the spatial relationship between heavy metals in soil and human activities based on landscape geochemical interpretation. *Journal of Geochemical Exploration, 146,* 136–148.

Zhang, A., Cortes, V., Phelps, B., van Ryswyk, H., & Srebotnjak, T. (2018). Experimental analysis of soil and mandarin orange plants treated with heavy metals found in oilfield-produced wastewater. *Sustainability, 10*(5), 1493.

Zhang, X., Zhong, T., Liu, L., Zhang, X., Cheng, M., Li, X., et al. (2016). Chromium occurrences in arable soil and its influence on food production in China. *Environmental Earth Sciences, 75*(3), 1–8.

Zhao, Y. F., Shi, X. Z., Huang, B., Yu, D. S., Wang, H. J., Sun, W. X., et al. (2007). Spatial distribution of heavy metals in agricultural soils of an industry-based peri-urban area in Wuxi, China1 1 project supported by the RURBIFARM (sustainable farming at the rural–urban interface) project of the European Union (No. ICA4-CT-2002-10021). *Pedosphere, 17*(1), 44–51.

Zhu, A. X., Hudson, B., Burt, J., Lubich, K., & Simonson, D. (2001). Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal, 65*(5), 1463–1472.

Modeling and Assessing the Spatial and Vertical Distributions of Potentially Toxic Elements in Soil and How the Concentrations Differ

Samuel Kudjo Ahado [1], Chukwudi Nwaogu [2,3,*] , Vincent Yaw Oppong Sarkodie [1] and Luboš Borůvka [1]

[1] Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamýcká 129, 16500 Prague, Czech Republic; ahados@af.czu.cz (S.K.A.); oppong_sarkodie@af.czu.cz (V.Y.O.S.); boruvka@af.czu.cz (L.B.)

[2] Department of Environmental Management, Federal University of Technology, Owerri, P.M.B. 1526, Owerri 460114, Nigeria

[3] Department of Forest Protection and Entomology, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamýcká 129, 16500 Prague, Czech Republic

* Correspondence: cnwaogu@gmail.com

Abstract: A healthy soil is a healthy ecosystem because humans, animals, plants, and water highly depend upon it. Soil pollution by potentially toxic elements (PTEs) is a serious concern for humankind. The study is aimed at (i) assessing the concentrations of PTEs in soils under a long-term heavily industrialized region for coal and textiles, (ii) modeling and mapping the spatial and vertical distributions of PTEs using a GIS-based ordinary kriging technique, and (iii) identifying the possible sources of these PTEs in the Jizerské Mountains (Jizera Mts.) using a positive matrix factorization (PMF) model. Four hundred and forty-two (442) soil samples were analyzed by applying the aqua regia method. To assess the PTE contents, the level of pollution, and the distribution pattern in soil, the contamination factor (CF) and the pollution load index load (PLI) were applied. ArcGIS-based ordinary kriging interpolation was used for the spatial analysis of PTEs. The results of the analysis revealed that the variation in the coefficient (CV) of PTEs in the organic soil was highest in Cr (96.36%), followed by Cu (54.94%) and Pb (49.40%). On the other hand, the mineral soil had Cu (96.88%), Cr (66.70%), and Pb (64.48%) as the highest in CV. The PTEs in both the organic soil and the mineral soil revealed a high heterogeneous variability. Though the study area lies within the "Black Triangle", which is a historic industrial site in Central Europe, this result did not show a substantial influence of the contamination of PTEs in the area. In spite of the rate of pollution in this area being very low based on the findings, there may be a need for intermittent assessment of the soil. This helps to curtail any excessive accumulation and escalation in future. The results may serve as baseline information for pollution assessment. It might support policy-developers in sustainable farming and forestry for the health of an ecosystem towards food security, forest safety, as well as animal and human welfare.

Keywords: heavy metals; positive matrix factorization; contamination factor; pollution load index; GIS-kriging

1. Introduction

Soil is an indispensable component of an ecosystem that directly or indirectly links and maintains the Earth's four spheres (namely the lithosphere, biosphere, hydrosphere, and atmosphere). However, this essential potential of soil has, in recent times, been threatened by heavy metals or potentially toxic elements (PTEs). Interestingly, the chemistry of soil makes it vulnerable to high concentrations of heavy metals or PTEs. At a required concentration, most PTEs such as Cr, Cu, Fe, Mn, Zn, Ni, Mo, Co, Se, and others are essential for plants, animals, or humans [1,2]. The presence of PTEs in soil has been attracting reasonable attention because of their ecological and biological risks. Several

studies in different biomes have been performed to identify the sources of PTEs in the soils [3–16].

Natural phenomena and anthropogenic activities are the two major sources that determine the concentrations of PTEs in soils [17–24]. Natural phenomena are described as the components generated from parent material, whereas anthropogenic sources primarily originate from acute human activities [17,18,25,26]. Many authors have revealed that natural sources of some PTEs (such as Pb, Cd, and Hg) have been surpassed by anthropogenic deposits into soils due to pedogenesis [17]. Industrial inputs, the combustion of fossil fuels, municipal wastewaters, and sewage sludge have been identified as anthropogenic sources of metals [11,14,27,28]. Furthermore, intensive agricultural practices have been reported to increase PTEs in soils [16,17]. In addition, automobile and vehicle emissions, road dusts, and military activities also account for increases in PTEs [4,29]. It has been estimated by some authors that agricultural practices contributed to 79.6%, 56%, and 63% of the annual concentrations of Cu, Zn, and Cd, respectively [30]. The authors further summarized that the total annual input of Pb (85%), Ni (67.5%), and Cr (43%) found in soil emanates from industrial atmospheric deposition.

The safety of plants, food, animals, and human health have been threatened by the accumulation of PTEs in soil. Toxic elements are discharged into the soil and subsequently absorbed by plants, which are consumed by animals and humans [12]. In some cases, the PTEs penetrate into surface and underground water, which are used by living organisms including humans [31–33].

In the Czech Republic, edible mushrooms grown in a smelting area were reported to have been contaminated by the atmospheric deposition of Pb [34]. In Germany, there has been an urgent call to address the Pb contents in plant-based foodstuffs including bread and potatoes, which are important suppliers of this metal in human meals [35]. The yearly deposition of Cr to soils in the UK was 327 tons [36]. In addition, the study reported that 126 tons out of the 327 tons were emanated from chemical fertilizers (mostly phosphate), while 83 tons originated from atmospheric deposition and 78 tons came from sewage sludge [36]. The effect of Cr is not only recorded in food crops but also in forest trees. The health of forest plants has been at risk because of exacerbated Cr content in the soil [37]. Globally, there have been reports on the effects of increased Cu, Fe, Mn, Mo, Zn, Ni, and other PTEs on soils, plants, animals, humans, and water. Therefore, the issue has become of critical concern to the government and the stakeholders, including decision makers.

The urgency of the situation demands a robust assessment with effective quantitative and qualitative analyses. An investigation of the PTEs in soil and their sources is the principal purpose for preserving and enhancing soil quality in most areas in the world. Thus, to develop reliable policies for a sustainable soil safety for an area, it is important to have good information on the soil and its contamination level. In recent years, several analyses including statistical, geostatistical, geo-accumulation index, multivariate and modeling, as well as potential ecological risk index analyses have been proposed and applied to investigate the source, degree, and spatiotemporal state of PTE pollution in the soil [3–6,12,14,24,28,29,31–33]. As reported by some authors, an assessment of the contents and distribution of PTEs in soil requires intensive and robust sampling to investigate the soil conditions under distinct soil types [38–42]. Furthermore, considering the high temporal, spatial, and vertical variability in the uppermost soil layers of a forest, a substantial number of samples need to be examined in order to thoroughly quantify the soil adequately along an extensive scale [43]. Routinely, the study of PTE content in soil has been performed following the regular laboratory chemical methods, including atomic absorption or inductively coupled plasma analysis. These methods are expensive and time-consuming and involve consecutive serial procedures with growing complications [44,45]. Thus, a systematically structured and affordable analytical method to monitor and assess the PTEs in soil on an appropriate vertical and spatial scale is necessary [46], especially when a tangible number of sampling points are considered. The flexibility and rapid accessibility of the positive matrix factorization (PMF) model in assessing soil pollutions is

remarkable [6,41]. This analytical method has high functionality for the investigation of the PTEs in soil. The PMF provides a great advantage in detecting and monitoring PTEs in organic and mineral soils: it is one of the best and latest models [6,41,47,48].

The study area is located in a northern part of the Czech Republic. The area is called Jizerské hory Mountains (Jizera Mts.). The area was polluted by past accumulation of PTEs from human and natural sources. However, there have been ongoing policies and efforts by the government and the people to ameliorate the problem, yet the impact is still prevailing in the ecosystem (mainly in the soil and vegetation). This is partly because, after soil is polluted, it takes a longer period for remediation and for ecosystem recovery processes to be completed [49]. Second, a large amount of the PTEs are enriched in the acidified forest soil, and these elements are still being discharged [5,50,51]. Some authors have reported health risks from the PTEs in high-altitude mountains in Europe including the Jizera Mts. [52–54]. For example, in a study performed by EscartÍn and Porte [52], the authors reported that a high percentage (76%) of polycyclic aromatic hydrocarbon (PAH) metabolites were detected in trout from the Central European high Mountains lake and that this has a high health risk. There have been many studies that focused on the acidification by sulfur and nitrogen oxides in the area. Studies focusing on the spatial and vertical distributions as well as the content and hotspots of PTEs are crucial for closing the gap in sustainable pollution assessment in the area. The benefits of applying PMF to investigate PTEs in the soil is commendable [55]. This study aims (i) to assess the concentrations of PTEs in the soil under Jizera Mts. in the Liberec region of Czech Republic after long-term, heavy industrialization; (ii) to model and map the spatial and vertical distributions of the PTEs using a GIS-based ordinary kriging technique; and (iii) to identify the possible sources of these PTEs and their contamination levels in the area using a PMF model. The findings from this study may serve as a baseline for the pollution assessment of farmland and forest soil quality in the Czech Republic and in Europe. The results might support policy-developers in sustainable farming and forestry for the health of the ecosystem and for food security, forest safety, as well as animal and human welfare.

2. Materials and Methods

2.1. Study Area

The study covered about 110 square kilometers in the Jizera Mts. The height above sea level of the area ranges from 600 to 1100 m. The average yearly temperature falls between 3 and 6 °C, which is contingent upon the altitude. The annual precipitation reaches about 1500 mm at the top of the mountains. Most areas are covered by forests (Figure 1), though in some areas, the regeneration of trees has been slow after intensive forest decline in the 1980s and 1990s [5,13]. Coniferous species, namely Norway spruce (*Picea abies*) and the European beech (*Fagus sylvatica*), are key forest trees. There are also areas with pockets of peatbogs. PTE pollution in the area is considered to have been emanated from atmospheric deposition released from the coal, textile, and steel industries and from agricultural activities. Geologically, the area is characterized by principal acidic bedrocks such as granite (granodiorite) and gneiss. Haplic/Entic Podzols, Stagnosols, and Cambisols are the predominant soils [56–58]. In most of the area, especially in the higher altitudes, the mor form of humus dominates while the moder humus type is observable only at lower altitudes [59]. The value of the soil pH was relatively low (Table 1), thus contributing to the high acidic condition of the area.
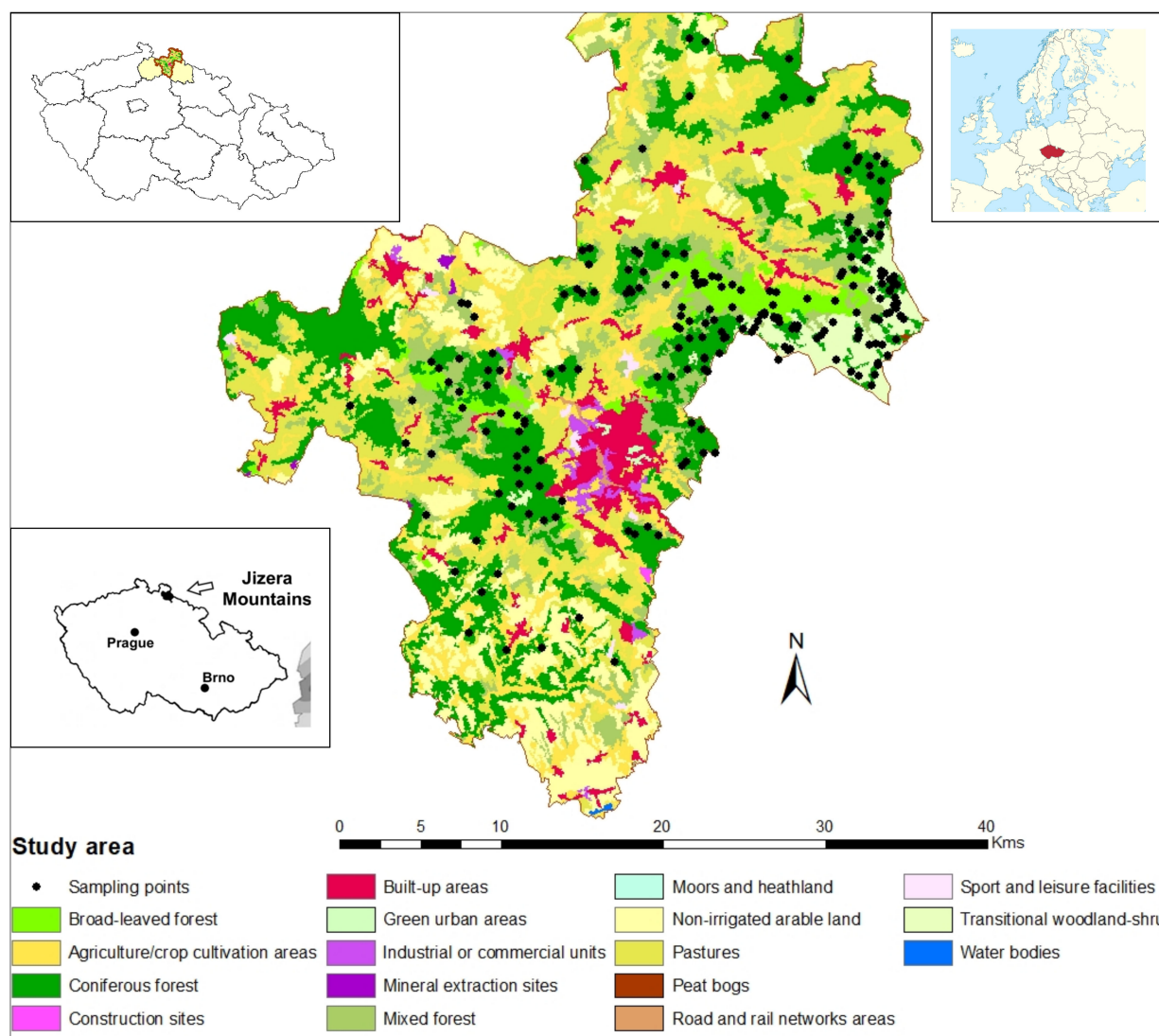
Figure 1. Sampling points and land use–land cover of the Jizera Mts. area derived from the CORINE database (**central** Map), the location of the Jizera Mts. in Liberec region in the northern part of the Czech Republic (**top left** and **down left** maps), and the location of the Czech Republic in Europe (**top right** map).

Table 1. Mean values of the physiochemical characteristics of the soil in the study area.

Properties (Unit)	Organic Soil	Mineral Soil
Sand (%)	29.7	28.2
Silt (%)	44.2	25.3
Clay (%)	26.1	46.5
Texture	Sandy clay-loam	Clay-loam
N (%)	1.6	0.5
C (%)	30.9	7.5
S (%)	0.34	0.26
P (mg kg^{-1})	946.9	386.2
K (mg kg^{-1})	811.6	935.3
Ca (mg kg^{-1})	915.2	327.9
Mg (mg kg^{-1})	839.5	1078.1
Al (mg kg^{-1})	9473.5	8614.4
pH	3.6	3.8

2.2. Soil Sampling and Laboratory Analysis

At every 3 km, soil samples were collected for both organic soil and mineral soil (to the depth of 30 cm). The samples were collected in 3 replicates for each sampling point, and the average value of the sampled points was used for the analysis. The sampling points were located using a handheld GPS system, while samples were collected using either a push probe or bucket auger depending on the terrain. A total of 221 samples each were collected from organic soil (org) and mineral soil (A) across the study area. The collected soil samples were stored in well-labelled plastic bags and taken to the laboratory. The collected samples were air-dried, ground, and sieved with a mesh of size 2.0 mm to obtain a pulverized sample.

Chemical Analysis and Instrument

The presence of elements such as Cr, Cu, Pb, Mn, and Fe in the soil were extracted using the aqua regia standard method (ISO 11466:1995, 1995) to determine their pseudo-total content [60]. For quality control (QC) of the method, the standard addition technique was adopted. For example, the QC of the concentration determination was guaranteed using the SRM 2711 (Montana II soil) reference material (National Institute of Standards and Technology, Gaithersburg, MD, USA). The values achieved were consistent with the reference data. The recovery differences were generally < 10% ($n = 3$). The detection limits for the elements based on the applied method were as follows: Cr (0.03 mg L^{-1}), Cu (0.015 mg L^{-1}), Pb (0.05 mg L^{-1}), Mn (0.05 mg L^{-1}), and Fe (0.15 mg L^{-1}). The presence of Fe and Mn in the soils were also investigated; their concentrations posed no threat in the area because their concentrations were far below the EU and world recommended limits.

2.3. Contamination Level Analysis for PTEs

The PTE pollution status of the study area was assessed through various contamination assessment indices, namely the contamination factor (CF) and the pollution index (PLI).

2.3.1. Contamination Factor

CF is defined as the ratio of metal content in the sample to the background value of the same metal. It is given by the following:

$$CF = C \ (metal) \ Sample/C \ (metal) \ background \ value \tag{1}$$

where C (metal) is the concentration of metal analyzed from sampled soil and where C (metal) background value is the geochemical background value (or concentration) of that metal.

It is important to state here that the baseline values used were the world values [10].

2.3.2. Pollution Load Index (PLI)

The PLI is an estimation and was first proposed by [60]. The pollution load index has been in use for the detection of pollution. It is robust and effective in the comparison of pollution levels in space and time. The PLI was calculated based on the concentration factor of each PTE by focusing on the background value in the soil, where CF is the contamination factor earlier stated (Equation (1)) and the letter 'n' signifies the number of metals studied. A pollution load index less than 1 indicates the optimal soil quality, and a PLI that is equal to 1 proves that only the baseline levels of contaminants are present, while a PLI greater than 1 infers the degradation of the quality of the site by [61].

The pollution load index (PLI) equation is given by the following:

$$PLI = \sqrt[n]{(CF_1 \times CF_2 \times CF_3 \times \ldots \times CF_n)} \tag{2}$$

where CF is the contamination factor derived for each metal and where n is the number of metals.

2.4. Source Apportionment Using a Positive Matrix Factorization (PMF) Model

The positive matrix factorization (PMF) model is an effective method acquired from the software EPA-PMF v 5.0, Washington DC, USA [55]. It was applied to determine the contribution of PTE sources to contamination in the study area. The mathematical method is a receptor model used in calculating the contribution of the sources to samples built on the composition or fingerprints of the sources. The PMF model apportions the collaborations of elements in soil composition by solving chemical mass balance:

$$C_{ij} = \sum_{K=1}^{p} G_{ik} + F_{kj} + E_{ij} \tag{3}$$

where C_{ij} signifies the content of PTEs j in soil sample I, p represents the number of factors (i.e., pollution sources), G_{ik} shows contribution of factor k to soil sample I, F_{kj} denotes the content of PTEs j in factor k, and E_{ij} stands for the residual.

Additional information on the procedures, methods, and formulas used in this study for determining the soil or site contamination level through the PMF model was followed as specified by [55,62] and as applied by [3,6].

2.5. Statistical Analysis and Spatial Modeling

Basic statistical parameters (such as mean, median, minimum, maximum, standard deviation, and coefficient of variance) were first calculated for each soil property based on horizon. Positive matrix factorization (PMF, EPA version 5.0, Washington, DC, USA) was used for the estimation of source apportionment and contamination level of the PTEs. To determine the relationship between the PTEs in organic and mineral soils, an ANOVA and correlation analysis were used. Ordinary kriging interpolation was used in determining the differences and/or similarities among sites with a proportional distance among them. The interpolation technique enhanced the creation of the spatial distribution maps of the PTEs of the study area. ArcGIS, version 10.7.1, CA, USA [63], was used for processing and visualizing of the spatial data. By applying the ordinary kriging technique, maps of the spatial distribution of these soil properties were generated [64]. The result was validated using the mean error [65,66]. In other words, to determine the accuracy of the produced maps, the mean error (ME) was used for the validation. The formula is shown below in Equation (4):

$$ME = \sum_{i=1}^{n} (x_{1.i} - x_{2.j})/n \tag{4}$$

where x_1 is prediction of the variable x, x_2 is measure of that variable, and n is number of records.

3. Results and Discussion

3.1. General Description of PTEs Concentrations and Their Spatial Distribution in the Soil

The basic statistical characteristics of the studied PTEs including Cr, Cu, Fe, Mn, and Pb for the organic soil and mineral soil have been described in Table 2. The coefficient of variation (CV) defines the degree of variations within PTE concentrations [67]. A coefficient of variation value less than 20% represents low variability, and a CV that falls between 21–50% is regarded as moderate variability. On the other hand, a CV ranging from 50–100% signifies high variability, while a CV greater than 1 (that is >100%) is described as having extremely high heterogeneity. In this study, the CV of the PTEs in the organic soil increased in the following order: Fe < Pb < Cu < Cr < Mn, accounting for 46.31%, 49.40%, 54.94%, 96.36%, and 97.06%, respectively.

Table 2. Basic statistical characteristics of the soil PTE concentrations in the study area.

Soil Horizons	Parameters [†]	Cr	Cu	Fe [‡]	Mn [‡]	Pb
Organic soil	Count	221	221	221	221	221
	Mean	11.0	16.2	7357.8	149.6	99.2
	Median	9.1	15.5	7010	73	92.9
	Mode	7.2	18.5	10,200	32	104
	Minimum	3.1	2.3	1004	1.0	7.1
	Maximum	85.2	81.9	21,000	1650	339
	Std dev	10.6	8.9	3407.5	145.2	49
	Coef of Var. (CV)	96.36	54.94	46.31	97.06	49.4
Mineral soil	Count	221	221	221	221	221
	Mean	4.5	6.4	6744.3	168.0	65.6
	Median	3.8	3.8	6194.4	68.4	58.8
	Mode	3.9	1.0	3610	248	111
	Minimum	0.4	0.2	159.3	0.5	6.7
	Maximum	26.5	38.3	24,274.0	1940.0	281.0
	Std dev	3.0	6.2	4054.5	137.6	42.3
	Coef of Var. (CV) *	66.7	96.88	60.12	81.9	64.48
	Czech Republic	<11.0	<16.0	>8000	<150.0	<60.0
	** European mean value	94.8	17.3	38,000	524	32
	** World mean value	59.5	38.9	35,000	488	27
	** Crati Basin	90.54	44.36	54,700	1300	63.67

* Authors' estimates from most publications in the Czech Republic on the issue as there was no official existing baseline; ** Kabata-Pendias [10]. [†] All parameters and numbers are reported in mg kg^{-1}, while CV is reported in %. [‡] Fe and Mn also showed reasonable variability; they posed no threat because their concentrations were far below the EU and world limits.

The CV of the PTEs for mineral soil was also in ascending order: Fe (60.12%) < Pb (64.48%) < Cr (66.70%) < Mn (81.9%) < Cu (96.88%). The results derived from the CV revealed a high variability between the PTEs in the mineral soil. Similarly, in the organic soil, the CV for Cr, Mn, and Cu indicated a high variability (Table 2). In general, both the organic soil and the mineral soil revealed high heterogeneity (or variability). All of the PTEs showed relatively high variability in both soil horizons except for the Fe (46.31%) in the organic soil. The spatial distribution of the heterogeneity of the PTEs suggest that the metals are enriched by intensive sources of from the industrial, commercial, domestic and agricultural sectors [3,5,13]. However, the content of the PTEs varies between the soil horizons, yet the organic soil had higher mean values across the metals, excluding Mn. The content of manganese was 18.4 mg kg^{-1} higher in the mineral soil compared with its content in the organic soil. This finding agrees with a report by other authors in the same region [21]. Studies have shown that, in addition to human activities and their associated soil acidifications [13], the geological bedrock of the area also contributes to accruing PTE concentrations [15]. The study area has Podzols and Dystric Cambisols as the prevailing soils [5], and this might have contributed to the high contents of Pb, Fe, and Mn. The concentration of Fe in the study area is remarkable when compared with other metals. This could be attributed to the high acidic soil status of the area (Table 1). As has been earlier reported, the concentration of Fe in the soil solution at optimal soil pH falls between 30 and 550 µg L^{-1}, but in high acid soils, it may exceed 2000 µg L^{-1} [68]. Higher concentrations of the PTEs were found in this study area relative to the neighboring regions in the country [3]. In comparison with the European value [10], the world value [10], and the Crati Basin value [69], the contents of Pb in both the organic soil and the mineral soil were higher. The exceptional content of Pb in the study area might be attributed to past intensive anthropogenic activities and the prevailing geological formation of Cambisols. Lead has been reported to exhibit the highest content in a Cambisols soil group [10].

3.2. Relationships among the PTE Concentrations in the Organic and Mineral Soils

The correlation analysis for the PTE concentrations in the organic and mineral soils revealed that the content of Cr in the mineral soil (Cr_tot_A) showed a significant and strong positive correlation with Cu, Fe, Mn, and Pb in the mineral soil (Table 3). Chromium has a strong relationship with other elements because it is easily mobilized in acidic soils and our study area is highly acidic [5,10]. Furthermore, in the mineral soil, Pb is significantly correlated with Cu and Mn. In the organic soil, Pb has a significant and strong positive relation with Cu and Fe. The strong relation between Pb, Fe, and Mn was documented earlier [10]. The concentrations of Pb in Fe–Mn nodules can be as high as 20,000 mg kg^{-1} [70]. Most of the negative correlations between the PTEs occurred in the inter-horizon and not within the same horizon. The correlation between the elements in the same soil horizon showed more positive relationships. This could be described by the likelihood that they shared the same origin. Furthermore, the correlation of the PTEs in the mineral soil revealed stronger relationships when compared with the correlation in the organic soil. This probably proved that these PTEs are more closely associated in the mineral soil relative to the organic soil. This finding was consistent with a recent report by other authors on the same issue [3].

Table 3. Summary of correlation analyses between the PTE concentrations in the organic and mineral soils.

Parameters	Cr_tot_org	Cu_tot_org	Fe_tot_org	Mn_tot_org	Pb_tot_org	Cr_tot_A	Cu_tot_A	Fe_tot_A	Mn_tot_A	Pb_tot_A
Cr_tot_org	1.00									
Cu_tot_org	0.03	1.00								
Fe_tot_org	0.77 *	0.43 *	1.00							
Mn_tot_org	0.52 **	−0.40	0.19	1.00						
Pb_tot_org	0.00	0.71 **	0.64 *	−0.56 *	1.00					
Cr_tot_A	0.53 *	−0.14	0.10	0.58	−0.21	1.00				
Cu_tot_A	0.10	−0.13	−0.10	0.20	−0.26 *	0.78 *	1.00			
Fe_tot_A	0.03	0.16	0.50 *	−0.06	0.43	0.56 **	0.00	1.00		
Mn_tot_A	0.10	−0.22 *	−0.05	0.73 *	−0.59 **	0.60 *	0.58 *	0.03	1.00	
Pb_tot_A	0.05	−0.12	−0.08	0.10	−0.10	0.76 **	0.81 **	0.19	0.54 *	1.00

* = Correlation is significant at the 0.01 *p*-value, at <0.05; ** = correlation is significant at the 0.05 *p*-value; tot_org = total concentration in organic soil; tot_A = total concentration in the mineral soil.

The ANOVA in Table 4 was used to analyze the distribution of PTE contents in relation to the organic soil and the mineral soil horizons. It was revealed that all of the elements, with the exception Mn, showed significant relationships in both the organic soil and the mineral soil. There have been few studies within and outside the study area that focused on the relationship between soil horizon [8,21]. Consistent with our study, many authors have reported a significant relationship between soil horizon, elevation, and the concentrations of metals. For example, in the Suxian district of Chenzhou City in Hunan Province of China, it was revealed that heavy metal concentrations decreased at low elevation but increased considerably with increasing elevations [8]. Other studies have affirmed that fine-particle metals including Cr and Cu accumulate more at lower elevations [11].

Table 4. Summary of ANOVA for PTE concentrations for the tot-org and tot-A horizons.

Soil Parameters	df	F-Statistics	*p*-Value *
Cr_tot_org	220	2.12	0.019
Cu_tot_org	220	−3.73	<0.001
Fe_tot_org	220	−1.40	<0.001
Mn_tot_org	220	3.31	0.685
Pb_tot_org	220	0.63	<0.001
Cr_tot_A	220	2.06	0.016
Cu_tot_A	220	−1.94	0.021
Fe_tot_A	220	−1.13	<0.001
Mn_tot_A	220	4.91	0.283
Pb_tot_A	220	0.82	0.041

* Figures (or values) in bold are significant at the 0.05 confidence level. tot_org = total concentration in organic soil; tot_A = total concentration in the mineral soil.

In the organic soil, the highest contents of Cr, Pb, Fe, and Mn were found in the northern and central parts of the area (Figure 2). Studies have revealed a close association between Mn and Fe. Manganese is described as a member of the iron family, and both elements are closely linked in geochemical processes [10]. The author further stressed that Mn cycles follow Fe cycles in various terrestrial environments. Copper on the other hand had a concentration hotspot that spread from the northeast to the northwest.
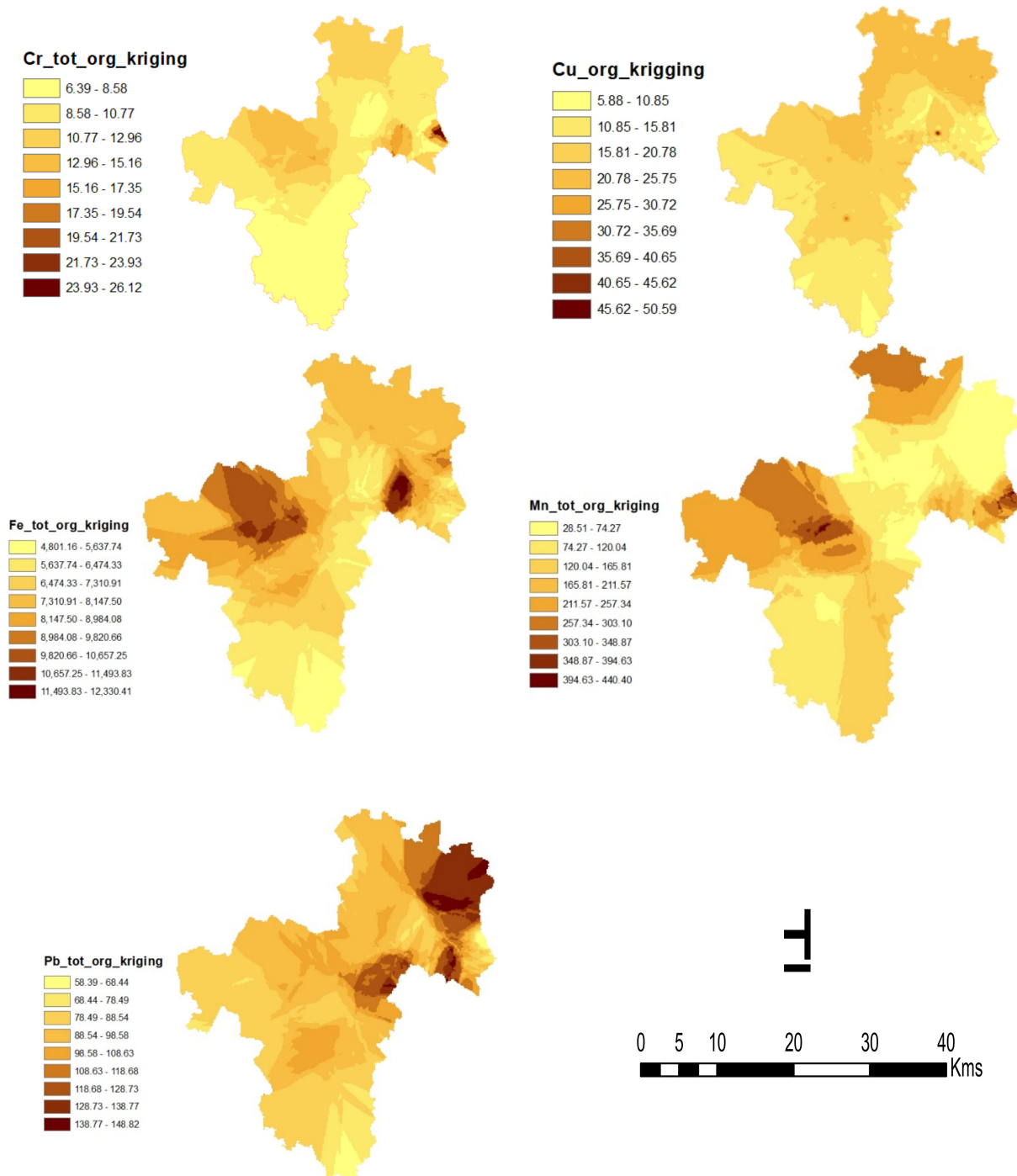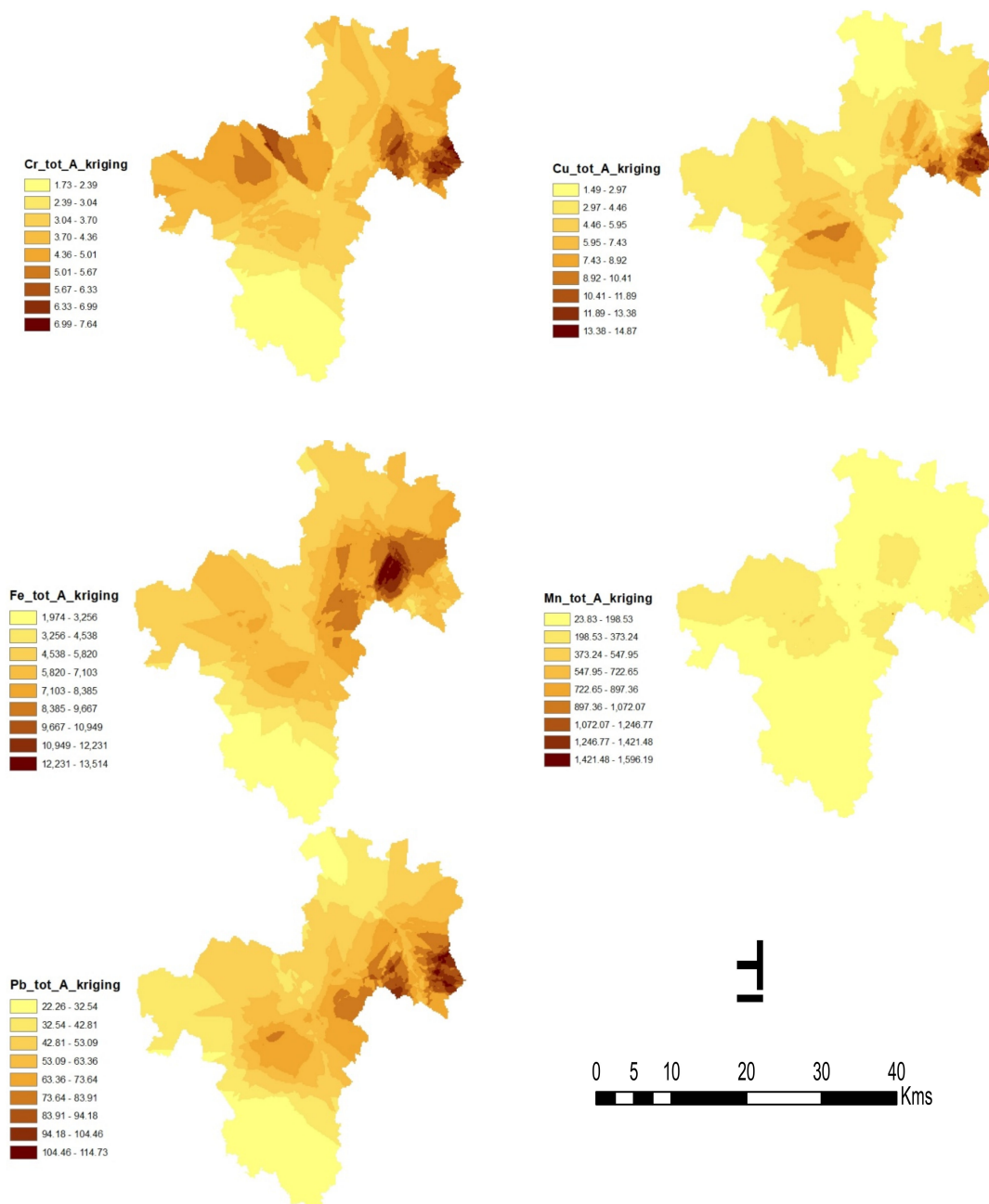

Figure 2. Spatial and vertical distributions of soil characteristics in the organic soil (org) assessed using ordinary kriging (all of the elements are reported in mg kg^{-1}). Though Fe and Mn posed no threat in the region, they were mapped/modeled to draw inferences on Cr, Cu, and Pb. In other words, the study attempted to assess if the presence of Fe and Mn in the soil influenced the vertical and spatial distributions of the three other PTEs (namely Cr, Cu, and Pb) in the different soil horizons.

The kriged map of Cu and Pb distributions showed almost the same pattern in the mineral soil (Figure 3). They tend to have higher concentrations towards the east in this area. In this mineral soil, Fe and Cr showed extensive spatial distribution patterns that spread from the northeast, through the central region, and to the northwestern part of the mapped area. On the other hand, the kriged map of Mn distribution showed higher concentrations within the northwest and north-central parts of the area. In general, for both organic and mineral soils, the northern and central parts of the kriged maps revealed more distribution of the elements when compared with the southern part. This could be explained by the historical distribution pattern of industrial and agricultural activities in the study area, which were mainly located in the northern and central parts [71].

Figure 3. Spatial and vertical distributions of soil characteristics in the surface mineral soil (A) made by ordinary kriging. (all of the elements are in mg kg^{-1}). Though Fe and Mn posed no threat in the region, they were mapped/modeled to draw inferences on Cr, Cu, and Pb. In other words, the study attempted to assess if the presence of Fe and Mn in the soil influenced the vertical and spatial distribution of the other three PTEs (namely Cr, Cu, and Pb) in the different soil horizons.

3.3. Source Apportionment by the Positive Matrix Factorization (PMF) Model

There have been many reports and studies that affirmed that our study area is located within the vicinity of various anthropogenic activities including mining, intensive agriculture, automobile gas emissions, and acute biological sludge, which might affect the soil [4,52,53,72,73]. It is important to examine the sources of PTEs in the study area. Therefore, the PMF model was adopted as one of the best and latest models with high functionalities for effective PTE source identification [6]. The validity and reliability of the analysis are based on minimum Q to model the residual matrix that influences a substantial number of variables. To derive the best result, the PMF model was run 20 times, while the best outputs (which were Run 8 and Run 20) were selected following the software developer's guide [55]. The PMF analysis produced six factors (see Figures 4 and 5) and disclosed the origin of contributions based on each PTE [62].
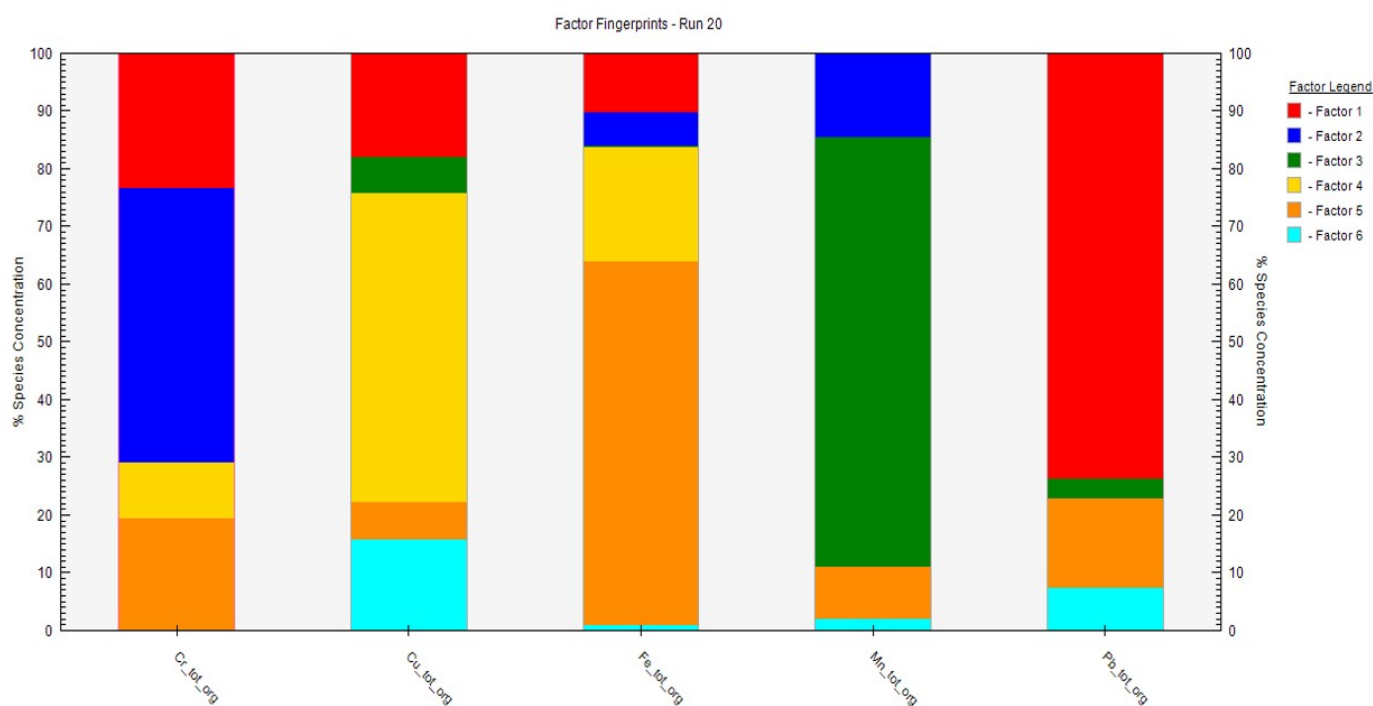
Figure 4. Source fingerprint of the total organic soil (tot-org) from the PMF model analysis showing the percentage contribution of PTEs. Note that the source and availability of Fe and Mn posed no risk to the soils.

In the organic soil, factor 1 was dominated by Pb and Cr, with a factor loading of 73.5% and 23.7%, while in the mineral soil, Cu and Pb had factor loads of 92.2% and 18.3%, respectively (Figures 4 and 5). Factor 2 was characterized by Cr and Mn, and these accrued factor loadings of 45.1% and 15.1%, respectively, in the organic soil layer. On the other hand, in the mineral soil, Mn and Cu accumulated factor loadings of 28.3% and 6.7%, respectively. The factor 3 load consisted mainly of Mn (72.6%) and Cu (4.9%) in the organic soil, while Fe (64.8%) and Cr (11.9%) accrued in the mineral soil. Furthermore, factor 4 load was characterized by Cu (51.7%) and Fe (18.3%) in the organic soil while Mn (71.1) and Pb (26.4%) accumulated in the mineral soil. The factor 5 had Fe (42.5%), Cr (19.6%), and Pb (14.2%), accounting for the highest elements in the organic soil, while Cr (72.9%) and Fe (20.1%) accrued in the mineral soil. The 6th factor (which is the last factor) revealed that Cu (15.7%) had the highest factor load in the organic and mineral soils relative to all of the studied elements. Cr, Cu, and Pb accumulations in the study probably confirmed intensive pollution from many sources such as industrial, agricultural, commercial, and municipal activities and wastes [28]. These PTEs might have also been deposited during weathering because, in mineral forms, most of the elements are oxidized, released, and reprecipitated in the soil [74]. However, the sources of Mn and Fe in the soil are natural sources, posing

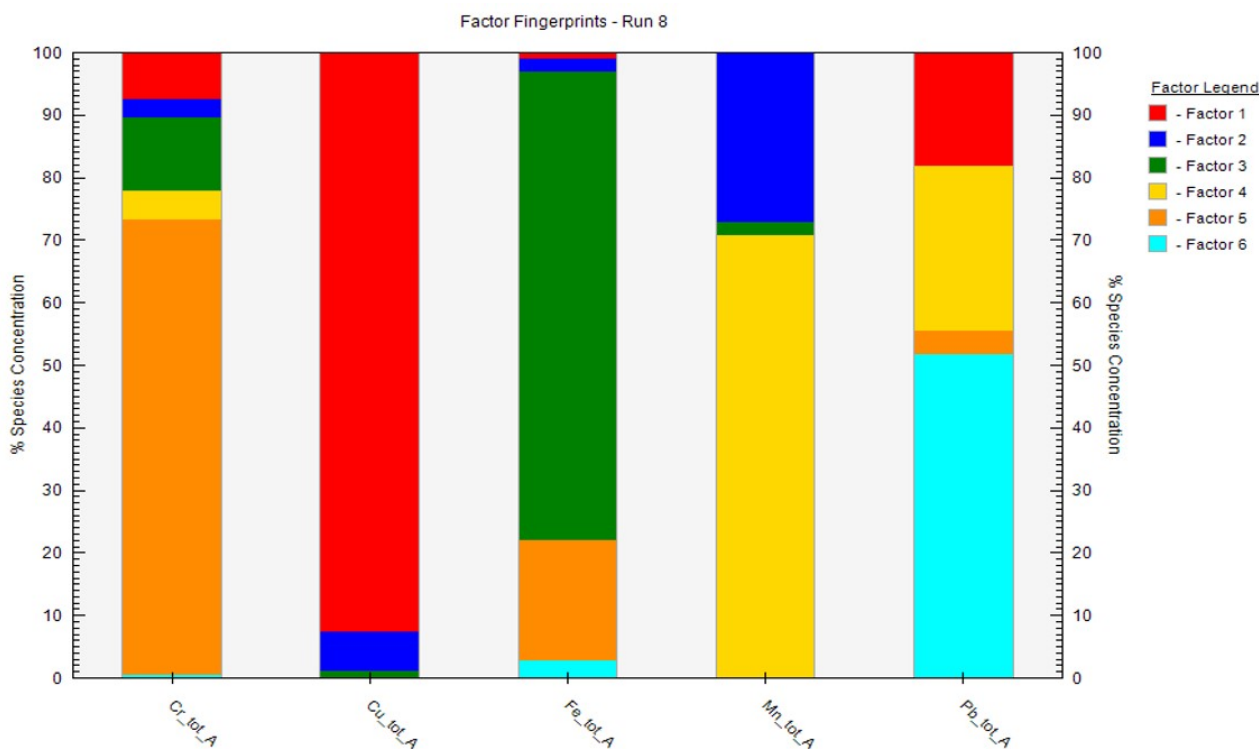no threats if the concentrations do not exceed the maximum allowable limits. However, the sources of Mn and Fe in the soil are natural sources, posing no threats if the concentrations do not exceed the maximum allowable limits.

Figure 5. Source fingerprint of the total mineral soil (tot-A) from the PMF model analysis showing the percentage contribution of PTEs. Note that the source and availability of Fe and Mn posed no risk to the soils.

3.4. Contamination Factor and Pollution Load Index of PTEs for Organic Soil and Mineral Soil

The research revealed that, in organic soil, there was no obvious pollution recorded in any of the soil sample points except in samples 33, 36, 65, 81, 103, 113, 147, 157, and 193. On the other hand, in the mineral horizon, the soil at some sites were polluted, observed only in samples 111, 126, 213, and 221 horizon (Table S1 and Table S2). This finding, however, is in contrast with the reports from previous studies in the same area. For example, there are studies showing that, in the past few decades, the area was among the major pollution zones caused by industrial and agricultural activities [5,75]. Furthermore, as one of the regions located in northern Bohemia, the study area has been documented as a region characterized by power and coal production from the 1950s to the 1980s [75]. In addition to intensive agriculture, industrial activities, and geological processes, the study area has some peatbogs, and these increased the pollution of the area by PTEs to a large extent [76,77]. This study area lies in the "Black Triangle", which is affected by industrial activities linked to the extraction and exploitation of coal and other natural resources on the sides of the Jizera Mts areas [20,21,72,73]. The study area is historically susceptible to pollution, but the current results revealed otherwise. This could be due to measures put in place by authorities to ameliorate the impact of industrial activities in the area.

4. Conclusions

In both organic and the mineral soils, a high variability in the PTEs was observed. The spatial distribution and the heterogeneity of the PTEs suggested that there were widely distributed sources of the metals' enrichment from the industrial, commercial, domestic, and agricultural sectors. Lead revealed a high concentration level. Chromium showed a strong relationship to other elements investigated, while Pb has a significant and strong

positive relation with Cu. This was probably because Cr is easily mobilized in acidic soils and our study area is highly acidic.

The correlation between the elements in the same soil layer showed more positive relationships, while the correlation of the PTEs in the mineral soil revealed stronger relationships when compared with the correlation in the organic soil. The findings revealed that all the elements in exemption of Mn indicated significant relationships in both the organic soil and the mineral soil. Meanwhile, the concentrations of Mn and Fe were not harmful in the study area. On the other hand, Mn and Fe we below the EU and World limits. In both the organic soil and the mineral soil, the northern part of the kriged maps revealed more distribution of the elements when compared with either the southern part. This implied that the concentrations of the elements were oriented towards the northern part of the study region.

The applications of the Positive Matrix Factorization (PMF) model, ArcGIS-based ordinary kriging, and contamination level analysis were effective for source identification, hotspot location, and assessment of the contamination level of the PTEs. In the organic soil, there was no obvious pollution recorded in any of the soil sample points except in samples 33, 36, 65, 81, 103, 113, 147, 157, and 193. On the other hand, in the mineral horizon, some deteriorated site quality was observed only in samples 111, 126, 213, and 221. The method and results presented might be applicable in coniferous and broad-leaf tree-dominated highlands with a history of industrial activities and atmospheric acidifications. The methods are suitable for measuring the distribution and concentration of elements.

The current result revealed that there is no evidence of pollution by PTEs in the Jizera Mts area. In contrast, this area lies within the "Black Triangle", which was affected by industrial activities linked with the extraction and exploitation of coal and other natural resources in Central Europe. However, the rate of pollution in the area is very low based on the findings of this study. There may be a need for intermittent assessment of the soil. This regular assessment will help to curtail the possibility of any excessive accumulation and escalation in the future. The findings from this study may serve as a baseline for pollution assessments of farmland and forest soil quality in the Czech Republic and in Europe. The results might support policy-developers in sustainable farming and forestry for the health of the ecosystem to achieve food security, forest safety, as well as animal and human welfare.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alloway, B.J. *Micronutrient Deficiencies in Global Food Production*; Springer: Heidelberg, Germany, 2008.
2. Eriksson, J.; Dahlin, S.A.; Sohlenius, G.; Söderström, M.; Öborn, I. Spatial patterns of essential trace element concentrations in Swedish soils and crops. *Geoderma Reg.* **2017**, *10*, 163–174. [CrossRef]
3. Agyeman, P.C.; Ahado, S.K.; Kingsley, J.; Kebonye, N.M.; Biney, J.K.M.; Borůvka, L.; Vasat, R.; Kocarek, M. Source apportionment, contamination levels, and spatial prediction of potentially toxic elements in selected soils of the Czech Republic. *Environ. Geochem. Health* **2020**, *43*, 601–620. [CrossRef] [PubMed]
4. Barker, B.J.; Clausen, J.L.; Douglas, T.L.; Bednar, A.J.; Griggs, C.S.; Martin, W.A. Environmental impact of metals resulting from military training activities: A review. *Chemosphere* **2021**, *265*, 129110. [CrossRef]
5. Borůvka, L.; Vašát, R.; Němeček, K.; Novotný, R.; Šrámek, V.; Vacek, O.; Pavlů, L.; Fadrhonsová, V.; Drábek, O. Application of regression-kriging and sequential Gaussian simulation for the delineation of forest areas potentially suitable for liming in the Jizera Mountains region, Czech Republic. *Geoderma Reg.* **2020**, *21*, e00286. [CrossRef]
6. Chen, L.; Lowenthal, D.H.; Watson, J.G.; Koracin, D.; Kumar, N.; Knipping, E.M.; Wheeler, N.; Craig, K.; Reid, S. Toward effective source apportionment using positive matrix factorization: Experiments with simulated PM2.5 data. *J. Air Waste Manag. Assoc.* **2010**, *60*, 43–54. [CrossRef]
7. Chien, C.S.; Wang, H.; Chen, Y.; Wang, M.; Liu, C. Removal of heavy metals from contaminated paddy soils using chemical reductants coupled with dissolved organic carbon solutions. *J. Hazard. Mater.* **2021**, *403*, 123549. [CrossRef] [PubMed]
8. Ding, Q.; Cheng, G.; Wang, Y.; Zhuang, D. Effects of natural factors on the spatial distribution of heavy metals in soils surrounding mining regions. *Sci. Total Environ.* **2017**, *578*, 577–585. [CrossRef]
9. Gerdol, R.; Bragazza, L. Effects of altitude on element accumulation in alpine moss. *Chemosphere* **2020**, *64*, 810–816. [CrossRef] [PubMed]
10. Kabata-Pendias, A. *Trace Elements in Soils and Plants*, 4th ed.; CRC Press Taylor & Francis Group: New York, NY, USA, 2011.
11. Liu, X.; Shi, H.; Bai, Z.; Zhou, W.; Liu, K.; Wang, M.; He, Y. Heavy metal concentrations of soils near the large opencast coal mine pits in China. *Chemosphere* **2020**, *244*, 125360. [CrossRef]
12. Marcantonio, R.A.; Field, S.P.; Sesay, P.B.; Lamberti, G.A. Identifying human health risks from precious metal mining in Sierra Leone. *Reg. Environ. Chang.* **2021**, *21*, 1–12. [CrossRef]
13. Pavlů, L.; Drábek, O.; Stejskalová, Š.; Tejnecký, V.; Hradilová, M.; Nikodem, A.; Borůvka, L. Distribution of aluminium fractions in acid forest soils: Influence of vegetation changes. *iForest Biogeosci. For.* **2018**, *11*, 721–727. [CrossRef]
14. Peli, M.; Bostick, B.C.; Barontini, S.; Lucchini, R.G.; Ranzi, R. Profiles and species of Mn, Fe and trace metals in soils near a ferromanganese plant in Bagnolo Mella (Brescia, IT). *Sci. Total Environ.* **2021**, *755*, 143123. [CrossRef]
15. Salvador-Blanes, S.; Cornu, S.; Bourennane, H.; King, D. Controls of the spatial variability of Cr concentration in topsoils of a central French landscape. *Geoderma* **2006**, *132*, 143–157. [CrossRef]
16. Sungur, A.; Kavdir, Y.; Özcan, H.; İlay, R.; Soylak, M. Geochemical fractions of trace metals in surface and core sections of aggregates in agricultural soils. *Catena* **2021**, *197*, 104995. [CrossRef]
17. Kalkhajeh, Y.K.; Huang, B.; Hu, W.; Ma, C.; Gao, H.; Thompson, M.L.; Hansen, H.C.B. Environmental soil quality and vegetable safety under current greenhouse vegetable production management in China. *Agric. Ecosyst. Environ.* **2021**, *307*, 107230. [CrossRef]
18. Tian, H.; Zhang, C.; Qi, S.; Kong, X.; Yue, X. Concentration and Spatial Distribution of Potentially Toxic Elements in Surface Soil of a Peak-Cluster Depression, Babao Town, Yunnan Province, China. *Int. J. Environ. Res. Public Health* **2021**, *18*, 3122. [CrossRef]
19. Shar, S.; Reith, F.; Ball, A.S.; Shahsavar, E. Long-term Impact of Gold and Platinum on Microbial Diversity in Australian Soils. *Microb. Ecol.* **2021**, *81*, 977–989. [CrossRef]
20. Matys Grygar, T.; Nováková, T.; Bábek, O.; Elznicová, J.; Vadinová, N. Robust assessment of moderate heavy metal contamination levels in floodplain sediments: A case study on the Jizera River, Czech Republic. *Sci. Total Environ.* **2013**, *452–453*, 233–245. [CrossRef]
21. Kváčová, M.; Ash, C.; Borůvka, L.; Pavlů, L.; Nikodem, A.; Němeček, K.; Tejnecký, V.; Drábek, O. Contents of Potentially Toxic Elements in Forest Soils of the Jizera Mountains Region. *Environ. Model. Assess.* **2015**, *20*, 183–195. [CrossRef]
22. Demková, L.; Jezný, T.; Bobuľská, L. Assessment of soil heavy metal pollution in a former mining area before and after the end of mining activities. *Soil Water Res.* **2017**, *12*, 229–236. [CrossRef]
23. Kabala, C.; Galka, B.; Jezierski, P. Assessment and monitoring of soil and plant contamination with trace elements around Europe's largest copper ore tailings impoundment. *Sci. Total Environ.* **2020**, *738*, 139918. [CrossRef]
24. Juhos, K.; Czigány, S.; Madarász, B.; Ladányi, M. Interpretation of soil quality indicators for land suitability assessment using multivariate approach for Central European arable soils. *Ecol. Indic.* **2019**, *99*, 261–272. [CrossRef]
25. Ahmadi, M.; Akhbarizadeh, R.; Haghighifard, N.J.; Barzegar, G.; Jorfi, S. Geochemical determination and pollution assessment of heavy metals in agricultural soils of south western of Iran. *J. Environ. Health Sci. Eng.* **2019**, *17*, 657–669. [CrossRef]
26. Hoaghia, M.A.; Levei, E.A.; Cadar, O.; Senila, M.; Hognogi, G.G. Assessment of metal contamination and ecological risk in urban soils situated near a metallurgical complex. *Environ. Eng. Manag. J.* **2017**, *16*, 1623–1630. [CrossRef]
27. Alloway, B. *Heavy Metals in Soils*; Chapman and Hall: London, UK, 1995.

28. Nwaogu, C.; Ogbuagu, D.H.; Abrakasa, S.; Olawoyin, M.A.; Pavlů, V. Assessment of the impacts of municipal solid waste dumps on soils and plants. *Chem. Ecol.* **2017**, *33*, 589–606. [CrossRef]

29. Mondal, S.; Singh, G. Pollution evaluation, human health effect and tracing source of trace elements on road dust of Dhanbad, a highly polluted industrial coal belt of India. *Environ. Geochem. Health* **2021**, *43*, 2081–2103. [CrossRef] [PubMed]

30. Luo, X.S.; Yu, S.; Zhu, Y.G.; Li, X.D. Trace metal contamination in urban soils of China. *Sci. Total Environ.* **2012**, *421–422*, 17–30. [CrossRef] [PubMed]

31. Bhuiyan, M.A.H.; Bodrud-Doza, M.; Rakib, M.A.; Saha, B.B.; Didar-Ul Islam, S.D. Appraisal of pollution scenario, sources and public health risk of harmful metals in mine water of Barapukuria coal mine industry in Bangladesh. *Environ. Sci. Pollut. Res.* **2021**, *28*, 22105–22122. [CrossRef]

32. Simionov, I.A.; Cristea, D.S.; Petrea, S.M.; Mogodan, A.; Nicoara, M.; Plavan, G.; Baltag, E.S.; Jijie, R.; Strungaru, S.A. Preliminary investigation of lower Danube pollution caused by potentially toxic metals. *Chemosphere* **2021**, *264*, 128496. [CrossRef]

33. Tamilmani, A.; Venkatesan, G. Assessment of trace metals and its pollution load indicators in water and sediments between Upper and Grand Anicuts in the Cauvery. *Int. J. Environ. Sci. Technol.* **2021**. [CrossRef]

34. Komárek, M.; Ettler, V.; Chrastný, V.; Mihaljevic, M. Lead isotopes in environmental sciences: A review. *Environ. Int.* **2008**, *34*, 562–577. [CrossRef]

35. Brüggemann, J.; Dörfner, H.H.; Hecht, H.; Kumpulainen, J.T.; Westermair, T. Status of Trace Elements in Staple Foods from Germany 1990–1994. In *Trace Elements, Natural Antioxidants and Contaminants in European Foods and Diets*; Kumpulainen, J.T., Ed.; FAO: Rome, Italy, 1996; Volume 49, p. 5.

36. Nicholson, F.A.; Smith, S.R.; Alloway, B.J.; Carlton-Smith, C.; Chambers, B.J. An inventory of heavy metals inputs to agricultural soils in England and Wales. *Sci. Total Environ.* **2003**, *311*, 205–219. [CrossRef]

37. Staniszewski, P.; Bilek, M.; Szwerc, W.; Tomusiak, R.; Osiak, P.; Kocjan, R.; Moskalik, T. The effect of tree age, daily sap volume and date of sap collection on the content of minerals and heavy metals in silver birch (Betula pendula Roth) tree sap. *PLoS ONE* **2020**, *15*, e0244435. [CrossRef]

38. Ren, H.Y.; Zhuang, D.F.; Singh, A.; Pan, J.J.; Qiu, D.S.; Shi, R.H. Estimation of as and Cu contamination in agricultural soils around a mining area by reflectance spectroscopy: A case study. *Pedosphere* **2009**, *19*, 719–726. [CrossRef]

39. Gholizadeh, A.; Boruvka, L.; Saberioon, M.; Vasat, 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. *Appl. Spectrosc.* **2013**, *67*, 1349–1362. [CrossRef]

40. Gholizadeh, A.; Boruvka, L.; Vasat, R.; Saberioon, M.; Klement, A.; Kratina, J.; Tejnecky, V.; Drabek, O. Estimation of potentially toxic elements contamination in anthropogenic soils on a Brown coal mining dumpsite by reflectance spectroscopy: A case study. *PLoS ONE* **2015**, *10*, e0117457. [CrossRef]

41. Gholizadeh, A.; Saberioon, M.; Ben-Dor, E.; Rossel, R.A.V.; Boruvka, L. Modelling potentially toxic elements in forest soils with viseNIR spectra and learning algorithms. *Environ. Pollut.* **2020**, *267*, 115574. [CrossRef]

42. Sun, X.; Zhang, L.; Lv, J. Spatial assessment models to evaluate human health risk associated to soil potentially toxic elements. *Environ. Pollut.* **2021**, *268*, 115699. [CrossRef] [PubMed]

43. Chodak, M.; Khanna, P.; Horvath, B.; Beese, F. Near infrared spectroscopy for determination of total and exchangeable cations in geologically heterogeneous forest soils. *J. Near Infrared Spectrosc.* **2004**, *12*, 315–324. [CrossRef]

44. Hang, X.; Wang, H.; Zhou, J.; Ma, C.; Du, C.; Chen, X. Risk assessment of potentially toxic element pollution in soils and rice (Oryza sativa) in a typical area of the Yangtze River Delta. *Environ. Pollut.* **2009**, *157*, 2542–2549. [CrossRef]

45. Liu, M.; Wang, T.; Skidmore, A.K.; Liu, X. Heavy metal-induced stress in rice crops detected using multi-temporal Sentinel-2 satellite images. *Sci. Total Environ.* **2018**, *637–638*, 18–29. [CrossRef] [PubMed]

46. Arellano, P.; Tansey, K.; Balzter, H.; Boyd, D.S. Detecting the effects of hydrocarbon pollution in the Amazon forest using hyperspectral satellite images. *Environ. Pollut.* **2015**, *205*, 225–239. [CrossRef] [PubMed]

47. Kong, F.; Chen, Y.; Huang, L.; Yang, Z.; Zhu, K. Human health risk visualization of potentially toxic elements in farmland soil: A combined method of source and probability. *Ecotoxicol. Environ. Saf.* **2021**, *211*, 111922. [CrossRef]

48. Wang, Z.; Chen, X.; Yu, D.; Zhang, L.; Wang, J.; Lv, J. Source apportionment and spatial distribution of potentially toxic elements in soils: A new exploration on receptor and geostatistical models. *Sci. Total Environ.* **2021**, *759*, 143428. [CrossRef]

49. Skeffington, R.A.; Cosby, B.J.; Whitehead, P.G. Long-term predictions of ecosystem acidification and recovery. *Sci. Total Environ.* **2016**, *568*, 381–390. [CrossRef]

50. Akselsson, C.; Hultberg, H.; Karlsson, P.E.; Pihl Karlsson, G.; Hellsten, S. Acidification trends in south Swedish forest soils 1986–2008-Slow recovery and high sensitivity to sea-salt episodes. *Sci. Total Environ.* **2013**, *444*, 271–287. [CrossRef] [PubMed]

51. Marx, A.; Hintze, S.; Sanda, M.; Jankovec, J.; Oulehle, F.; Dusek, J.; Vitvar, T.; Vogel, T.; van Geldern, R.; Barth, J.A.C. Acid rain footprint three decades after peak deposition: Long-term recovery from pollutant sulphate in the Uhlirska catchment (Czech Republic). *Sci. Total Environ.* **2017**, *598*, 1037–1049. [CrossRef]

52. EscartÍn, E.; Porte, C. Biomonitoring of PAH Pollution in High-Altitude Mountain Lakes through the Analysis of Fish Bile. *Environ. Sci. Technol.* **1999**, *33*, 406–409. [CrossRef]

53. Kopáček, J.; Hejzlar, J.; Krám, P.; Oulehle, F.; Posch, M. Effect of industrial dust on precipitation chemistry in the Czech Republic (Central Europe) from 1850 to 2013. *Water Res.* **2016**, *103*, 30–37. [CrossRef] [PubMed]

54. Křeček, J.; Palán, L.; Stuchlík, E. Impacts of land use policy on the recovery of mountain catchments from acidification. *Land Use Policy* **2019**, *80*, 439–448. [CrossRef]

55. U.S. EPA. *EPA Positive Matrix Factorization (PMF) 3.0 Model*; U.S. Environmental Protection Agency: Research Triangle Park, NC, USA, 2010. Available online: http://www.epa.gov/heasd/products/pmf/pmf.html (accessed on 20 March 2021).

56. IUSS Working Group WRB. World Reference Base for Soil Resources 2014, Update 2015. In *International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*; World Soil Resources Reports No. 106; FAO: Rome, Italy, 2015. Available online: http://www.fao.org/publications/card/en/c/942e424c-85a9-411d-a739-22d5f8b6cc41 (accessed on 2 December 2019).

57. Vacek, O.; Vašát, R.; Borůvka, L. Quantifying the pedodiversity-elevation relations. *Geoderma* **2020**, *373*, 114441. [CrossRef]

58. Tessier, A.; Campbell, P.G.C.; Bisson, M. Sequential extraction procedure for the speciation of particulate trace metals. *Anal. Chem.* **1979**, *51*, 844–851. [CrossRef]

59. Pavlů, L.; Borůvka, L.; Drábek, O.; Nikodem, A. Effect of natural and anthropogenic acidification on aluminium distribution in forest soils of two regions in the Czech Republic. *J. For. Res.* **2021**, *32*, 363–370. [CrossRef]

60. Melo, B.A.G.; Motta, F.L.; Santana, M.H.A. Humanic acids: Structural properties and multiple functionalities for novel technical development. *Mater. Sci. Eng.* **2016**, *62*, 967–974. [CrossRef] [PubMed]

61. Tomlinson, D.L.; Wilson, J.G.; Harris, C.R.; Jeffrey, D.W. Problems in the assessment of heavy-metal levels in estuaries and the formation of a pollution index. *Helgoländer Meeresuntersuchungen* **1980**, *33*, 566–575. [CrossRef]

62. Norris, G.; Duvall, R.; Brown, S.; Bai, S. *EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide*; U.S. Environmental Protection Agency, Office of Research and Development: Washington, DC, USA, 2014.

63. ESRI. *ArcGIS Desktop: Release 10*; Environmental Systems Research Institute: Redlands, CA, USA, 2019.

64. Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists*, 2nd ed.; John Wiley & Sons Ltd.: Chichester, UK, 2007.

65. Brus, D.J.; Kempen, B.; Heuvelink, G.B.M. Sampling for validation of digital soil maps. *Eur. J. Soil Sci.* **2011**, *62*, 394–407. [CrossRef]

66. Vašát, R.; Pavlů, L.; Borůvka, L.; Drábek, O.; Nikodem, A. Mapping the topsoil pH and humus quality of forest soils in the north bohemian Jizerské hory Mts. Region with ordinary, universal and regression kriging: Cross-validation comparison. *Soil Water Res.* **2013**, *8*, 97–104. [CrossRef]

67. Karami, M.; Afyuni, M.; Khoshgoftarmanesh, A.H.; Papritz, A.; Schulin, R. Grain zinc, iron, and copper concentrations of wheat grown in central Iran and their relationships with soil and climate variables. *J. Agric. Food Chem.* **2009**, *57*, 10876–10882. [CrossRef]

68. Kabata-Pendias, A.; Wiacek, K. Excessive uptake of heavy metals by plants from contaminated soil. *Soil Sci. Ann.* **1985**, *36*, 33.

69. Guagliardi, I.; Cicchella, D.; De Rosa, R. A geostatistical approach to assess concentration and spatial distribution of heavy metals in urban soils. *Water Air Soil Pollut.* **2012**, *223*, 5983–5998. [CrossRef]

70. Kabata-Pendias, A.; Pendias, H. *Biogeochemistry of Trace Elements*, 2nd ed.; Wyd. Nauk PWN: Warsaw, Poland, 1999.

71. Černik, J.; Kunc, J.; Martinat, S. Territorial-technical and socio-economic aspects of successful brownfield regeneration: A case study of the liberec region (Czech Republic). *Geogr. Tech.* **2016**, *11*, 22–38. [CrossRef]

72. Novák, M.; Erel, Y.; Zemanová, L.; Bottrel, S.H.; Adamová, M. A comparison of lead pollution record in Sphagnum peat with known historical Pb emission rates in the British Isles and the Czech Republic. *Atmos. Environ.* **2008**, *42*, 8997–9006. [CrossRef]

73. Oulehle, F.; Hruska, J. Tree species (*Picea abies* and *Fagus sylvatica*) effects on soil water acidification and aluminium chemistry at sites subjected to long term acidification in the Ore Mts., Czech Republic. *J. Inorg. Biochem.* **2005**, *99*, 1822–1829. [CrossRef]

74. Wei, W.; Ma, R.; Sun, Z.; Zhou, A.; Bu, J.; Long, X.; Liu, Y. Effects of Mining Activities on the Release of Heavy Metals (HMs) in a Typical Mountain Headwater Region, the Qinghai-Tibet Plateau in China. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1987. [CrossRef] [PubMed]

75. Maján, G.; Kozak, M.; Püspöki, Z.; McIntosh, R.; Miko, L. Environmental geological examination of chromiumcontamination in Eastern Hungary. *Environ. Geochem. Health* **2001**, *23*, 229–233. [CrossRef]

76. Glasshiem, E. Most, the Town that Moved: Coal, Communists and the 'Gypsy Question' in Post-War Czechoslovakia. *Environ. Hist.* **2007**, *13*, 447–476. [CrossRef]

77. Zuna, M.; Mihaljevic, M.; Sebek, O.; Ettler, V.; Handley, M.; Navrátil, T.; Goliás, V. Recent lead deposition trends in the Czech Republic as recorded by peat bogs and tree rings. *Atmos. Environ.* **2011**, *45*, 4950–4958. [CrossRef]

Health risk assessment and the application of CF-PMF: a pollution assessment–based receptor model in an urban soil

Prince Chapman Agyeman[1] · Samuel Kudjo Ahado[1] · Kingsley John[1] · Ndiye Michael Kebonye[1] · Radim Vašát[1] ·
Luboš Borůvka[1] · Martin Kočárek[1] · Karel Němeček[1]

Abstract

Purpose This study was carried out to assess human health risk exposure, to apply a novel pollution assessment–based receptor model CF-PMF (contamination factor-positive matrix factorization), and to estimate the extent of contamination across seven cities in the Frydek-Mistek district. Nevertheless, the impact of agricultural production and industrial activities on urban soil and the livelihood of the indigenous peoples in the study area as well as the source contribution of the individual PTEs is unknown.

Methods This study collected 49 soil samples across seven towns in the Frydek-Mistek district, which are primarily agricultural and industrially oriented urbanized communities. The samples were air-dried, and the potentially toxic elemental (PTEs) (i.e., Pb, As, Cr, Ni, Mn, Cu, and Zn) concentrations measured using portable x-ray fluorescence.

Results Nemerow Pollution index and modified contamination degree indicated that the urban contamination levels were between low and moderate contamination level with a few cases of high contamination levels. The degree of contamination and the contamination factor showed varying levels of contamination for PTEs, with a high level of contamination and a low to high level of contamination, respectively. PTEs displayed a low to high pattern of spatial distribution in urban soil around Trinec, Bystice, Hrcava, and Harirov. The source of the PTEs was detected using principal component analysis, and the source apportionment of the PTEs was further assessed using CF-PMF (contamination factor-positive matrix factorization). Comparison of the CF-PMF receptor model and the EPA-PMF receptor model revealed that the novel receptor model performed better. The root mean square error (RMSE) and the mean absolute error (MAE) of the new receptor model marginal errors reduced significantly. RMSE and MAE for the CF-PMF receptor model for all the PTEs for instance As, Cr, Cu, Mn, Ni, Pb, and Zn are 11.56, 97.85, 17.30, 527.26, 37.16, 32.12, and 68.02 (RMSE) and 11.58, 95.00, 17.26, 520.85, 37.04, 32.13, and 68.03 (MAE) were lesser than the EPA.PMF receptor model respectively.
Health risk computed indicated that there was no potential carcinogenic and non-carcinogenic risk being exposed to the people living within the study area.

Conclusion We propose using the novel receptor model CF-PMF because its output has shown to be optimal with minimal error and improved efficiency when compared to the parent model EPA-PMF. In general, continuous introduction of agro-related inputs and other anthropogenic activities surges PTEs levels in urban soils. Thus, constructive yet efficient steps, appropriate control, and mitigation measures are required to abate pollution sources that may be sowed to the soil.

Keywords Contamination factor-positive matrix factorization · Spatial distribution · Urban soil · Health risk assessment · Principal component analysis

---

✉ Prince Chapman Agyeman
   agyeman@af.czu.cz

Extended author information available on the last page of the article

1 Introduction

Potential toxic elements (PTEs) have always been at the center stage of soil pollution due to human activities that sustain human life and survival on earth. Agyeman et al. (2020) stated that PTE is a generic lexicon given to hazardous

metal(loids) that damages the human health or the ecosystem. PTE's soil pollution has escalated over the globe (Solgi et al. 2012; Yang et al. 2018) and has gained a tremendous spotlight in modern times (Chen et al. 2015; Wang et al. 2016). Mineral ore and rocks are the origin of PTEs (Alloway 2013), which are usually natural formation like those of weathering of rocks including mineral formations and from anthropogenic impacts correlated to urbanization, industrial activities, mining, agriculture, and natural resource extraction (e.g., gold) (Alyazichi et al. 2017; Jones et al. 2019; Bhuiyan et al. 2021). According to Alloway (2013), PTEs can typically be small in nature with an elevated concentration level that is often recurrent and has a sustained bioavailable duration. These PTEs due to their bioavailability in nature and long residence period are found to be injurious to human health and the ecosystem at large. According to Burges et al. (2015), PTEs are pervasive and may be detrimental to the environment and human health due to their degree of noxiousness and tenacity in nature. Anthropogenic processes such as industrial development, urban sprawl, mining, and agriculture have long been at the forefront of soil pollution. According to Wei and Yang (2010) and Agyeman et al. (2020), over the last decades, toxic substance from various sources, including PTEs, have been significantly introduced into the soil because of accelerated industrial development and urban sprawl. Kabata-Pendias (2011) have reported that soil has the greatest natural environmental effect because it monitors the distribution of PTEs to the air, the hydrosphere, and the biosphere, not only as a geochemical sink for the use of pollutants, but also as a natural safety valve. However, much research has been done on PTEs, since it is a potential threat to health in society (Agyeman et al. 2020) in terms of human activities, especially in the urban areas, agricultural land, and industrial areas (Guagliardi et al. 2013). Ferri et al. (2012) reported that the soil tends to remain in its condition for a period of time following contamination due to sorption and mobility of soil particles by PTEs.

According to USEPA (1996), PTEs that are mostly uncovered in most polluted site such as urban soil in excess are As, Cd, Cr, Cu, Hg, Pb, and Zn. The rate of reaction, transportation, and the fate of the these PTEs are largely dependent on the metal speciation and the chemical forms of the metals (Wuana and Okieimen 2011). Shiowatana et al. (2001) and Buekers (2007) outlined that PTEs are adsorbed in seconds or minutes by a rapid reaction, followed by a gradual adsorption reaction process that can take days or weeks, and are disseminated in various forms such as toxicity, mobility, and bioavailability. PTEs may possibly alter soil properties particularly biological soil properties (Friedlova 2010). Nevertheless, the toxicity of the PTEs has a rippling effect on the flora and fauna of the soil. These are profoundly influenced by factors such as pH, organic matter, soil temperature, clay minerals, inorganic cation and anion

ratios, and the chemical types of PTEs (Giller et al. 1998; Šmejkalová et al. 2003). According to Levy et al. (1992), the rate of reaction of PTEs in soils such as urban soil, can potentially control the following processes: biological immobilization and mobilization, mineral precipitation and dissolution, plant uptake, aqueous complexation and ion exchange, adsorption, and desorption.

Mamut et al. (2017) and Eziz et al. (2018) argued that PTEs could potentially have an effect on humans, flora, fauna, and the food chain in the environment. Substantial research in recent years has based its studies on both health risk and ecological risk as well as the distribution of environmental impacts (Xu et al. 2014; Eziz et al. 2018; Doabi et al. 2018; Rinklebe et al. 2019; Baltas et al. 2020). PTE pollution to human from either anthropogenic or natural sources most often has a devastating health implication. Human exposure to PTEs is realized in several forms, be it dermal, ingestion, or inhalation, as the surest orifice or pathways in which the pollutants gets into human body (Ayantobo et al. 2014). The procedural and standardization procedures that human use in the field of medicine have characterized human exposure to PTEs at any level and can lead to carcinogenic and non-carcinogenic health effects (Lim et al. 2008).

Even through there are many pieces of literature published across the globe on health threats, there is a lack of documentation and research in the study area. Quite apart from that, some papers published by Gržetić and Ghariani (2008), Wang et al. (2010), Maria Figueiredo et al. (2011), Luo et al. (2012), and Bhuiyan et al. (2021) claim that health-related risk evaluation is limited in various cities such as Belgrade in Serbia, Changsha in China, Sao Paulo in Brazil, Xiamen in China, and Bangladesh, respectively. Health risk assessment is a realistic cardinal approach to assessing and evaluating risk to human health posed by PTEs by diverging pathways of exposure according to Kampa and Castanas (2008) and Bempah and Ewusi (2016).

A number of receptor models are consistently applied in the source allocation study, involving positive matrix factorization (PMF), UNMIX, principal component analysis/ absolute principal component score analysis-multi-linear regression (PCA/APCS-MLR), and chemical mass balance (CMB). In recent papers published by Salim et al. (2019), Fei et al. (2020), Wu et al. (2020), Zhang et al. (2020), Agyeman et al. (2020), and Bhuiyan et al. (2021) relied mainly on PMF or APCS/PCA-MLR or both to calculate and identify the elemental source distribution of PTEs. PMF and APCS/ PCA-MLR are chosen due to the following reasons: (i) the use of effective monitoring procedures, along with the intention of establishing a substantial database, has recently become a universal practice; (ii) these receptor models do not need pre-quantified source profiles (i.e., backward tracking) in disparity with CMB; and (iii) the capacity of the receptor models is capable of coping with significant monitoring datasets

(Lee et al. 2016). Even though PMF and APCS/PCA-MLR are frequently used, some authors have also raised concerns on the effectiveness and the efficiency of the receptor models based on their applicability to their intended purpose. Some of the constraints outlined by some of the authors are the differences in computing source contribution, differences in estimated contribution for each potential pollution source (Gholizadeh et al. 2016), inability to identify more sources (Zhang et al. 2019), and high percent error Salim et al. (2019).

The study area has a strong emphasis on indigenous health, and therefore, it is necessary and reasonable to evaluate the quality and the detail of the risk of the soil across the cities, predicated on the priority for human life, livestock, and soil health (example urban soil). The aim of this study is to estimate the degree of soil contamination and the pattern of spatial distribution of PTEs and proposes and applies a novel pollution assessment–based receptor model (contamination factor-positive matrix factor-CF-PMF) for source distribution and the assessment of carcinogenic and non-carcinogenic health risks to humans. This study seeks to answer the following research questions: What is the impact of agricultural production and industrial activities on urban soil and the livelihood of the indigenous peoples in the study area? What is the source contribution of the individual PTEs? We hypothesized that the chemical composition of the soil in the study area is less favorable to the health of the indigenes and PTE levels above the normal threshold. Nevertheless, both industrial and agricultural activities were actively carried out in the study area, and therefore, the chemical composition of prevailing pollutants and the spatial distribution across the area need to be investigated. The findings of this study would significantly contribute to understanding the risks resulting from human and livestock exposure to PTEs in cities and towns within the Moravian-Silesian Region in the Czech Republic. Furthermore, the results will motivate interested parties', indigenes of the study area, and legislators to raise awareness of the soil toxicity and health risk exposure level of the urban soil, allowing them to take corrective actions to ensure a safer environment.

2 Materials and methods

2.1 Study area

The area under study is situated in the district of Frydek-Mistek within the foothill of the Moravian-Silesian Region in the Czech Republic (Fig. 1). The community is a combination of a previous two independent towns, namely, Silesian Frydek and the Moravian Mistek, which were put together in the year 1943 and stayed since 1955. The area under study is positioned within the geographical coordinates

49° 41′ 0″ North and 18° 20′ 0″ East at an altitude between 225 and 327 m above sea level, characterized by a cold temperate climate and a high amount of rainfall even in dry months. In Frýdek-Místek, the summers are hot and partially cloudy, and the winters are cold, dry, windy, and mainly cloudy. Over the course of the year, temperatures usually range from 24 to 75 °F and are rarely below 8 °F or above 86 °F while the average annual precipitation ranges from 685 to 752 mm (Weather Spark 2016). The area survey of the district is measured at 1208 km^2 with 39.38% of the land size designated for agricultural activities and 49.36% for forestlands. The study area comprises of the following cities: Havirov, Terlicko, Trinec, Bystrica, Jablunkov, mostly Jablunkov, and Hrcava, which are affected by intensive urban farming and active industries such as the steel industry. Trinec and Vitkovice, a part of Ostrava city, where the steel industry is located, becomes an essential area for the assessment of PTEs distribution and health risk within and around neighboring communities (Agyeman et al. 2020). The soil's properties are differentiated evidently from color, structure, and carbonate content. The soil shows a medium and fine texture material that is derived from parent materials. It is mostly colluvial, alluvial, or aeolian deposits. Some part of the soils shows mottles in the top and subsoil that is primarily accompanied by concretions and bleaching. The potential toxic element pollution in the area is anticipated to occur from atmospheric deposition emitted from the steel industry nearby, vehicular emission, abrasion from tires, and agricultural activities (e.g., pesticide and insecticide applications) (Agyeman et al. 2020). Nevertheless, the dominant soil types are cambisols and stagnosols (Kozák 2010). Cambisols soil type dominate the Czech Republic; they are found at the elevation range of 455.1 to 493.5 m (Vacek et al. 2020).

2.2 Soil sampling and analysis

A total sample of 49 topsoil was obtained across seven towns (Havirov, Terlicko, Trinec, Bystrice, Jablunkov, mostly Jablunkov, and Hrcava) situated within the district of Frydek-Mistek. The sample design adopted for sampling was the regular grid, and the soil sample intervals were 2 x 2 km using the handheld GPS device (Leica Zeno 5 GPS) at a depth 0 to 20 cm. The collected sample was placed in Ziploc bags, well labeled accordingly, and transported to the laboratory. The collected samples were air-dried, crushed by a mechanical device (Fritsch disc mill pulverize), and then sieved (< 2 mm), to obtain a pulverized sample. These samples were then scanned including a three-beam system (Weindorf et al. 2013) for the elemental concentration of As, Cr, Cu, Mn, Pb, Ni, and Zn using a portable X-ray fluorescence spectrometer (Delta Premium XPD 6000, OLYMPUS INNOV-X,
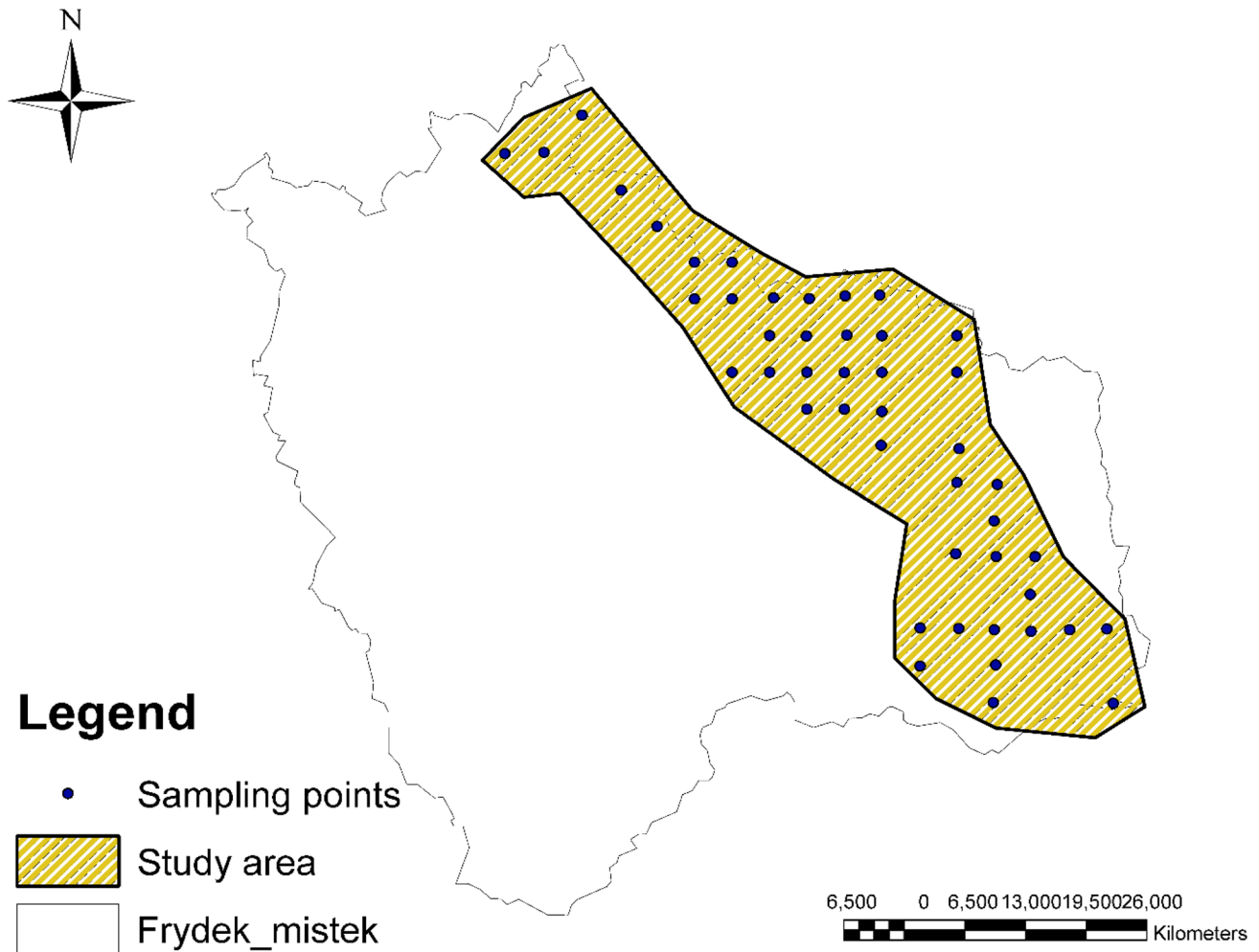
Fig. 1 Study area map showing sampling points

USA). Each sample was measured in triplicates with the average measurement computed for each sample at the end of the analysis.

2.3 Quality assurance and quality control (QA/QC)

The quality assurance and control process, the standard reference material for a portable device (i.e., XRF 2711a NIST, the National Institute of Standards and Technology), was used in the analysis to ensure quality compliance. Reference material was measured intermittently together with the soil samples to ensure that the analysis remained precise until completion. The detection limits for the elements been research on are < 10 mg/kg (Ni), < 10 mg/kg (Cu), < 5 mg/kg (As), < 10 mg/kg (Mn), < 10 mg/kg (Cr), < 5 mg/kg (Pb), and < 5 mg/kg (Zn).

2.4 Contamination assessment indices

2.4.1 Contamination degree (C_{deg})

This is a contamination assessment tool used in computing the degree of contamination index. This was introduced by Håkanson (1980), and it is defined as the summation of contamination factors for all PTEs analyzed reflecting the degree of environmental contamination. The equation is given by

$$C_{deg} = \sum_{i=1}^{n} C_f \tag{1}$$

whereby C_f represents contamination factor and n the number of PTEs analyzed.

C_f is employed to compute the contamination level of toxic chemicals such as PTEs in soil or sediments centered

on the concentration level in the sample to the geochemical background level. The equation is given as

$$c_f = \frac{c_{sample}}{c_{geo-background}} \tag{2}$$

whereby C_{sample} denotes the PTE concentration in the soil, and $C_{geo-background}$ refers to the geochemical background level. The geochemical background level used was selected from the European average value (EAV) (Kabata-Pendias 2011) (refer to Table 1). The contamination level can be categorized according to their values from 1 to 6: if CF < 1, low pollution; 1 < CF < 3, moderate pollution; 3 < CF < 6, considerable pollution; CF > 6, very high pollution. The interpretation of C_{DEG} values is given as $C_d < 6 =$ low degree, $6 < C_d < 12 =$ moderate degree of contamination, $12 < C_d < 24 =$ considerable degree of contamination, and $C_d > 24 =$ high degrees of contamination.

2.4.2 mCd

Abrahim and Parker (2008) first used this index. The index allows the assessment of the total contamination of the soil PTEs corresponding to the sum of the total contamination factor (Cf) to the given number of PTEs divided by the number of PTEs analyzed. This index is calculated by the sum of the content of PTEs at a given location. This is given by

$$mCd = \frac{\sum_{i=1}^{n} Cn}{n} \tag{3}$$

in which n denotes the number of analyzed PTEs and Cn the individual PTEs concentration.

The classification of modified degree of contamination (mCd) values is given as < 1.5 very low contamination, 1.5–2 = low contamination, 2–4 = moderate contamination, 4–8 = high contamination, 8–16 = very high contamination, 16–32 = extremely high contamination and > 35 = ultra-high contamination.

2.4.3 Nemerow pollution index ($PI_{Nemerow}$)

$PI_{Nemerow}$ computes the overall degree of pollution of the soil that consists of the concentration of all analyzed PTEs (Qingjie et al. 2008). The index is used in the assessment of for both the O and A horizons. The formula is given by

$$PI_{Nemerow} = \frac{\sqrt{(1/n \sum_{i-1}^{n} PI)^2 + PI_{max}^2}}{n} \tag{4}$$

where PI represents the computed values for the single pollution index and P_{max} the maximum values for the single pollution index of all the PTEs; the interpretation of $PI_{Nemerow}$ class values is given as $\leq 0.7 =$ clean, 0.7–1 = warning list, 1–2 = slight pollution, 2–3 = moderate pollution and $\geq 3 =$ heavy pollutions. However, single pollution index (PI) is defined as the concentration of the PTE in the sample to its geochemical background level. The equation is given by

$$PI = \frac{C_n}{B_n} \tag{5}$$

in which the c_n represents the concentration of the sampled PTEs, and B_n also denotes the geochemical background level of the same elements. The geochemical background level used was selected from the European average value (EAV).

2.4.4 PMF receptor model

Positive matrix factorization (PMF), EPA–PMF v5.0 (U.S. Environmental Protection Agency 2014), is a mathematical technique, a receptor model used in computing the contribution of the source of samples built on the composition or fingerprints of the sources. The input files are composed of the concentration and uncertainty values of the samples. Matrix X concentration is given as

$$X = GF + E \tag{6}$$

Table 1 The concentration of PTEs in the study area, basic statistics, and background level of toxic elements

Elements	Cr	Ni	Cu	As (mg/kg)	Mn	Pb	Zn
Mean	62.14	16.25	21.25	9.76	674.97	50.23	95.69
*UCC	100.00	20.00	55.00	1.80	900.00	15.00	70.00
*WAV	59.50	29.00	38.90	6.83	488.00	27.00	70.00
*EAV	94.80	37.00	17.30	11.60	524.00	32.00	68.10
CV %	22.95	63.47	43.68	46.41	31.19	50.17	38.75
Minimum value	44.00	0.00	8.87	5.60	319.67	28.47	53.17
Maximum value	111.67	43.00	58.33	28.67	1335.67	180.67	270.33

*Kabata [11] (page 41 and 42) Upper continental crust (UCC), world average value (WAV)

European average value (WAV), coefficient of variability (CV

in which G ($m \times p$), F ($p \times n$), and E ($m \times n$) represents the concentration factor matrices, for the source profile species and uncertainty.

The determination of the contribution, as well as profiles factors, is given by this equation

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{\varepsilon_{ij}}{u_{ij}} \right)^2 \qquad (7)$$

whereby m refers to the number of PTEs investigated, n signifies the number of soil samples, and U_{ij} means the uncertainty of PTEs j in soil sample i.

The authors have previously described the function of the minimum Q and the uncertainty and explaining the parameters involved as well as the implementation techniques (Agyeman et al. 2020).

2.4.4.1 CF-PMF The pollution assessment–based receptor model, contamination factor receptor (CF-PMF), is a novel receptor model which is based on the PMF model, but its determination utilizes the computed CF values of the respective PTEs under investigation instead of the raw data gotten from the field. The CF-PMF receptor model equation is given as

$$C_{f = \frac{(C_{sample})_{ij}}{(C_{geo-background})_i}} \qquad (8)$$

in which the $(C_{sample})_{ij}$ is the calculated total contamination factor of the PTEs from the jth source in the ith sampling area, $(C_{geo-background})_{ij}$ also represents the examined single PTE concentration in the assessed environment in the jth source from the ith sampling location, and $(B_n)_i$ denotes the geochemical background values of the respective analyzed PTEs in the reference environment of the reference PTEs.

2.5 Health risk assessment

The presence of industries, productive agriculture, and other anthropogenic factors exposes individuals within the study area to PTEs. Based on the risk of people being exposed, inhalation, ingestion, and dermal are three known pathways that the residents can be exposed to. According to Wang et al. (2017), in urban, peri-urban, and rural areas, three ways can be used to evaluate the risk of PTEs described below. The following equations specify the pathways of human exposure by PTEs.

$$CDI_{ing} = \frac{C \times IR_{ing} \times EF \times ED}{BW \times AT} 10^{-6} \qquad (9)$$

$$CDI_{inh} = \frac{C \times IR_{inh} \times EF \times ED}{PEF \times BW \times AT} \qquad (10)$$

$$CDI_{derm} = \frac{C \times SA \times AF \times ABS \times EF \times ED}{BW \times AT} \times 10^{-6} \qquad (11)$$

$$CDI_{total} = CDI_{ing} + CDI_{inh} + CDI_{derm} \qquad (12)$$

The definition of the parameters for CDI_{ing}, CDI_{inh}, and CDI_{derm} and reference values of the indices above Eqs. (8, 9, 10, and 11) are listed in the table (see ST1).

2.6 Non-carcinogenic risk assessment

The potential non-cancerous risk of one PTE was determined as the hazard quotient (H.Q), in which the equation is given by

$$HQ = \frac{CDI_{total}}{RfD} \qquad (13)$$

where Rfd implies the reference dosage (mg/kg/day) and is the projected daily humans' exposure. The computational HQ values were used to assess the detailed risk to health of all the PTEs studied. The values were summed up and expressed by equation as a hazard index (HI), which is given by Eq. 14:

$$HI = \sum HQ = HQ_{ing} + HQ_{inh} + HQ_{derm} \qquad (14)$$

In which HQ_{ing}, HQ_{inh}, and HQ_{derm} correspond to the hazard quotient for ingestion, inhaling, and dermal. USEPA report (US EPA 1989) specifically asserted that if HI < 1, there is a possibility to have a negative effect upon a person's health who is exposed to PTEs. Eziz et al. (2018), however, reported that there are also, non-carcinogenic health risks when the HI > 1 occurs.

2.7 Carcinogenic risk assessment

The USEPA report (EPA 2002) stressed that a human exposed to carcinogenic risk (CR) could increase the likelihood of developing cancer of any form. Equations 13 and 14 were used to calculate PTEs such as As, Ni, and Cr carcinogenic risk.

$$CR = DCI \times SF \qquad (15)$$

$$TCR = \sum CR = CR_{ing} + CR_{inh} + CR_{derm} \qquad (16)$$

where CR, TCR, and SF values symbolize carcinogenic risk (no unit), total carcinogenic risk (no unit), and slope factor for carcinogenic PTEs (mg/kg/day) respectively. The value of the TCR should differ from 1 to 10–6 to 1 to 10–4 in value. This is the acceptable criterion which shows that human health is not significantly endangered (Hu et al. 2012).

2.8 Data analysis

Statistical analyses were conducted using kyplot, PMF EPA 5.0 for source distribution estimation, Excel for possible health risk estimation, RStudio for mapping, principal component analysis (PCA), and Pearson's correlation matrix estimation. PCA is applied with the aim of finding a collection of low-sized orthogonal base functions known as principal components (PCs) (Jollife and Cadima 2016). PCA also shows the principle of similarity of findings and variables by showing them in maps as points. Furthermore, there is a smaller new collection of uncorrelated variables, also called PCA scores, which represent the original variables of interest (John et al. 2021).

The multiple linear regression model (MLR) is a regression model that encompasses the relationship between a response variable and a large number of predictor variables by using linearly integrated parameters computed using the least squares method. Following the selection of an explanatory variable, the least square model is a prediction function directed toward a soil property in MLR. In order to construct a linear relationship using the explanatory variable, PTE was used as a response variable. The factor scores served as predictors, and the PTEs served as the response variable. The number of samples used in this analysis was 49, and the scale was set between 0 and 1, indicating low and high values. A random approach was used to divide the data into a test dataset (with 25% for validation) and a training dataset (75% for calibration). The models were subjected to a tenfold cross-validation procedure, which was repeated five times. Mean absolute error (MAE), root mean square error (RMSE), and R square or coefficient determination (R^2) were used to assess the receptor models. To evaluate the best receptor model using the validation parameters, the R^2 value must be high, and the closer the value is to 1, the higher the accuracy.

The ordinary kriging (OK) geostatistical interpolation technique was used. This interpolation technique enabled us to estimate the spatial distribution of PTEs in the location under investigation. Kriging is an interpolation that predicts values of a variable at locations where data are not available based on the spatial pattern of the available data (Bishop and McBratney 2001). It is expressed by this equation:

$$Z'(\S_0) = \sum_{i=1}^{n} \lambda_i . Z(\S_i) \tag{17}$$

in which $Z'(\S_0)$ is the interpolated value for point $\S_0$, $Z(\S_i)$ denotes the known value, and λ_i represents the kriging weight for the $Z(\S_i)$ values. It can be computed by the semi-variance function of the variables on the condition that the estimated value is unbiased and optimal. The semivariogram model is expressed as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{n} [Z(X_i) - Z(X_i + h)]^2 \tag{18}$$

whereby $\gamma(h)$ signifies semi-variance, $N(h)$ denotes point group number at distance h, $Z(x_i)$ represents numerical value at position x_i, and $Z(x_i + h)$ is the numerical value at a distance $(x_i + h)$.

3 Results

3.1 PTE concentration in soil

The concentration of the PTEs in the soil decreased in the following order $Mn > Zn > Cr > Pb > Cu > Ni > As$ (see Table 1). The general mean concentration of the PTEs of the current study juxtaposed with the European average background (EAV) level, particularly Cu, Mn, Pb, and Zn was higher than the EAV tolerable limit. The mean concentration of copper in the present study is 1.23 higher than EAV (see Table 1), likewise Mn (1.28), Pb (1.59), and Zn (1.4). Alternatively, the mean concentration of the following PTEs Cr, As, Mn, Pb, and Zn in Table 1 indicated that the world average value (WAV) (Kabata-Pendias 2011) of the same elements was lower than the mean concentration of the elements of this present study. The present research PTE (Cr, As, Mn, Pb, and Zn) concentration levels are higher with a magnitude of 1.04, 1.4, 1.38, 1.86, and 1.36 times than respective values of WAV. Furthermore, the elements studied juxtaposing with the same elements of upper continent crust (UCC) (Table 1) exceeded some of the PTEs of the present study in exception of As, Pb, and Zn. Comparatively, As, Pb and Zn concentration exceeded that of UCC by the size of 5.44, 3.33, and 1.36 time higher than the respective values.

According to Karimi Nezhad et al. (2015), the coefficient of variance (CV) suggests the degree of variability within the concentrations of PTEs. $CV \leq 20\%$ indicates low variability, $21\% \leq CV \leq 50\%$ is considered as moderate variability, $50\% \leq CV \leq 100\%$ suggests high variability, and CV above 100% is regarded as exceptionally high variability. The coefficient of variation (CV %) of the PTEs in the current soils decreases in this order $Ni > Pb > As > Cu > Zn > Mn > Cr$ accruing 63.47%, 50.17%, 46.41%, 43.68%, 38.75%, and 22.95%, respectively.

3.2 Chemometric approach

3.2.1 PCA and Pearson's correlation matrix

PCA was used in the pattern recognition of the principal source of PTEs pollution in the study area. It is a useful tool that can provide informative suggestion concerning PTE pathways and primary sources (Hou et al. 2013). In this research, the principal component loading's significant correlation value

was fixed at 0.65 or higher (Table 2). Based on the eigenvalues that should be 1 or more than 1, PC 1 and 2 were found to be statistically significant, accounting for 74.23% of the total data variance. The first principal component (PC1) accrued 54.58% that explains the variation in total, which comprises the following PTEs in the order Pb, Zn, As, Mn, and Cu. PC1 origination can be ascribed to a multiplicity of sources such as geogenic and anthropogenic components. Principal component 2 loading explained that 19.65% of the total variance and demonstrated that the concentration of PTEs (Cr and Ni) is associated. Hence, it suggested that Cr and Ni share a common source of contamination and more of geogenic origin with an anthropogenic boost.

The correlation matrix (see Fig. SF1) between the investigated PTEs indicated that there is a nexus between the PTEs. The correlation between the PTEs illustrated a stronger connection between the elements. Zinc (Zn) and lead (Pb) showed a stronger positive correlation of $r = 0.92$, as well as Pb and As, $r = 0.88$; Zn and As $r = 0.75$; and Mn and As, $r = 0.72$. With this, therefore, it is vital to accentuate that they probably share the same or closely related sources. In a like manner, other correlation between PTEs such as Pb and Mn ($r = 0.58$), Zn and Cu ($r = 0.55$), Cu and Ni ($r = 0.48$), Ni and Cr ($r = 0.48$), and Pb and Cu ($r = 0.46$) also showed a resilient connection specifying that the source of pollution might be related or close. Zn and Pb showed the strongest positive correlation, and the least positive correlated element was between Ni and Mn with $r = 0.06$. All the PTEs showed a positive collection without any negation.

3.2.2 CF-PMF pollution assessment-based receptor model

The CF-PMF model was used in the identification of the source in the soil and the apportionment of PTE contribution (Fig. 2). The minimum Q controls the residual matrix that ensures that the reasonable number of factors is produced. The CF-PMF used in the current paper discharged factor loadings

Table 2 The total contribution of PTEs in the principal component of the study area

Elements	PC1	PC 2
Cr	0.37	0.73
Ni	0.57	0.68
Cu	0.67	0.29
As	0.88	−0.29
Mn	0.71	−0.35
Pb	0.91	−0.26
Zn	0.9	−0.12
Eigen value	3.82	1.38
% Variance explained	54.58	19.65
Cumulative % total		74.23

that run for 20 runs. Run 8 was the selected run among the 20 runs to discharge the factor loadings as well as the percentage contribution of each PTEs in the study. Factor 1 gave high factor loading values that comprised Ni and Cr (42.7% and 46.3%, respectively). Factor 2 was dominated by Cr, Mn, and Cu with 50.6%, 39.7%, and 31.0% factor loadings, respectively. Factor 3 loading comprised As, Zn, and Pb having the factor loadings of 51.7%, 50.6%, and 60.4%, respectively. The fourth-factor loading was dominated by Ni, Mn, and Cu accruing 53.7%, 32.9%, and 36.5% loadings, respectively.

3.3 Contamination assessment of PTEs

3.3.1 Contamination and modified contamination degree

The calculated contamination assessment indices such as contamination factor, contamination degree, and modified contamination degree values of the PTEs showed a diverse degree of contamination (Table 3). The estimated contamination factor showed a contamination level from low to moderate among all the PTEs. However, Zn, Cu, and Pb showed considerable high contamination at sample point 41(FM-468–01), which is displayed on a box plot (Fig. 3 and see Table ST3). Contamination degree (C_{deg}) computation is given in Table 3, and Ni showed a considerable degree of contamination. The other PTEs indicated a very high degree of contamination, and Zn calculated C_{deg} was very high. The degree of contamination in the soil as assessed by modified contamination degree (mCd) indicated that the Ni level of contamination was moderate; nonetheless, Cr and As showed high contamination level in the soils. Moreover, Cu, Pb, Zn, and Mn degree of contamination was very high.

3.3.2 Nemerow pollution index (PI_nemerow)

Nemerow pollution of PTEs is shown in Fig. 4 and Table ST3. Application of Nemerow pollution index to interpret the pollution level in the soil showed that some areas were least polluted by PTEs as displayed in Fig. 4. Nevertheless, some of the regions revealed the tendency of warning limit, as well as other areas, showed slight to the moderate pollution level. The northeastern (Trinec and Bystrice) and some parts of the central regions (Bystrice and Jablunkov) showed moderate to high spatial distribution pollution class as indicated in the map. This hotspot shows an active heterogeneous pollution distribution with a multiplicity of sources. The spatial distribution pattern of PI_nemerow of soils in the study area showed a sectorial distribution pattern in the soils in these towns: Trinec, Bystrice, and Jablunkov. The non-polluted regions distributed in the western and some parts of the central, north-western, and southwestern areas showed evidence of clean to low polluted areas distributed spatially in other parts of the study area.
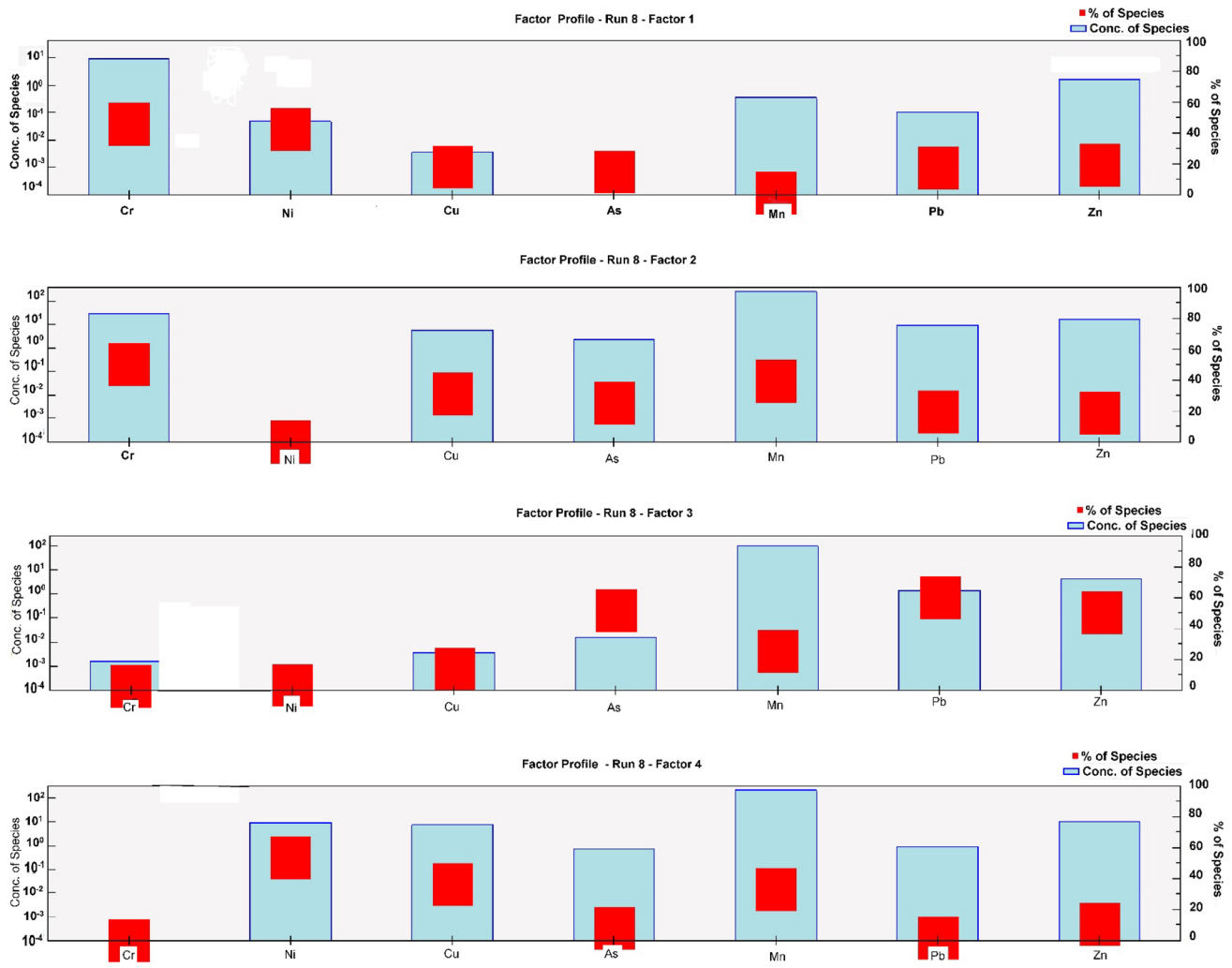
Fig. 2 Factor profile of PTEs from CF-PMF model analysis showing percentage contributions

3.4 Potential health risk

3.4.1 Non-carcinogenic risk

The computed chronic daily intake (CDI), HQ, and HI values are displayed in Table 4. The CDI_{total} distribution for the PTEs in the soils in the present studies (adult and children) is given in the following descending order: $Mn > Zn > Cr < Pb < Cu < Ni < As$ (Table 4). The computed hazard quotient (HQ) of the children seems to be higher than the adult's HQ (see Table 4), and it falls within the range of $4.7E-01$ to $6.81-03$ while that of the adults' span between the range $4.47E-02$ to $7.31E-04$. In ascending order, the estimated HQ values for the PTEs (both adults and children) is as follows: $As < Cr < Pb < Mn < Ni < Cu < Zn$ accounting for 27.96%, 1.10%, 0.72%, 43.87%, 6.49%, 19.43%, and 0.43% for children and 28.22%, 1.09%, 0.71%, 43.71%, 6.47%, 19.37%, and 0.43% for adults, respectively.

3.4.2 Carcinogenic risk

Regarding carcinogenic risk CDI, CR, and TCR were computed as shown in Table 5. The chronic daily intake was calculated for Cr, Ni, Pb, and As. The CDI total for children and

Table 3 Computed modified contamination degrees values of PTEs

Elements	Cr	Ni	Cu	As	Pb	Zn	Mn
C_{deg}	32.12	21.52	60.20	41.22	63.12	76.91	68.86
mCd	4.59	3.07	8.60	5.89	9.02	10.99	9.84

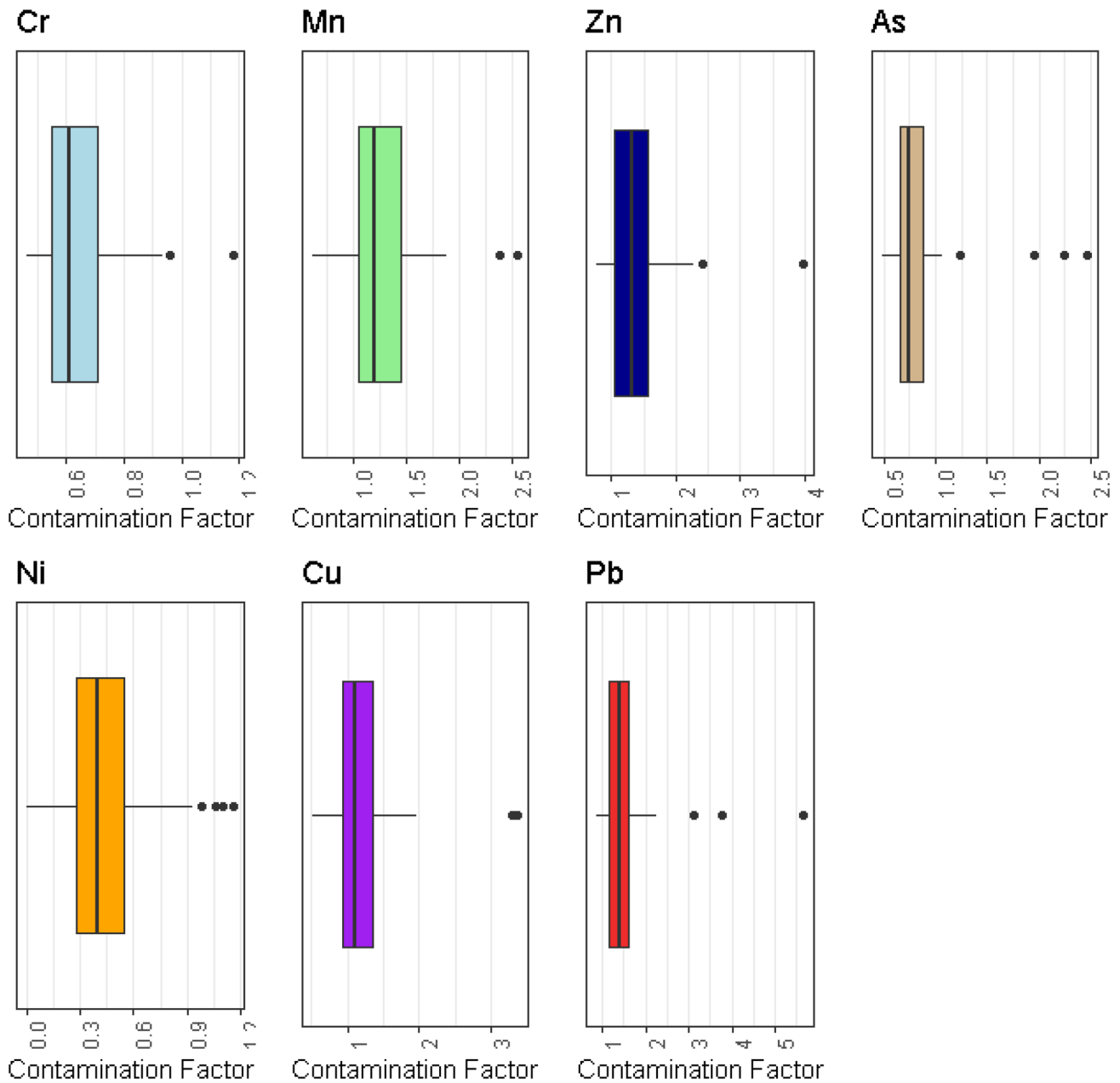**Fig. 3** Box plot showing for contamination factor (CF) of PTEs significance levels in the selected towns

adults is given in this ascending order: Cr < Pb < Ni < As. The CDI for children ranges between 5.51E − 05 and 6.82E − 05 whereas that for adults from 2.36E − 05 to 2.92E − 05.

3.5 Spatial prediction of PTEs

The spatial distribution of PTEs in the study area is shown in Fig. 5. The distribution pattern of the PTEs showed a sectorial trend of spatial variability that is skewed toward the east northern part of the map for Cu and Ni (this was toward Trinec and Bysrice town). The southeastern part (Trinec and Hrcava) and a little part of the central part skewed toward the east northern part (Trinec and Bysrice) of the map showed a hotspot for As and Mn. A larger area of the central part of that map showed a high concentration and a spatial variability pattern for Cr (Trinec and Bysrice). On the other hand, Pb and Zn showed a spatial distribution pattern at the east northern part moving downward to the southeastern part of the map (Trinec, Bysrice and Jablunkov).
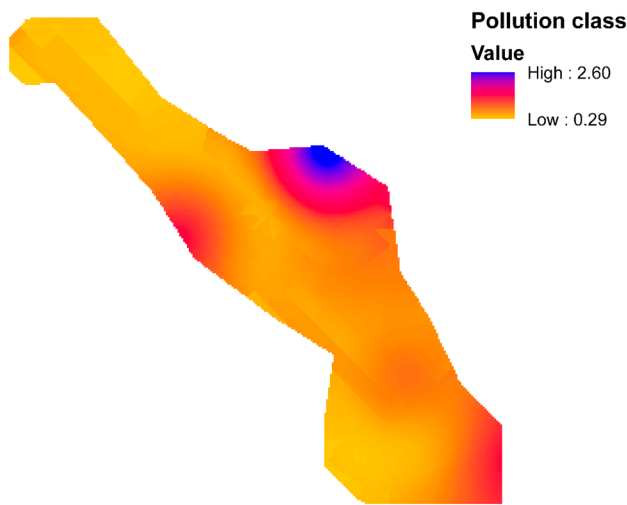
Fig. 4 Distribution map of PI$_{nemerow}$_kriging showing pollution levels in soils of the study area

4 Discussion

4.1 PTE concentration in soil

The PTEs such as Cu, Mn, Pb, and Zn showed a high pollution in the study area. The higher content of the PTEs signifies that they have multiple sources. The geogenic source cannot be ruled out; hence, there is enough evidence that proves that anthropogenic activities (steel industry and intensive agriculture) and other factors are accounting for the upsurge of the PTE concentrations. Hossain Bhuiyan et al. (2021) reported that human-related pollution is mostly caused by anthropogenic activities such as agriculture and sewage drainage, as well as industrial and air emissions. According to Jia et al. (2018), the excess of some of the PTE values in the present study to the UCC respective PTE values provides a clear indication that the source of pollution might also have an anthropogenic origin. Comparatively, the current soil concentration likened to the soil concentration of Sweden (Kabata-Pendias 2011) portrays that the present

PTE concentration of the study area exceeded similar PTE concentrations in all levels. The coefficient of variability results explained that there is a moderate variability between As, Cu, Zn, Mn, and Cr and that they are more homogeneous. Ni and Pb showed high variability, which indicates a non-homogenous distribution of Ni and Pb, which explained a probable human-related activity. The spatial distribution of the non-homogeneity of Ni and Pb foretells the presence of locally enrichment source.

4.2 Pollution assessment

The contamination factor of the PTEs revealed that 97.95 % (48 samples) of the 49 urban soils sampled had low chromium contamination, with Ni displaying 93.87 % (46 samples), Cu 34.69 % (17 samples), As 87.75 % (43 samples), Mn 18.38 % (9 samples), Pb 2.05 % (1 sample), and Zn 20.4 % (10 samples). Lead, manganese, zinc, and copper exhibited elevated moderate contamination distribution level representing 91.84%, 81.64% 77.55%, and 61.22% of the total sample, respectively. Similarly, chromium, nickel, and arsenic moderate contamination level was relatively low representing 2.05%, 6.13%, and 12.25% of the total sampled data. Contamination levels of lead, zinc, and copper of the urban soil in some locations were considerably high: 6.47%, 2.05%, and 4.09% of the overall sample that were sampled from the 49 locations, respectively.

Modified contamination degree suggests a moderate enrichment based on the cumulative and average of all the PTEs analyzed. The overall enrichment of urban soil and the resultant impact of PTEs on the soil were pervasive in the study area based on the results (see Table 3). The mCd result of Ni compared to the proposed gradations of Abrahim and Parker (2008) establishes that the cumulative average calculation of Ni is moderately contaminated which might be attributed to a geo-anthropogenic source (blend of geogenic and anthropogenic sources (steel industry)). However, arsenic and chromium showed a high level of contamination that can be related to parent materials and intensive farming. Cu, Pb, Mn, and Zn also showed a very high degree of pollution that might be associated

Table 4 Comparison assessing of model quality using multiple linear regression

	CF_PMF						
	CR_As	CR_Cr	CR_Cu	CR_Mn	CR_Ni	CR_Pb	CR_Zn
R^2	0.830	0.993	0.478	0.930	0.983	0.993	0.905
RMSE	0.179	0.013	0.433	0.113	0.038	0.073	0.180
MAE	0.100	0.011	0.228	0.081	0.023	0.062	0.127
	PMF						
	As	Cr	Cu	Mn	Ni	Pb	Zn
R^2	0.831	0.993	0.478	0.929	0.983	0.992	0.905
RMSE	2.070	1.272	7.493	59.580	1.412	2.345	12.244
MAE	1.158	1.045	3.936	42.189	0.852	1.992	8.640

Table 5 The non-carcinogenic risk index of PTEs in soils in the study area

Elements	Cr	Ni	Cu	As	Mn	Pb	Zn
Children							
CDI_{ing}	7.94E − 04	2.08E − 04	2.72E − 04	1.25E − 04	8.63E − 03	6.42E − 04	1.22E − 03
CDI_{inh}	2.23E − 08	5.84E − 09	7.64E − 09	3.51E − 09	2.43E − 07	1.81E − 08	3.44E − 08
CDI_{derm}	6.83E − 07	1.79E − 07	2.34E − 07	1.07E − 07	7.42E − 06	5.52E − 07	1.05E − 06
CDI_{total}	7.95E − 04	2.08E − 04	2.72E − 04	1.25E − 04	8.64E − 03	6.43E − 04	1.23E − 03
Adult							
CDI_{ing}	8.51E − 05	2.23E − 05	2.91E − 05	1.34E − 05	9.25E − 04	6.88E − 05	1.31E − 04
CDI_{inh}	1.25E − 08	3.27E − 09	4.28E − 09	1.97E − 09	1.36E − 07	1.01E − 08	1.93E − 08
CDI_{derm}	9.12E − 08	2.38E − 08	3.12E − 08	1.43E − 08	9.9E − 07	7.37E − 08	1.40E − 07
CDI_{total}	8.52E − 05	2.23E − 05	2.92E − 05	1.34E − 05	9.26E − 04	6.89E − 05	1.31E − 04
Children							
HQ_{ing}	2.65E − 01	1.04E − 02	6.79E − 03	4.16E − 01	6.16E − 02	1.85E − 01	4.08E − 03
HQ_{inh}	7.81E − 04	2.84E − 07	1.91E − 07	1.17E − 08	4.85E − 06	5.13E − 06	1.15E − 07
HQ_{derm}	2.28E − 05	3.31E − 05	1.95E − 05	8.72E − 04	9.28E − 06	1.05E − 03	1.75E − 05
HQ	2.66E − 01	1.04E − 02	6.81 − 03	4.7E − 01	6.17E − 02	1.85E − 01	4.10E − 03
HI	0.95						
Adult							
HQ_{ing}	2.83E − 02	1.11E − 03	7.28E − 04	4.46E − 02	6.60E − 03	1.96E − 02	4.37E − 04
HQ_{inh}	4.38E − 04	1.59E − 07	1.07E − 07	6.55E − 09	2.72E − 06	2.87E − 06	6.43E − 08
HQ_{derm}	3.04E − 05	4.42E − 06	2.6E − 06	1.16E − 04	1.24E − 06	1.40E − 04	2.34E − 06
HQ	2.88E − 02	1.12E − 03	7.31E − 04	4.47E − 02	6.61E − 03	1.98E − 02	4.39E − 04
HI	0.10						

to the intensive urban crop production on urban soil, the application of livestock manure on the urban farmland, and the steel industry. Despite the fact that the parent material's contribution to higher levels of Cu, Pb, Mn, and Zn in urban soil is undeniable fact. However, Bhuiyan et al. (2011) posited that PTEs with higher pollution levels demonstrate an anthropogenic impact.

The Nemerow pollution index displayed various color patterns indicating differing levels of pollution on the urban soil. But the most contaminated area on the PInemerow distribution map is where the steel plant is situated. The areas that are in proximity with the steel plant also revealed a relatively high pollution pattern, implying that the steel plant seems to be the major pollutant source within that environ.

4.3 PCA

Figure 6 shows the projection of the clustered PTEs and the relationship fostered between the PTEs. The high *r* values of Pb and As (PC 1) indicate that they may share the same or close related source which might be more of anthropogenic than geogenic (see Fig. 6). However, lead and arsenic are agronomically related in agrochemicals, such as lead arsenate pesticides or herbicides, which are an essential source of chemicals in urban agricultural soil (Franco-Uría et al. 2009). Previous studies from Nicholson et al. (2003) and Luo et al. (2009) outlined that fertilizer and livestock manures are an essential source for both Pb and As, and this is coherent with

the present findings in the urban soil. Zn, Mn, and Cu (0.90, 0.71, and 0.67 respectively) source of occurrence may be attributed to a combination of the geogenic and anthropogenic source (liming). Zinc displayed a more definite correlation matrix with the other two PTEs (Mn and Cu). According to Mantovi et al. (2003), Cu and Zn concentrations in soil surges in relation to the application of wastes derived from animal farming and fertilizer application. The enrichment of the PC 2 PTEs (Cr and Ni) proposed that PC 2 might be controlled primarily by a parent material with a hinge to an anthropogenic source. The anthropogenic source of Cr and Ni could be appropriated to agricultural fertilizer that is in accordance with research carried by USEPA (2002) as part of the central metal contaminants.

4.4 CF-PMF receptor model

The dominance of Cr and Ni (factor 1) in the urban soil can be ascribed to geogenic- and anthropogenic-related sources (such as the steel industry where it is predominantly used for alloy for formation and other agricultural-related activities such as slaked lime). The current results are in accordance with similar studies by Veit et al. (2009) and Saha et al. (2011) outlining that slaked lime NPK plays a role in enriching the soils with Cr. Nevertheless, previous studies by Zhang et al. (2016) revealed that high Cr concentration in agricultural soils that surpasses the maximum acceptable
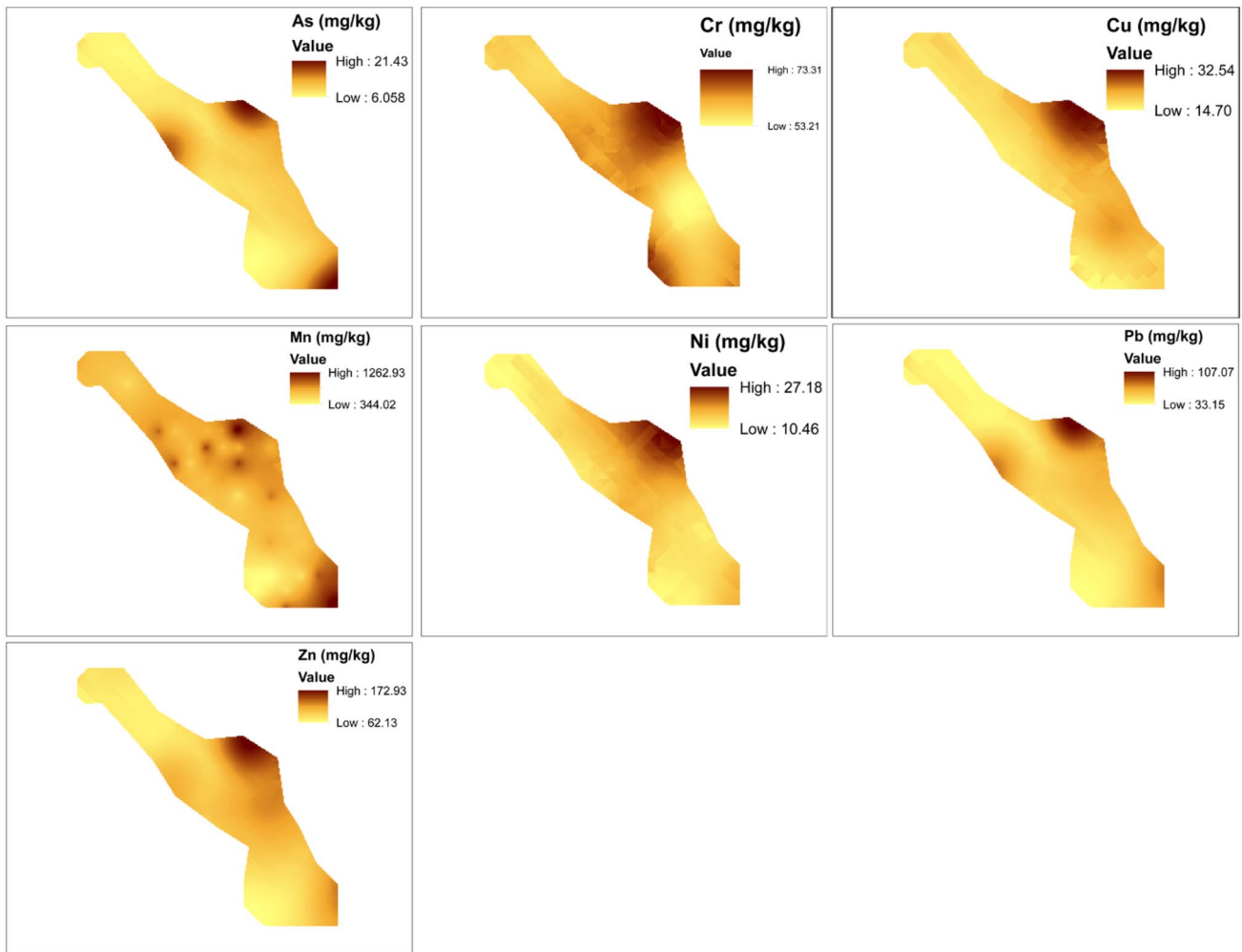
**Fig. 5** The spatial distribution of PTEs in the soil

limit is not limited to the agro-related source. However, higher Cr level is due to collaboration with other tenants such as the geogenic source. This assertion (that geogenic source) was confirmed by Manta et al. (2002) and Mamat et al. (2014) in their studies. Beyond that, other studies by Li et al. (2009) and Liu et al. (2015) mentioned that Cr concentration in agricultural soil increases by the application of sewage irrigation to farmlands. Industrial activities such as steel industries and smelting ores are also contributing to Cr and Ni enrichments in soil. Several studies in China, for instance, Gansu province, discovered that non-ferrous metal and steel production pollutes the soil with Cr and Ni coming from industrial activities (steel industries, smelting ore) (Chen et al. 2016; Qu et al. 2013).

Copper accumulation in factor 2 primarily may be related to livestock manure (Nicholson et al. 2003) because combining Zn to Cu serves as an additive that improves microbial activities (anti-bacterial agent to the gut) (Rosen and Roberts 1996) as well as control, scours after weaning (Holm 1990). According to Cheng et al. (2014) and Xiong et al. (2010),

livestock manure (particularly from pigs) and phosphate fertilizers are rich in Cu which may eventually contribute to its enrichment in agricultural soil. Manganese enrichment is of geogenic source, and according to International Manganese Institute, it is the 4th most used element in tonnage after Fe, Al, and Cu (Das et al. 2011). According to Goncalves et al. (2014), Mn is ubiquitous and the 2nd and the 12th most abundant element in the Earth's crust.

The source of Pb and Zn from factor 3 largely might have originated from agriculture, vehicular traffic, fuel knocking, and abrasion from car tires. Previous studies from Li et al. (2001) and Tepanosyan et al. (2016) share similar assertion that Pb and Zn enrichment may be attributed to road traffic, tire abrasion, and fuel knocking as well as the minimal geogenic source. Most pesticides and herbicides such as calcium arsenate, lead arsenate, and sodium arsenate are rich in As and are used in diverse ways for agricultural production. Research conducted by Bhattacharya et al. (2007) attests to the fact that agrochemicals of such nature are highly rich in inorganic As. Similar studies accentuate that livestock

Fig. 6 The use of principal component analysis in the projection of PTE components
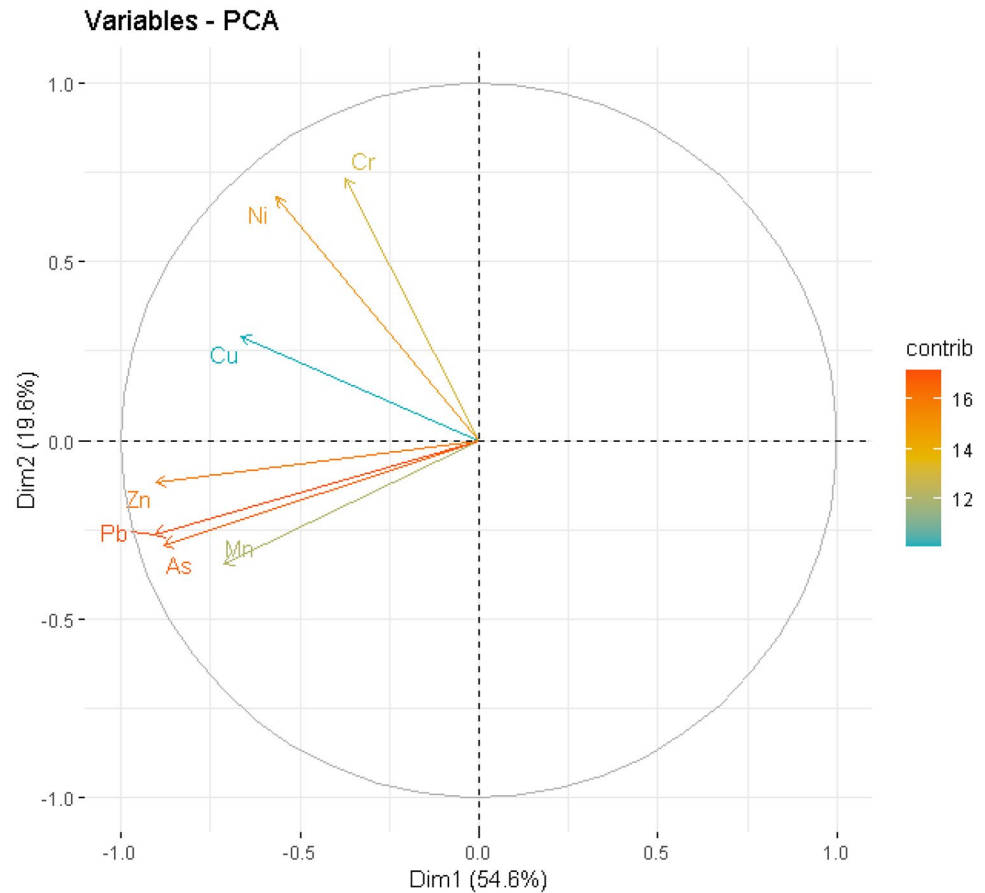
manure is also a potential source for As enrichment in the soil (Micó et al. 2006; Fang et al. 2011). Factor 4 is a blend of anthropogenic and geogenic sources.

4.5 Comparison of CF-PMF receptor model to EPA-PMF receptor model

One of the most used pollution assessment indices in assessing the soil quality and extent of contamination with specific PTEs in an urban area is the contamination factor. According to Kowalska et al. (2018), in assessing the contamination level of a specific PTE, CF is one of the most analytical techniques in assessing the soil quality of an urban area. The positive matrix factorization receptor model is a robust receptor model, but the hybridization of PMF and CF increases the source apportionment efficiency and minimizes the error. Comparative assessment the hybrid model and PMF exhibited that consistently, the hybrid model performed better than the parent model. The estimated coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) suggested that out of the seven PTEs evaluated, CF-PMF showed superior performance in all the seven PTEs (Table 4). All the PTEs analyzed in CF-PMF have lower error level as compared to PMF. According to Li et al. (2016), for

a model to deemed as a good model, the prediction value R^2 value should be 0.75 or greater. However, RMSE and MAE values on the other hand should be close to 0 or infinitesimally small. Evaluating the models, it was evident that CF-PMF receptor model performed better than PMF receptor model. The CF-PMF receptor model is an improvement of the PMF receptor model thereby amplifying the efficiency of source apportionment estimation as well as decreasing marginal error significantly. The errors with regard to RMSE for the CF-PMF receptor model for all the PTEs such as As, Cr, Cu, Mn, Ni, Pb, and Zn are 11.56, 97.85, 17.30, 527.26, 37.16, 32.12, and 68.02 lesser than the EPA.PMF receptor model respectively. Similarly, the MAE error with the CF-PMF receptor model for all PTEs such as As, Cr, Cu, Mn, Ni, Pb, and Zn is 11.58, 95.00, 17.26, 520.85, 37.04, 32.13, and 68.03 less than the EPA.PMF receptor correspondingly.

4.6 Potential health risk

The non-carcinogenic intake (CDI_{total}) of adult and children is presented in Table 5, and the CDI_{total} values of the PTEs of children compared to that of the adults indicates that that of the children is a bit higher to that of the adults, since children are more exposed than the adults and are more prone

to possible hurt than the adults. Children and adult CDI oral ingestion was the highest among the other CDIs computed. Earlier reports by Fang et al. (2011) and Karim and Qureshi (2014) and a more recent report by Bhuiyan et al. (2021) confirm the same results and proceed to report that oral or ingestion remains the utmost exposure pathway of PTEs into the human body. In this present study, the total HI value estimated for children is 9.5 times higher than that of the adults (see Table 5). It presupposes that children are more susceptible and more sensitive to the health effects of PTEs because their mouth and finger practices tend to increase their rate of exposure to PTEs. Numerous studies regarding health risk have reported similar high HI results for children (Baltas et al. 2020; Rinklebe et al. 2019; Varol et al. 2020; Wu et al. 2018). For instance, Varol et al. (2020) reported 8.44E−01 for children to 9.85E−02 for adults, and Baltas et al. (2020) reported 1.21 for children to 0.131 for adults. The computed HI for children was 0.95, which is less than the threshold of 1; therefore, it implies that it is unlikely for the PTEs to have a negative impact to an exposed individual (Kusin et al. 2018). Similarly, the calculated HI for an adult is not significant because it is equally less than the threshold 1, which points out that it is unlikely for non-carcinogenic negative impact to befall an individual if exposed. It is vital to note that the calculated HI results are from summing up all the elemental HQs assessed; therefore, if most HQs are high, this may result in a high HI and vice versa.

The CDI total for children (CR) is higher than that of the adults, irrespective of the estimated value of the PTEs. Children are open to multiple exposure pathways than adults, and children being exposed to PTEs leads to diverse health issues such as cardiovascular disease, poor respiratory function, neurodevelopmental deficits, and skeletal damage as well as reproductive toxicity (Madrigal et al. 2018). The CDI oral ingestion for adults and children were also found higher than the other computed CDIs. Moreover, the CDI$_{total}$ for children was found higher than that for the adults (Table 6). The CDI$_{total}$ and TCR for As of the adults were found lower than that of the children. Computed TCR for children was found 2.33 times higher than that of the adults. The total carcinogenic risk for the adults and children was 6.9E−06 and 1.61E−05, respectively. The computed TCR for both adults (6.9E−06) and children (1.61E−05) pointed out that carcinogenic health risk within the study area falls within the acceptable limits (TCR values should range between 1×10^{-6} and 1×10^{-4}). Therefore, the propensity for indigenes within the enclave of the urban soils to be exposed to carcinogenic related health risk is unlikely.

4.7 Spatial prediction of PTEs

The concentration of Cu and Ni pointed out that its enrichment primarily can be attributed to the steel industry and

Table 6 The carcinogenic risk index of PTEs in soils in the study area

Adults	Elements			
	Cr	Ni	As	Pb
CDI$_{ing}$	2.92E−05	7.63E−06	4.58E−06	2.36E−05
CDI$_{inh}$	4.29E−09	1.12E−09	6.74E−10	3.47E−09
CDI$_{derm}$	3.13E−08	8.18E−09	4.91E−09	2.53E−08
CDI$_{total}$	2.92E−05	7.64E−06	4.59E−06	2.36E−05
Children				
CDI$_{ing}$	6.81E−05	1.78E−05	1.07E−05	5.5E−05
CDI$_{inh}$	1.92E−09	5.01E−10	3.01E−10	1.55E−09
CDI$_{derm}$	5.86E−08	1.53E−08	9.20E−09	4.73E−08
CDI$_{total}$	6.82E−05	1.78E−05	1.07E−05	5.51E−05
Adult				
CR$_{ing}$			6.87E−06	
CR$_{inh}$	1.80E−07	9.43E−10	1.02E−08	
CR$_{derm}$			1.80E−08	
TCR			6.9E−06	
Children				
CR$_{ing}$			1.60E−05	
CR$_{inh}$	8.04E−08	4.21E−10	4.54E−09	
CR$_{derm}$			3.37E−08	
TCR			1.61E−05	

agro-related sources (livestock manure); this is coherent with previous research conducted by Facchinelli et al. (2001). Moreover, Salonen and Korkka-Niemi (2007) outlined that some PTEs such Ni and Cu can be present in the parent materials of the soil with minute temporal and spatial distribution in worldwide soils. The enrichment of As in the soil is due to the potential application of lead arsenate and sodium arsenate to boost yield (i.e., increase yield in fruits and potatoes) which upsurges the levels of the PTEs (Frank et al. 1976). Manganese is of natural origin, but the continuous application of manganese sulfate in agricultural soil to increase yield in crops such as vegetables and beans elevate the concentration levels of PTEs (Frank et al. 1976). The enrichment of Cr is due to multiplicity of anthropogenic sources. The hotspot of Cr in the map is as a result of the steel industry usage of chromium in alloy formation as well as sewage discharge. Goovaerts (1997) hinted the source of PTEs such as Cr, Cu, and Ni; the geochemical background of these elements is normal in general, but sometimes, their elevation in soils may be influenced by agro-anthropogenic-related sources. The spatial distribution of Zn and Pb primarily is linked toward agricultural fertilizer, vehicular traffic, and fuel knocking. This is coherent with previous research by Kachenko and Singh (2006), Perez-de-Mora et al. (2006), and Rodríguez et al. (2008) reporting that Pb and Zn higher levels in urban agricultural soil are as a result of anthropogenic component constituted by human-related activities. Pb pollution is one of the critical concerns for almost half

of the sites of the US Superfund Environmental Protection Agency (EPA) according to Hettiarachchi and Pierzynski (2004), McBride et al. (2014), and Brown et al. (2016).

5 Conclusion

This study showed that some PTE concentrations, such as Cr, As, Mn, Pb, and Zn, exceeded the WAV, whereas the mean concentrations of As, Pb and Zn were higher than the UCC. However, the concentration of Cu, Mn, Pb, and Zn also exceeded the tolerable EAV limit. The PCA established the prime source of pollution in the study area and clarified that with the significant statistics of 74.23%. It suggested that the source of pollution originated from a multiplicity of origin that is from anthropogenic (mostly agricultural practices and steel industry) and geogenic sources. The CF-PMF pollution assessment–based receptor model discharged four factors, and the source distribution revealed the dominance of Ni and Cr (factor 1); Cr, Mn, and Cu (factor 2); As, Zn, and Pb (factor 3); and Ni, Mn, and Cu (factor 4). The contamination factor exhibited low to medium level of contamination for all the PTEs except for Pb, Zn, and Cu that further displayed a considerable contamination level. The contamination degree also indicated that the PTEs for Ni were considerably contaminated and contamination degrees of the other PTEs such as Cr, Cu, As, Pb, Zn, and Mn were considerably high. However, the mCd also specified that Ni was rather moderately contaminated. In addition, it placed Cr and As in the same high contamination degree categories and Cu, Pb, Zn, and Mn were very highly contaminated in the urban soil.

Nevertheless, the Nemerow pollution displayed a low to moderate level of pollution pattern, but the northeastern (Trinec and Bystrice) and certain parts of the central regions (Bystrice and Jablunkov) indicated moderate to high spatial distribution pollution class. The risk assessment of both noncarcinogenic and carcinogenic health for adults and children suggested that it is unlikely that the exposure to PTEs would have a negative effect, and there is no carcinogenic risk to the residents living within the enclave of the study area. Spatial distribution of PTEs in the study area suggested a hotspot along Trinec, Bystrica, Jablunkov, and Hrcava. Continuous use of agro-related inputs and other anthropogenic tenants, such as the steel industry, is likely to raise the urban soil PTE levels.

The comparison assessment of the novel CF-PMF receptor model based on PMF showed that combining CF to PMF has improved the receptor model's accuracy. Multiple linear regression analysis of both the EPA.PMF model and the CF.PMF model using cross validation evaluation such as coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) has consistently shown that the error level has been reduced significantly across all the PTEs analyzed. The CF-PMF receptor model has shown to be effective

and useful in the discovery and distribution of the percentage contribution of the PTEs under investigation.

In parallel with the health risk assessment, pollution assessment and the CF-PMF receptor model highlighted hotspot and risk-prone areas within the urban area, which are of great concern to the communities under investigation. However, it is important for the cities to take pragmatic measures to reduce and protect the soil from PTEs accumulation.

Declarations

Competing interests The authors declare no competing interests.

References

Abrahim GMS, Parker RJ (2008) Assessment of heavy metal enrichment factors and the degree of contamination in marine sediments from Tamaki Estuary, Auckland, New Zealand. Springer 136:227–238. https://doi.org/10.1007/s10661-007-9678-2

Agyeman PC, Ahado SK, Kingsley J et al (2020) Source apportionment, contamination levels, and spatial prediction of potentially toxic elements in selected soils of the Czech Republic. Environ Geochem Health 1–20. https://doi.org/10.1007/s10653-020-00743-8

Alloway B (2013) Heavy metals in soils: trace metals and metalloids in soils and their bioavailability

Alyazichi YM, Jones BG, McLean E et al (2017) Geochemical assessment of trace element pollution in surface sediments from the Georges River, Southern Sydney, Australia. Arch Environ Contam Toxicol 72:247–259. https://doi.org/10.1007/s00244-016-0343-z

Ayantobo et al (2014) Non-cancer human health risk assessment from exposure - Google Scholar. In: Am J Environ Sci. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Non-cancer+human+health+risk+assessment+from+exposure+to+heavy+metals+in+surface+and+groundwater+in+Igun+Ijesha%2C+Southwest+Nigeria.%22+American+Journal+of+Environmental+Sciences+10%2C+no.+3+%282014. Accessed 26 Nov 2020

Baltas H, Sirin M, Gökbayrak E, Ozcelik AE (2020) A case study on pollution and a human health risk assessment of heavy metals in agricultural soils around Sinop province, Turkey. Chemosphere 241. https://doi.org/10.1016/j.chemosphere.2019.125015

Bempah CK, Ewusi A (2016) Heavy metals contamination and human health risk assessment around Obuasi gold mine in Ghana. Environ Monit Assess 188. https://doi.org/10.1007/s10661-016-5241-3

Bhattacharya P, Welch AH, Stollenwerk KG et al (2007) Arsenic in the environment: biology and chemistry. Sci Total Environ 379:109–120. https://doi.org/10.1016/j.scitotenv.2007.02.037

Bhuiyan MA, Karmaker SC, Bodrud-Doza M et al (2021) Enrichment, sources and ecological risk mapping of heavy metals in agricultural soils of Dhaka district employing SOM. PMF and GIS Methods Chemosphere 263:128339. https://doi.org/10.1016/j.chemosphere.2020.128339AQ

Bhuiyan MAH, Bodrud-Doza M, Rakib MA et al (2021) Appraisal of pollution scenario, sources and public health risk of harmful metals in mine water of Barapukuria coal mine industry in Bangladesh. Environ Sci Pollut Res 1–18. https://doi.org/10.1007/s11356-020-11999-z

Bhuiyan MAH, Suruvi NI, Dampare SB et al (2011) Investigation of the possible sources of heavy metal contamination in lagoon and canal water in the tannery industrial area in Dhaka, Bangladesh. Environ Monit Assess 175:633–649. https://doi.org/10.1007/s10661-010-1557-6

Bishop TFA, McBratney AB (2001) A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma 103:149–160. https://doi.org/10.1016/S0016-7061(01)00074-X

Brown SL, Chaney RL, Hettiarachchi GM (2016) Lead in urban soils: a real or perceived concern for urban agriculture? J Environ Qual 45(1):26–36

Buekers J (2007) Fixation of cadmium , copper , nickel and zinc in soil : kinetics , mechanisms and its effect on metal bioavailability

Burges A, Epelde L, Garbisu C (2015) Impact of repeated single-metal and multi-metal pollution events on soil quality. Chemosphere 120:8–15. https://doi.org/10.1016/j.chemosphere.2014.05.037

Chen H, Teng Y, Lu S et al (2015) Contamination features and health risk of soil heavy metals in China. Sci Total Environ 512–513:143–153. https://doi.org/10.1016/j.scitotenv.2015.01.025

Chen T, Chang Q, Liu J et al (2016) Identification of soil heavy metal sources and improvement in spatial mapping based on soil spectral information: a case study in northwest China. Sci Total Environ 565:155–164. https://doi.org/10.1016/j.scitotenv.2016.04.163

Cheng Q, Guo Y, Wang W, Hao S (2014) Spatial variation of soil quality and pollution assessment of heavy metals in cultivated soils of Henan Province, China. Chem Speciat Bioavailab 26:184–190. https://doi.org/10.3184/095422914X14042081874564

Das A, Sukla L, Pradhan N et al (2011) Manganese biomining: a review. Elsevier

Doabi SA, Karami M, Afyuni M, Yeganeh M (2018) Pollution and health risk assessment of heavy metals in agricultural soil, atmospheric dust and major food crops in Kermanshah province. Iran Ecotoxicol Environ Saf 163:153–164. https://doi.org/10.1016/j.ecoenv.2018.07.057

EPA U (2002) Zinc fertilizers made from recycled hazardous secondary materials

Eziz M, Mohammad A, Mamut A, Hini G (2018) A human health risk assessment of heavy metals in agricultural soils of Yanqi Basin, Silk Road Economic Belt, China. Hum Ecol Risk Assess 24:1352–1366. https://doi.org/10.1080/10807039.2017.1412818

Facchinelli A, Sacchi E, Mallen L (2001) Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. Environ Pollut 114:313–324. https://doi.org/10.1016/S0269-7491(00)00243-8

Fang F, Wang H, Lin Y (2011) Spatial distribution, bioavailability, and health risk assessment of soil Hg in Wuhu urban area, China. Environ Monit Assess 179:255–265. https://doi.org/10.1007/s10661-010-1733-8

Fei X, Lou Z, Xiao R et al (2020) Contamination assessment and source apportionment of heavy metals in agricultural soil through the synthesis of PMF and GeogDetector models. Sci Total Environ 747. https://doi.org/10.1016/j.scitotenv.2020.141293

Ferri R, Donna F, R. Smith D et al (2012) Heavy metals in soil and salad in the proximity of historical ferroalloy emission. J Environ Prot (irvine, Calif) 03:374–385. https://doi.org/10.4236/jep.2012.35047

Franco-Uría A, López-Mateo C, Roca E, Fernández-Marcos ML (2009) Source identification of heavy metals in pastureland by multivariate analysis in NW Spain. J Hazard Mater 165:1008–1015. https://doi.org/10.1016/j.jhazmat.2008.10.118

Frank R, Ishida K, SUDA P (1976) Metals in agricultural soils of Ontario. Can J Soil Sci 56:181–196. https://doi.org/10.4141/cjss76-027

Friedlova M (2010) The influence of heavy metals on soil biological and chemical properties

Giller KE, Witter E, Mcgrath SP (1998) Toxicity of heavy metals to microorganisms and microbial processes in agricultural soils: a review. Soil Biol Biochem 30:1389–1414. https://doi.org/10.1016/S0038-0717(97)00270-8

Goncalves AC, Nacke H, Schwantes D, Coelho GF (2014) Heavy metal contamination in Brazilian agricultural soils due to application of fertilizers. Environ Risk Assess Soil Contam. https://doi.org/10.5772/57268

Goovaerts P (1997) Geostatistics for natural resources and evaluation

Gržetić I, Ghariani RHA (2008) Potential health risk assessment for soil heavy metal contamination in the central zone of Belgrade (Serbia). J Serb Chem Soc 73:923–934. https://doi.org/10.2298/JSC0809923G

Guagliardi et al (2013) A multivariate approach for anomaly separation of - Google Scholar. soil sediments

Gholizadeh MH, Melesse AM, Reddi L (2016) Water quality assessment and apportionment of pollution sources using APCS-MLR and PMF receptor modeling techniques in three major rivers of South Florida. Sci Total Environ 566–567:1552–1567. https://doi.org/10.1016/j.scitotenv.2016.06.046

Håkanson L (1980) An ecological risk index for aquatic pollution control-a sedimentological approach. Water Res 14:975–1001

Hettiarachchi GM, Pierzynski GM (2004) Soil lead bioavailability and in situ remediation of lead-contaminated soils: A review. Environ Prog 23(1):78–93

Holm A (1990) E. coli associated diarrhoea in weaner pigs: zinc oxide added to the feed a preservative measure. cabdirect.org

Hou D, He J, Lü C et al (2013) Distribution characteristics and potential ecological risk assessment of heavy metals (Cu, Pb, Zn, Cd) in water and sediments from Lake Dalinouer, China. Ecotoxicol Environ Saf 93:135–144. https://doi.org/10.1016/j.ecoenv.2013.03.012

Hu X, Zhang Y, Ding Z et al (2012) Bioaccessibility and health risk of arsenic and heavy metals (Cd Co, Cr, Cu, Ni, Pb, Zn and Mn) in TSP and PM2.5 in Nanjing. China Atmos Environ 57:146–152. https://doi.org/10.1016/j.atmosenv.2012.04.056

Jia Z, Li S, Wang, L (2018) Assessment of soil heavy metals for eco-environment and human health in a rapidly urbanization area of the upper Yangtze Basin. Sci Rep 8. https://doi.org/10.1038/s41598-018-21569-6

John K, Abraham II, Kebonye NM et al (2021) Soil organic carbon prediction with terrain derivatives using geostatistics and sequential Gaussian simulation. J Saudi Soc Agric Sci. https://doi.org/10.1016/j.jssas.2021.04.005

Jollife IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philos Trans R Soc A Math Phys Eng Sci 374. https://doi.org/10.1098/rsta.2015.0202

Jones BG, Alyazichi YM, Low C et al (2019) Distribution and sources of trace element pollutants in the sediments of the industrialised Port Kembla Harbour, New South Wales Australia. Environ Earth Sci 78. https://doi.org/10.1007/s12665-019-8358-1

Kabata-Pendias A (2011) Trace Elements in Soils and Plants. 4th edn. CRC Press. Taylor and Francis Group. ISBN: 978-1-4200-9368-1

Kachenko AG, Singh B (2006) Heavy metals contamination in vegetables grown in urban and metal smelter contaminated sites in Australia. Water Air Soil Pollut 169(1):101–123

Kampa M, Castanas E (2008) Human health effects of air pollution. Environ Pollut 151:362–367. https://doi.org/10.1016/j.envpol.2007.06.012

Karim Z, Qureshi BA (2014) Health risk assessment of heavy metals in urban soil of Karachi, Pakistan. Hum Ecol Risk Assess 20:658–667. https://doi.org/10.1080/10807039.2013.791535

Karimi Nezhad MT, Tabatabaii SM, Gholami A (2015) Geochemical assessment of steel smelter-impacted urban soils, Ahvaz. Iran J Geochemical Explor 152:91–109. https://doi.org/10.1016/j.gexplo.2015.02.005

Kowalska JB, Mazurek R, Gąsiorek M, Zaleski T (2018) Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination–a review. Environ Geochem Health 40:2395–2420

Kozák J (2010) Soil Atlas of the Czech Republic. 150

Kusin FM, Azani NNM, Hasan SNMS, Sulong NA (2018) Distribution of heavy metals and metalloid in surface sediments of heavily-mined area for bauxite ore in Pengerang, Malaysia and associated risk assessment. CATENA 165:454–464. https://doi.org/10.1016/j.catena.2018.02.029

Lee DH, Kim JH, Mendoza JA et al (2016) Characterization and source identification of pollutants in runoff from a mixed land use watershed using ordination analyses. Environ Sci Pollut Res 23:9774–9790. https://doi.org/10.1007/s11356-016-6155-x

Levy DB, Barbarick KA, Siemer EG, Sommers LE (1992) Distribution and partitioning of trace metals in contaminated soils near Leadville, Colorado. J Environ Qual 21:185–195. https://doi.org/10.2134/jeq1992.00472425002100020006x

Li J, He M, Han W, Gu Y (2009) Analysis and assessment on heavy metal sources in the coastal soils developed from alluvial deposits using multivariate statistical methods. J Hazard Mater 164:976–981. https://doi.org/10.1016/j.jhazmat.2008.08.112

Li L, Lu J, Wang S et al (2016) Methods for estimating leaf nitrogen concentration of winter oilseed rape (Brassica napus L.) using in situ leaf spectroscopy. Ind Crops Prod 91:194–204. https://doi.org/10.1016/j.indcrop.2016.07.008

Li X, Poon CS, Liu PS (2001) Heavy metal contamination of urban soils and street dusts in Hong Kong. Appl Geochemistry 16:1361–1368. https://doi.org/10.1016/S0883-2927(01)00045-2

Lim HS, Lee JS, Chon HT, Sager M (2008) Heavy metal contamination and health risk assessment in the vicinity of the abandoned Songcheon Au-Ag mine in Korea. J Geochemical Explor 96:223–230. https://doi.org/10.1016/j.gexplo.2007.04.008

Liu Y, Wang H, Li X, Li J (2015) Heavy metal contamination of agricultural soils in Taiyuan, China. Pedosphere 25:901–909. https://doi.org/10.1016/S1002-0160(15)30070-9

Luo L, Ma Y, Zhang S et al (2009) An inventory of trace element inputs to agricultural soils in China. J Environ Manage 90:2524–2530. https://doi.org/10.1016/j.jenvman.2009.01.011

Luo XS, Ding J, Xu B et al (2012) Incorporating bioaccessibility into human health risk assessments of heavy metals in urban park soils. Sci Total Environ 424:88–96. https://doi.org/10.1016/j.scitotenv.2012.02.053

Madrigal J, Persky V, Pappalardo A, Argos M (2018) Association of heavy metals with measures of pulmonary function in youth: findings from the 2011–2012 National Health and Nutrition Examination Survey (NHANES) ISEE Conf Abstr 2018. https://doi.org/10.1289/isesisee.2018.o03.03.26

Mamat Z, Yimit H, Ji RZA, Eziz M (2014) Source identification and hazardous risk delineation of heavy metal contamination in Yanqi basin, northwest China. Sci Total Environ 493:1098–1111. https://doi.org/10.1016/j.scitotenv.2014.03.087

Mamut A, Eziz M, Mohammad A, Anayit M (2017) Human and ecological risk assessment: An International Journal The spatial distribution, contamination, and ecological risk assessment of heavy metals of farmland soils in Karashahar-Baghrash oasis, northwest China. Int J 23:1300–1314. https://doi.org/10.1080/10807039.2017.1305263

Manta DS, Angelone M, Bellanca A et al (2002) Heavy metals in urban soils: a case study from the city of Palermo (Sicily), Italy. Sci Total Environ 300:229–243. https://doi.org/10.1016/S0048-9697(02)00273-5

Mantovi P, Bonazzi G, Maestri E, Marmiroli N (2003) Accumulation of copper and zinc from liquid manure in agricultural soils and crop plants

Maria Figueiredo AG, Tocchini M, dos Santos S, TF (2011) Metals in playground soils of SÃ£o Paulo city, Brazil. Procedia Environ Sci 4:303–309. https://doi.org/10.1016/j.proenv.2011.03.035

McBride MB, Shayler HA, Spliethoff HM, Mitchell RG, Marquez-Bravo LG, Ferenz GS, Bachman S (2014) Concentrations of lead, cadmium and barium in urban garden-grown vegetables: the impact of soil variables. Environ Pollut 194:254–261

Micó C, Recatalá L, Peris M, Sánchez J (2006) Assessing heavy metal sources in agricultural soils of an European Mediterranean area by multivariate analysis. Chemosphere 65:863–872. https://doi.org/10.1016/j.chemosphere.2006.03.016

Nicholson FA, Smith SR, Alloway BJ et al (2003) An inventory of heavy metals inputs to agricultural soils in England and Wales. Sci Total Environ 311:205–219. https://doi.org/10.1016/S0048-9697(03)00139-6

Pérez-De-Mora A, Burgos P, Madejón E, Cabrera F, Jaeckel P, Schloter M (2006) Microbial community structure and function in a soil contaminated by heavy metals: Effects of plant growth and different amendments. Soil Biol Biochem 38:327–341. https://doi.org/10.1016/j.soilbio.2005.05.010

Qingjie G, Jun D, Yunchuan X et al (2008) Calculating pollution indices by heavy metals in ecological geochemistry assessment and a case study in parks of Beijing. J China Univ Geosci 19:230–241. https://doi.org/10.1016/S1002-0705(08)60042-4

Qu MK, Li WD, Zhang CR, et al (2013) Source apportionment of heavy metals in soils using multivariate Statistics and Geostatistics

Rinklebe J, Antoniadis V, Shaheen SM et al (2019) Health risk assessment of potentially toxic elements in soils along the Central Elbe River, Germany. Environ Int 126:76–88. https://doi.org/10.1016/j.envint.2019.02.011

Rodríguez JA, Nanos N, Grau JM et al (2008) Multiscale analysis of heavy metal contents in Spanish agricultural topsoils. Chemosphere 70:1085–1096. https://doi.org/10.1016/j.chemosphere.2007.07.056

Rosen GD, Roberts PA (1996) Comprehensive survey of the response of growing pigs to supplementary copper in feed

Saha R, Nandi R, Saha B (2011) Sources and toxicity of hexavalent chromium. J Coord Chem 64:1782–1806

Salim I, Sajjad RU, Paule-Mercado MC et al (2019) Comparison of two receptor models PCA-MLR and PMF for source identification and apportionment of pollution carried by runoff from catchment and sub-watershed areas with mixed land cover in South Korea. Sci Total Environ 663:764–775. https://doi.org/10.1016/j.scitotenv.2019.01.377

Salonen VP, Korkka-Niemi K (2007) Influence of parent sediments on the concentration of heavy metals in urban and suburban soils in Turku, Finland. Appl Geochemistry 22:906–918. https://doi.org/10.1016/j.apgeochem.2007.02.003

Shiowatana J, McLaren RG, Chanmekha N, Samphao A (2001) Fractionation of arsenic in soil by a continuous-flow sequential extraction method. J Environ Qual 30:1940–1949. https://doi.org/10.2134/jeq2001.1940

Šmejkalová M, Mikanová O, Borůvka L (2003) Effects of heavy metal concentrations on biological activity of soil micro-organisms

Solgi E, Abbas ES, Alireza RB, Hadipour M (2012) Soil contamination of metals in the three industrial estates, Arak, Iran. Bull Environ Contam Toxicol 88:634–638. https://doi.org/10.1007/s00128-012-0553-7

Tepanosyan G, Sahakyan L, Belyaeva O, Saghatelyan A (2016) Origin identification and potential ecological risk assessment of potentially toxic inorganic elements in the topsoil of the city of Yerevan, Armenia. J Geochemical Explor 167:1–11. https://doi.org/10.1016/j.gexplo.2016.04.006

U.S.Environmental Protection Agency (2014) Positive matrix factorization (PMF) 5.0-Fundamentals and User Guide

US EPA (1989) Risk assessment guidance for superfund. human health evaluation manual part A, Interim Final,. United States Environ Prot Agency 1 part A:300

USEPA (1996) USEPA, Report: recent Developments for In Situ Treatment - Google Scholar. https://scholar.google.co.uk/scholar?hl=en&as_sdt=0%2C5&q=USEPA%2C+Report%3A+recent+Developments+for+In+Situ+Treatment+of+Metals+contaminated+Soils%2C+U.S.+Environmental+Protection+Agency%2C+Office+of+Solid+Waste+and+Emergency+Response%2C+1996.&btnG=. Accessed 6 May 2021

USEPA (2002) Supplemental guidance for developing soil screening levels for superfund sites, appendix D-dispersion factors calculations. USA. United States Environ Prot Agency, Washington, DC pp. 4–24

Vacek O, Vašát R, Borůvka L (2020) Quantifying the pedodiversity-elevation relations. Geoderma 373:114441. https://doi.org/10.1016/j.geoderma.2020.114441

Varol M, Sünbül MR, Aytop H, Yılmaz CH (2020) Environmental, ecological and health risks of trace elements, and their sources in soils of Harran Plain Turkey. Chemosphere 245. https://doi.org/10.1016/j.chemosphere.2019.125592

Veit MT, Da Silva EA, Fagundes-Klen MR et al (2009) Biosorption of nickel and chromium from a galvanization effluent using seaweed pre-treated on a fixed-bed column. Acta Sci - Technol 31:175–183. https://doi.org/10.4025/actascitechnol.v31i2.864

Wang C, Yang Z, Zhong C, Ji J (2016) Temporal-spatial variation and source apportionment of soil heavy metals in the representative river-alluviation depositional system. Environ Pollut 216:18–26. https://doi.org/10.1016/j.envpol.2016.05.037

Wang G, Liu HQ, Gong Y et al (2017) Risk assessment of metals in urban soils from a typical Industrial city, Suzhou, Eastern China. Int J Environ Res Public Health 14. https://doi.org/10.3390/ijerph14091025

Wang Z, Chai L, Yang Z et al (2010) Identifying sources and assessing potential risk of heavy metals in soils from direct exposure to children in a mine-impacted city, Changsha, China. Wiley Online Libr 39:1616–1623. https://doi.org/10.2134/jeq2010.0007

Weather Spark (2016) Average Weather in Frýdek-Místek, Czechia, Year Round - Weather Spark. https://weatherspark.com/y/83671/Average-Weather-in-Frýdek-Místek-Czechia-Year-Round. Accessed 14 Sept 2020

Wei B, Yang L (2010) A review of heavy metal contaminations in urban soils, urban road dusts and agricultural soils from China. Microchem J 94:99–107. https://doi.org/10.1016/j.microc.2009.09.014

Weindorf DC, Paulette L, Man T (2013) In-situ assessment of metal contamination via portable X-ray fluorescence spectroscopy: Zlatna, Romania. Environ Pollut 182:92–100. https://doi.org/10.1016/j.envpol.2013.07.008

Wu J, Lu J, Li L et al (2018) Pollution, ecological-health risks, and sources of heavy metals in soil of the northeastern Qinghai-Tibet Plateau. Chemosphere 201:234–242. https://doi.org/10.1016/j.chemosphere.2018.02.122

Wu J, Margenot AJ, Wei X et al (2020) Source apportionment of soil heavy metals in fluvial islands, Anhui section of the lower Yangtze River: comparison of APCS–MLR and PMF. J Soils Sediments 20:3380–3393. https://doi.org/10.1007/s11368-020-02639-7

Wuana RA, Okieimen FE (2011) Heavy metals in contaminated soils: a review of sources, chemistry, risks and best available strategies for remediation. ISRN Ecol 2011:1–20. https://doi.org/10.5402/2011/402647

Xiong X, Yanxia L, Wei L et al (2010) Copper content in animal manures and potential risk of soil copper pollution with animal manure use in agriculture. Resour Conserv Recycl 54:985–990. https://doi.org/10.1016/j.resconrec.2010.02.005

Xu X, Zhao Y, Zhao X et al (2014) Sources of heavy metal pollution in agricultural soils of a rapidly industrializing area in the Yangtze Delta of China. Ecotoxicol Environ Saf 108:161–167. https://doi.org/10.1016/j.ecoenv.2014.07.001

Yang Q, Li Z, Lu X et al (2018) A review of soil heavy metal pollution from industrial and agricultural regions in China: pollution and risk assessment. Sci Total Environ 642:690–700. https://doi.org/10.1016/j.scitotenv.2018.06.068

Zhang H, Li H, Yu H, Cheng S (2020) Water quality assessment and pollution source apportionment using multi-statistic and APCS-MLR modeling techniques in Min River Basin, China. Environ Sci Pollut Res 27:41987–42000. https://doi.org/10.1007/s11356-020-10219-y

Zhang J, Li R, Zhang X et al (2019) Vehicular contribution of PAHs in size dependent road dust: a source apportionment by PCA-MLR, PMF, and Unmix receptor models. Sci Total Environ 649:1314–1322. https://doi.org/10.1016/j.scitotenv.2018.08.410

Zhang X, Zhong T, Liu L et al (2016) Chromium occurrences in arable soil and its influence on food production in China. Environ Earth Sci 75:1–8. https://doi.org/10.1007/s12665-015-5078-z

Authors and Affiliations

Prince Chapman Agyeman[1] ⬤ **· Samuel Kudjo Ahado[1] · Kingsley John[1] · Ndiye Michael Kebonye[1] · Radim Vašát[1] · Luboš Borůvka[1] · Martin Kočárek[1] · Karel Němeček[1]**

[1]   Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, 16500 Prague, Czech Republic