

Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review



A. Horta^{*}, B. Malone, U. Stockmann, B. Minasny, T.F.A. Bishop, A.B. McBratney, R. Pallasser, L. Pozza

Department of Environmental Sciences, The University of Sydney, New South Wales 2006, Australia

ARTICLE INFO

Article history:

Received 29 August 2014

Received in revised form 20 November 2014

Accepted 25 November 2014

Available online 3 December 2014

Keywords:

Proximal sensing

Spatial statistics

Field spectroscopy

Soil contamination

ABSTRACT

There are tens of millions of contaminated soil sites in the world, and with an increasing population and associated risk there is a growing pressure to remediate them. A barrier to remediation is the lack of cost-effective approaches to assessment. Soil contaminants include a wide range of natural and synthetic metallic and organic compounds and minerals thus making analytical costs potentially very large. Further, soil contaminants show a large degree of spatial variation which increases the burden on sampling costs. This paper reviews potentially cost-effective methods for measurement, sampling design, and assessment. Current tiered investigation approaches and sampling strategies can be improved by using new technologies such as proximal sensing. Design of sampling can be aided by on-the-go proximal soil sensing; and expedited by subsequent adaptive spatially optimal sampling and prediction procedures enabled by field spectroscopic methods and advanced geostatistics. Field deployment of portable Visible & Near Infrared [wavelength 400–2500 nm] (Vis-NIR) and X-ray fluorescence (PXRF) spectroscopies will require special calibration approaches but show huge potential for synergistic use. The use of mid-infrared spectroscopy [wavelength 2500–25,000 nm, wavenumber 4000–400 cm^{-1}] (MIR) for field implementation requires further adaptive research. We propose an integrated field-deployable methodology as a basis for further developments.

Crown Copyright © 2014 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Contents

1.	Introduction	181
2.	Soil contamination assessment	183
2.1.	Soil contaminants and sources of contamination	183
2.2.	Contaminated site characterisation – current approach	183
2.3.	Sampling for contamination assessment	183
3.	Infrared spectroscopy for measuring soil contaminants	184
3.1.	Heavy metals	185
3.1.1.	Studies on contaminated sites	186
3.1.2.	Geochemical studies	186
3.2.	Organic contaminants	186
3.3.	Other applications	190
4.	PXRF for measuring soil contaminants	191
5.	Taking IR and PXRF to the field and the role of spectral libraries	194
6.	Spatial analysis for sampling, predicting and mapping contamination	195
7.	Prospective methodology for soil contamination assessment supported by field spectroscopy and advanced spatial analysis	202
7.1.	The synergistic use of Vis-IR and PXRF – data fusion	202
7.2.	In-field soil inference engine	203
7.3.	Spatial analysis – adaptive sampling and mapping	204

^{*} Corresponding author at: Faculty of Agriculture and Environment, Biomedical Building, 1 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015, Australia.
E-mail address: ana.seixashorta@sydney.edu.au (A. Horta).

8. Conclusions	205
Acknowledgments	205
References	205

1. Introduction

Mining, urbanisation, and agricultural and industrial processes have resulted in the contamination on the order of 10^7 sites¹ across the globe with contaminants such as heavy metals and organics. The European Union has 342,000 contaminated and 2.5 million potentially contaminated sites (Panagos et al., 2013). One of the leading European countries for developing legislation for soil quality assessment and protection, The Netherlands, has 265,000 sites potentially contaminated; from those, 11,000 sites are actually in need of urgent remedial action (Ministry of Housing, Spatial Planning and the Environment, 2010). Typical soil contaminants in Europe include Total Petroleum Hydrocarbons (TPH) and heavy metals which contribute to 60% of soil contamination. About 160,000 contaminated sites potentially exist across Australia (State of the Environment Committee, 2011). Sites impacted by TPH represent a significant proportion of Australia's contaminated land (Clements et al., 2009). In the United States, 1200 sites are on the National Priority List (NPL) for the treatment of contaminated soils, indicating the extensiveness of this problem (Mulligan et al., 2001). Approximately 63% of the sites on the NPL include contamination from toxic heavy metals.

Given the pressure on soil for food security and growing urbanisation, the identification and remediation of contaminated sites is of increasing importance (Liu et al., 2013; Chen, 2007; Luo et al., 2012; Cai et al., 2008). Many of these contaminated sites are now becoming attractive as high value commercial and residential land but health and environmental risks are at stake. However, remediation rates are quite modest, e.g. in 33 European countries where 127,000 sites have been recognised as contaminated and 1.17 million are potentially contaminated, only 58,000 (5%) have undergone some kind of remediation (Panagos et al., 2013). In Australia, the Cooperative Research Centre for Contamination Assessment and Remediation of the Environment estimates that Australian companies only clean up roughly 1000 sites each year, i.e. 0.5% per annum, which may be not much better than maintaining a steady-state situation in terms of overall numbers, i.e. cleaning up sites as fast as newly contaminated sites are created and/or identified. Estimated clean-up costs are of the order of \$2 billion per year, with the total remediation cost being much larger (State of the Environment Committee, 2011). Additional sites continue to be identified, as contamination is often not apparent until a site is prepared for sale or redevelopment or the land use changes (NSW EPA, 2013).

Worldwide, national and other jurisdictional protocols have been devised for contaminated site assessment, establishing decision-support systems to evaluate the need for remediation. Most legal frameworks propose decision models based on risks to human health, the soil ecosystem and food safety. Risk evaluation is site-specific and dependent on future land management. Regardless of the soil protection target, decisions are supported by a tiered site characterisation whereby a site undergoes a preliminary investigation followed by more intensive investigations in stages. This approach allows data from each stage to be assessed and fed into planning the next stage of investigation. In each progressive tier, the assessment becomes less conservative, is based

on more site-specific information and, hence, is more complex, time-consuming and often more expensive (Clements et al., 2009; Ministry of Housing, Spatial Planning and the Environment, 2010).

Remediation processes are expensive and rely on estimates of the amount of contaminated soil which needs to be removed and the type of contaminant (Mulligan et al., 2001; Lewandowski et al., 2006; Schultz, 1997). Therefore, accurate estimates of the spatial distribution of contaminants are essential (Markus and McBratney, 2001; Motelay-Massei et al., 2004; Imperato et al., 2003). In practice, the evaluation of the extent and source of site contamination requires soil sampling and further laboratory analysis to gather information about the type and degree of contamination (Cattle et al., 2002). Sampling at an individual location along with associated laboratory analytical costs can be as much as 1000 AUD (depending on the type of contamination), and some replications of samples may be needed to define the contamination, and even then this may not be entirely accurate. In Europe, costs for site investigations generally are between €5000 to €50,000 and costs for remediation projects usually fall in the range of €50000 to €500,000 (Van Liedekerke et al., 2014).

Legislation in The Netherlands, Australia and the United States advocates where possible the use of formal probabilistic sampling schemes (e.g. simple random, stratified random) for contaminated site assessments. These types of sampling design support an unbiased decision about whether contamination levels exceed a threshold of unacceptable risk which helps to identify the location of "hot spots" or plume delineation and to characterise the nature and extent of contamination at a site (US EPA, 2002). Often choosing the most resource-effective design is a trade-off between performance and cost that accounts for practical issues such as schedule and budget risks and health and safety risks. It is important that the necessary data are available to enable a statistical analysis to inform the decision of whether or not the site presents an unacceptable risk to human health or the environment, and is suitable for the intended future land use.

Contamination thresholds exist for a wide range of pollutants (heavy metals and organics) (Ferguson, 1999; Regan et al., 2002) and these determine whether the location from which it emanates requires remediation. This location can refer to an established volume of soil or a determined exposure area depending on the criteria defined in the decision model. Normally, decisions are made based on the mean or maximum concentration measured for a certain volume or area of contaminated soil. In the case of the mean, a 95% confidence interval is computed to determine if this concentration exceeds or is below the threshold. The more samples taken the more precise will be the estimates of the mean. These practical requirements make soil contamination assessments expensive, time-consuming and, as a result, often based on inadequate field data.

However, it is necessary to generalise the results to unsampled locations and to define homogeneous contamination zones (along with an estimate of the associated uncertainty) across the site being assessed (Gilbert, 1987; Bierkens, 1997). Geostatistical techniques can be used in a geographic information system (GIS) to provide information on the spatial distribution of contamination (Burrough, 2001).

In summary, there is room to optimise current decision-based models for soil contamination assessment namely by improving the quality and quantity of information provided in the preliminary investigation stages. The current situation with respect to measurement and delineation motivates a range of questions which will be addressed in this review.

There is a way forward; increasing advances in electronics, information technology, and spatial statistics can contribute significantly.

¹ An estimate may be obtained as follows: if we assume that developed economies have 0.005 contaminated and potentially contaminated sites per capita [Panagos et al., 2013 report 0.00502 sites per capita for the European Union] and a population of 1 billion, emerging economies have 0.0025 contaminated sites and potentially contaminated per capita and a population of 2 billion, developing economies have 0.001 contaminated sites and potentially contaminated per capita and a population of 4 billion, we get a total of 14 M. Given uncertainty this is probably between 10 M and 20 M.

Contextually, proximal soil sensing, improved sampling design aided by portable field spectrometers and cheaper laboratory measurements combine to suggest a new approach. This is already corroborated by the US EPA (<http://www.clu-in.org/characterization/>) and the Australian legislation for the assessment of site contamination (**National Environment Protection (Assessment of Site Contamination) Measure, as amended in 2013**). These guidance documents suggest that “real-time” sampling, field analytical methods and on-site interpretation and iteration of field data can expedite site characterisation in terms of the concentration and distribution of contaminants. Changing the importance of on-site analysis tools, from a screening-level stage to actually influence the decision-making process, was advocated by **Crumbling (2001)** in a document produced for the US EPA about the use of a “triad approach to improve the cost-effectiveness of hazardous waste site clean-ups”. This document aimed to set a framework for using an integrated triad of systematic planning where real-time measurement technologies were essential to plan and implement data collection and technical decision-making. According to **Crumbling (2001)**, this “modernization” of site characterisation offers cost savings of about 50%. Besides the budget issue, on-site analytical tools increase sampling density thus generating datasets more representative of site conditions (which is crucial since the largest source of uncertainty in contaminated sites is linked with the degree of contamination heterogeneity). Hence, when combined with sampling and laboratory analysis the potential to produce more accurate information increases enormously therefore improving decision-making (**Clements et al., 2009**). The concerns of formally adopting on-site analysis tools as part of the decision-making models are still related to lack of accuracy and reliability of the available instruments and techniques when compared with the analytical tests conducted in a laboratory.

Proximal soil sensing refers to all field methods that sense the soil from outside – i.e., where the probe does not enter into the soil volume

it is measuring (**McBratney et al., 2011a; Viscarra Rossel et al., 2011**). It includes ground-penetrating radar (GPR), electromagnetic induction (EMI) and gamma radiometrics (GR) as well as field spectroscopic methods such as visible-infrared (Vis-IR) and portable X-ray fluorescence (PXRF). All proximal soil sensing methods have a part to play in soil contamination assessment – although not generally contamination specific, the on-the-go or potentially mobile methods of GPR, EMI and GR are most useful for quick site characterisation to provide prior information for initial sampling design. They are generally not predictive of contaminant concentrations. The field spectroscopic methods such as Vis-IR and PXRF are targeted at measurements of contamination concentrations, and therefore are suitable for use in the adaptive sampling and mapping of contamination. They have the added advantage of being easy to operate and are non-hazardous in the work environment (**He et al., 2007**). Vis-IR and PXRF predictions are generally based on calibration equations for a large range of potential contaminants which require well-designed and populated spectral libraries from a large suite of contaminated and non-contaminated sites (**Shepherd and Walsh, 2002**).

The current approach for contaminated site characterisation can be generalised as the workflow presented on the left side of **Fig. 1**. We will outline this briefly in **Section 2** below. In recent times, there has been a plethora of new research that gives us techniques that run parallel with, or can be combined with, the conventional approaches. As mentioned above, spectroscopic techniques which are field deployable offer exciting possibilities; these are reviewed in **Sections 3, 4 and 5**. Much has been done in the past decade on understanding the spatial distributions of contaminants and creating new sampling and prediction methods. This work is reviewed in **Section 6**. A combination of approaches on the right-hand side of **Fig. 1** can be regarded as a potentially new approach – this proposed new methodology is outlined in **Section 7**.

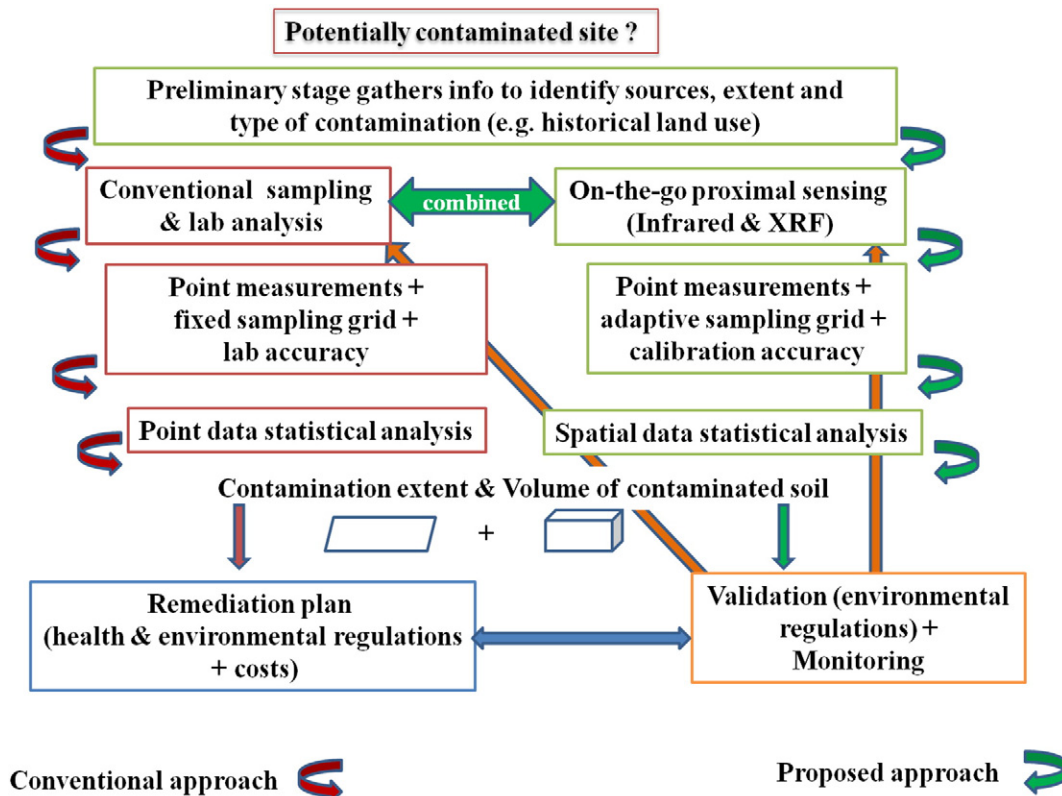


Fig. 1. Contamination assessment workflow: current and prospective approaches.

2. Soil contamination assessment

2.1. Soil contaminants and sources of contamination

Soil acts as a sink for almost all substances released into the environment by human activities: many pollutants accumulate in the soil due to its specific filtering and buffering properties. If the concentration of these substances is above a defined background value or so high that it potentially causes a risk to human health, plants, animals, ecosystems or other media (e.g. water), the soil is regarded as “contaminated” (EC, 2005). Soil contamination may have important consequences in terms of the soil's ability to function. Soils may fail to support vegetation and biomass production, may fail to provide valuable materials and substrate to human activities, ecological systems and biological cycling of nutrients or may be unable to act as filter and buffer, affecting the hydrosphere, compromising groundwater resources and threatening aquatic ecosystems (Rodrigues et al., 2009).

Major pollutants include organic contaminants such as chlorinated hydrocarbons, TPH and heavy metals (Alloway, 1995). If the natural soil functions of buffering, filtering and transforming are overexploited a variety of negative environmental impacts arise, the most problematic of which are pollution of water, acute or chronic toxicity and uptake of contaminants by plants.

Moreover soil may act as potentially medium of contamination exposure therefore becoming a major threat to human health. Although current research has not yet established a clear cause–effect between soil contamination and adverse human health outcomes there is an increasing body of evidence that this relationship exists (Brevik and Burgess, 2012). As a precautionary approach legislation for contamination assessment was developed taking health guidelines into consideration.

Contamination can be diffuse or localized (van der Perk, 2006). Diffuse soil contamination is the presence of a substance in the soil as a result of human activity emitted from moving sources, from sources with a large area, or from many sources. It can be caused by dispersed sources and occurs where emission, transformation and dilution of contaminants in other media has occurred prior to their transfer to soil. As a result, the relationship between the contaminant source and the level and spatial extent of soil contamination is indistinct. This is generally associated with atmospheric deposition, certain agricultural practices, inadequate waste and wastewater recycling and treatment or even contamination due to natural hazards such as floods and landslides. Atmospheric deposition of anthropogenic contaminants (including nutrients and acid deposition) are due to emissions from industry, transport, households and agriculture. Certain agricultural practices cause diffuse soil contamination by direct application of pesticides, sewage sludge, compost, fertilizers and manure. Currently, the most important problems from diffuse sources are acidification, the effects of a surplus of nutrients and contamination by heavy metals (State of the Environment Committee, 2011). Local soil contamination occurs where intensive industrial activities, inadequate waste disposal, mining, military activities or accidents introduce excessive amounts of contaminants. At industrial plants that are still operating, soil contamination may have its origin in the past but current activities still have significant impacts (EC, 2005).

2.2. Contaminated site characterisation – current approach

Knowing the type and source of contaminants is very useful for contamination management and it will likely influence the time and budget needed to establish the contamination extent and therefore the remediation which is adequate to minimise health and environmental risks. The main concern in developing legal and regulatory guidance for soil contamination assessment is to ensure that the quality and the quantity of data available is enough to support decision-makers in the process of determining if a site is contaminated or not and, if so, the risks it presents to human health, the soil ecosystem, and food safety. This risk evaluation is usually site-specific and dependent on the future land

use. Also, decision-making requires statistical evidence that the site does not present unacceptable risks and is suitable for the intended land use.

These issues are acknowledged by current legislation in practice namely in countries such as The Netherlands, Australia and the United States. These countries present similar guidance for establishing a contamination-assessment framework including best practices for sampling and data analysis. In the context of this review we will focus on these two issues using the information available² to describe the current approach for contaminated sites characterisation.

Standard practice for contaminated-site characterisation comprises a tiered approach in which a site undergoes a preliminary investigation followed by more intensive investigations in stages. The purpose of the preliminary investigation is to establish whether the site might be contaminated. This stage is essential to gather information on the sources of contamination, where it is located and what type of contaminants could be expected. Usually this can be inferred from the past and present land uses, previous reports of incidents and/or environmental assessments as well as the site's geological and hydrological situation.

Currently, soil sampling is not expected in the preliminary investigation. It should occur in a second investigation stage based on the information collected. The quality and the quantity of these data will determine the cost-efficiency of the sampling design therefore it is recognised that a sufficient amount of effort (and thus money) should be spent on the preliminary investigation.

As part of the decision-making process, a rule is established to determine if the site is contaminated, and to what extent it requires remediation. This decision rule is based on a mean or maximum concentration found for a specific contaminant. If the mean or maximum concentration exceeds (with a certain level of confidence) a specified threshold (known variously as Investigation level in The Netherlands, Health Investigation level in Australia and Soil Screening level in the United States) then further investigation is needed or a remediation plan is required. To obtain a best estimate of the contaminant's concentration, a minimum number of samples are required depending on the expected contamination variability across the site and the spatial scale of the decision-based support. Support refers to the measure (length, area or volume) usually associated with the physical representation of the data values. Commonly support refers to “points” that represent a location in space, i.e. the sampling location. However, for contamination assessment, this support can be specified as an exposure area for risk evaluation or as a volume of soil to be removed or remediated. The mean concentration estimated for the decision-based support must be representative regardless of the contamination variability. Therefore, it is all about designing a sampling plan that will ensure collecting representative, defensible and accurate data within the available time and budget.

2.3. Sampling for contamination assessment

The general purpose of a sampling design for contamination assessment is to support a decision about whether contamination levels exceed a threshold of unacceptable risk and, if so, what is the nature and extent of the contamination. Samples should be collected in an unbiased manner, i.e. all locations within the sampling area should have a pre-defined probability of being selected as a sampling point. Known

² The Netherlands: Soil Quality Decree 1999 (complemented by the Circular on Soil remediation 2009); NEN 5740 (2009).

Australia: National Environment Protection (Assessment of Site Contamination) Amendment Measure 2013 (Schedules B1 and B2); Standards Australia AS4482.1 (2005); NSW OEH Contaminated Sites: Guidelines for Consultants Reporting on Contaminated Sites (2011); NSW EPA Contaminated Sites: Sampling Design Guidelines (1995), EPA Victoria IWRG621 (2009); EPA South Australia Guidelines for the assessment and remediation of site contamination (2014).

United States: Guidance for Choosing a Sampling Design for Environmental Data Collection (EPA QA/G-5S) (2002); US EPA Risk characterisation handbook (2000).

constraints when deciding on a sampling design comprise sampling and analysis limitations, time restrictions, geographic barriers and budget. Examples of constraints include measurement instrument performance, weather conditions and the costs associated with collecting samples in the field or laboratory analysis.

Broadly there are two main categories of sampling designs: judgmental and probability-based (De Gruijter et al., 2006). The legal frameworks mentioned previously favour a probability-based design. Basically when using probabilistic sampling quantitative conclusions (statistical inference) can be drawn from the analytical results. Therefore the data analyst can provide to the decision-maker a 95% upper confidence level (UCL) of the mean contaminant's concentration in soil. The analyst will be able to state whether the data indicate that the concentration exceeds or is below the threshold with a certain level of confidence.

Within the category of probability-based sampling, it is possible to choose the design that is most suitable to site conditions and expected contamination heterogeneity. Commonly used sampling includes simple random, stratified, systematic grid, and adaptive cluster sampling (US EPA, 2002). The quality and quantity of the data collected during the preliminary investigation has a strong influence when choosing the sampling design. If the information is not accurate enough it is probably better to opt for systematic grid sampling (with a randomly selected location for the origin) to ensure complete site coverage and a homogeneous distribution of samples over the site. However it should be noted that the estimate of variance from such a design is over-estimated and so are the confidence intervals associated with the estimates of the mean (De Gruijter et al., 2006). On the other hand, if there is enough evidence to characterise the source and spatial variability of soil contamination, a stratified random sampling may be adequate. It combines both a pollution-oriented and a statistical random sampling. The site is first divided into several areas in which pollution is thought to be relatively homogeneous. Each area is then randomly sampled to obtain the mean concentration (Malherbe, 2002).

For sampling purposes and depending on the chosen design, the site can be divided into domains/areas or stockpiles representing similar material types, similar contamination, and other site-specific features as indicated by the land use history (NSW EPA, 1995). Hence, the minimum number of samples to collect can be assigned to a known area or volume. The US EPA recommends a minimum of 10 samples for a 95% UCL average calculation (US EPA, 2002). For large soil volumes (i.e. >2500 m³) the minimum sampling rate should not be less than 1 sample per 250 m³. In terms of the number of samples in function of the area it varies from 5 (0.05 ha) to 55 (5 ha) (EPA Victoria, 2009).

In summary,

- The preliminary investigation stage plays a crucial role in decision-making since this determines how the subsequent stages for contamination assessment should be conducted. If the information gathered in the preliminary investigation is not enough and/or not accurate more time, effort and budget need to be considered.
- Sampling is not generally included in the preliminary investigation. The sampling plan is designed after compiling the information available and it strongly depends on the quantity and quality of the data. Sample collection is necessary for laboratory analysis to determine the concentration of the contaminants.
- Depending on the sampling plan, the potentially contaminated site is divided into areas or stockpiles. This could depend on the spatial scale of the decision-based support. For each decision-based support, a minimum number of samples have to be collected to allow for the statistical analysis of the analytical results.
- Decision-making is based on the estimates for the mean or maximum concentration of the contaminants (measured for a certain volume or area of contaminated soil) and their 95% UCL. If the mean or maximum concentration exceeds a specified threshold then further investigation is needed or a remediation plan is required.

It is clear that improving the outputs of a preliminary investigation will benefit enormously from site characterisation and thus decision-making. Basically, it is all about providing more and accurate data to minimise lack of statistical representativeness, increase the spatial quantification, reduce the time needed to evaluate the presence and extent of contamination, add focus to the sampling plan and overall reduce the project costs. Real-time sampling, field analytical methods and on-site interpretation and replication of field data can expedite site characterisation in these terms (Crumbling, 2001; Clements et al., 2009).

In this context, proximal sensing namely field spectroscopic methods such as Vis-IR and PXRF can provide real-time measurements of contamination concentrations to support an on-the-go assessment. These IR and PXRF measurements can be a part of an adaptive sampling design whereby sampling efforts are directed to areas where the degree of contamination appears to be higher (Juang et al., 2005, 2008). This adaptive sampling strategy is somewhat similar to what was proposed by Englund and Heravi (1994). These authors analysed the benefits of phased sampling where the sampling process is interrupted until the data from prior sampling is available for interpretation. In two-phase sampling, preliminary estimates of contaminant concentrations based on data from the first phase are used to determine locations where additional data are needed most. When using field spectroscopy, prior data refers to IR and PXRF measurements and assuming that on-the-go software is available so that the period between sampling phases is reduced. The second phase could be the collection of samples to validate Vis-IR and PXRF measurements or to choose the locations where it is necessary to extract soil cores to verify contamination in the soil profile.

Instead of making decisions based on a single mean/maximum value for a specified support, a dense dataset as provided by Vis-IR and PXRF measurements can be easily integrated to provide a continuous map of contamination increasing the chances of detecting “hot-spots” (generally characterised as high contaminant concentrations that might be missed by a conventional sampling plan). Also, knowing the true concentration of the contaminant helps to select a better remediation technique. Advanced geostatistics and GIS tools can be used to develop a mapping framework based on on-the-go field measurements. Geostatistical mapping can bring out insufficiently sampled areas and/or high-risk areas based on the spatial continuity observed in the data (see Section 6). The uncertainties associated with the mapping output could also be quickly evaluated. Reducing the uncertainty of the result could be a matter of repeating field measurements and/or choosing specific locations to collect soil samples for validation.

3. Infrared spectroscopy for measuring soil contaminants

Infrared diffuse reflectance spectroscopy is based on the principle that molecules have specific frequencies at which they rotate or vibrate corresponding to discrete energy levels. Absorption spectra of compounds are a unique reflection of their molecular structure. Spectroscopy in both the visible, near (Vis-NIR, 400–700–2500 nm) and mid (MIR, 2500–25,000 nm) infrared ranges allows rapid acquisition of soil information. Spectral signatures of soil materials are characterised by their reflectance to particular wavelengths in the electromagnetic spectrum. Properties that are related to the surface area of the soil usually can be predicted well from Vis-NIR and MIR spectroscopy (Soriano-Disla et al., 2014; Stenberg et al., 2010; Viscarra Rossel et al., 2006).

Vis-NIR spectrometers are used extensively and gained popularity in soil science because they are available in a portable format, are easy and ready to use in the field and require minimal or even no sample preparation. With proper calibration, they can be used to estimate many soil properties, namely total C, total N, sand and clay content, cation exchange capacity, and pH (Soriano-Disla et al., 2014; Stenberg et al., 2010). Reviews on the principles and use of Vis-NIR for predicting soil properties can be found in Stenberg et al. (2010) and Soriano-Disla et al. (2014) thus will not be repeated here. Schwartz et al. (2011) summarised the use of Vis-NIR spectroscopy particularly for assessing

soil contamination, and Shi et al. (2014) gave a review specifically on Vis-NIR for predicting heavy metals. A review by Rathod et al. (2013) focussed on spectral sensing of soil and vegetation for monitoring phytoremediation of metal-contaminated soils, and Okparanma and Mouazen (2013a) specifically reviewed the use of Vis-NIR to predict Total Petroleum Hydrocarbon (TPH) concentration in soil.

Mid-infrared (MIR) spectroscopy usually produces better predictions than the Vis-NIR as it reflects the fundamental stretching and bending vibrations of molecules, which produces more spectral information and also results in more intense spectral peaks (Reeves, 2010; Soriano-Disla et al., 2014). However MIR generally requires more sample preparation as its accuracy and reproducibility is affected by sample homogeneity. Nevertheless a field portable MIR spectrometer has now become available (Sorak et al., 2012).

As soil is a complex mixture of materials it is difficult to assign specific features of the spectra to specific chemical components. In addition, due to the scattering effect, the useful part of the absorption spectra (related to elemental concentration) is relatively small when compared to the scattering effect (Gobrecht et al., 2014). Multivariate calibration techniques are used to calibrate the spectra to known measured soil properties. Digital soil spectra usually contain hundreds or thousands of reflectance values as a function of wavelength. As there are more predictor variables than the number of observations and predicted variables, methods that reduce the dimension of the spectra or techniques to select useful predictors are required. Principal component regression and the partial least squares (PLS) method are commonly utilised (Geladi and Kowalski, 1986). PLS extracts successive linear combinations of the predictors, which optimally address the combined goals of explaining response variation and explaining predictor variation. Other data-mining techniques that are capable of variable (wavelength) selection have also been found useful (Minasny and McBratney, 2008).

3.1. Heavy metals

Infrared spectroscopy can usually detect constituents or elements in the soil when the elements are present in amounts greater than the detection limit of spectroscopy and form vibrational bonds. Iron and other

cations such as Al and Mg have a distinct effect on the vibrations of OH bonds. As an example, Fe content in soils is linked with the depth and width of the absorption band between 750 and 1200 nm (Richter et al., 2009). Other elements or constituents in the soil, including the ones which have no absorbance, can be determined indirectly if they are present in a relatively high amount and correlated to constituents that absorb the infrared radiation.

While infrared studies are mostly used for the estimation of organic compounds, some inorganic compounds can produce an infrared signature. Metal carbonyl and organometallic compounds are commonly studied in infrared spectroscopy (Stuart, 2004). However, soil is a complex mixture of organic and inorganic materials, and the relatively low concentration of heavy metals usually does not show absorption features in the Vis-NIR or MIR spectrum. As an example, Fig. 2 shows a clay soil which has a relatively low concentration of carbon (0.7%) spiked with different amounts of lead (these results were obtained in an experiment conducted by the authors for this paper purpose). The absorbance spectra of the soil in the near infrared region are not affected even up to a very high Pb concentration of 16,000 mg kg⁻¹. The spectra only show visible features around 2270–2320 nm at a very high Pb concentration (>70,000 mg kg⁻¹). According to a study by Kemper and Sommer (2002) the infrared absorbance increasing with increasing Pb content appears to reflect the increasing content of organic matter in the soil.

The mechanism of detection of heavy metals via infrared spectroscopy is thus attributed to the association of heavy metals with organic matter, clay minerals, and Fe/Al oxides (Choe et al., 2009; Dupuy and Douay, 2001; Vohland et al., 2009; Wu et al., 2007). These indirect relationships between metal concentrations and soil constituents have previously been explored in pedotransfer functions (PTFs). For example, Chen et al. (2002), Elzinga et al. (1997) and Horn et al. (2006) presented various PTFs that predicted the adsorption of metal as a function of soil carbon content, CEC (cation exchange capacity) and pH.

Wu et al. (2007) investigated the spectral response by adding known amounts of heavy metals into an alluvial soil. The results showed that the spectral features for heavy metals only appear in the NIR spectrum for transitional elements (around 620 nm for Cr and 820 nm for

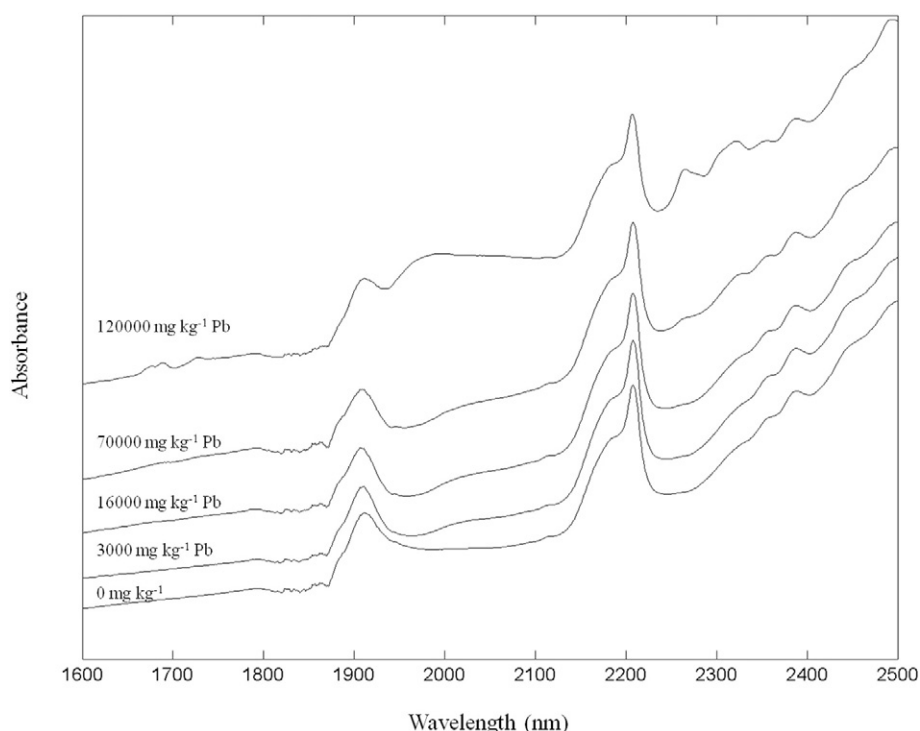


Fig. 2. NIR spectral reflectance of a clay soil (organic C concentration = 0.7%) with the addition of different amounts of Pb.

Cu), at very high concentrations ($>4000 \text{ mg kg}^{-1}$). Thus, [Kleinebecker et al. \(2013\)](#) working on aquatic sediments recommended that reliable NIR prediction of heavy metals is restricted to samples with high concentrations. [Choe et al. \(2008\)](#) proposed a binding mechanism based on the surface complexation model, where metals can bind to the hydroxylated mineral surfaces (such as Fe, Al, Si, or Mn) thus affecting the absorption peaks at around 500–1000 nm (Fe, Al, Mn oxides), and 2000 nm (clay minerals), or 400–600 nm (organic matter) ([Piccolo and Stevenson, 1982](#)). As a consequence, the applicability of the results from such studies is highly dataset-specific ([Middleton et al., 2011](#)).

[Goldberg and Johnston \(2001\)](#) studied As adsorption mechanisms on amorphous Al and Fe oxides using Raman and Fourier Transform Infrared (FTIR) spectroscopy. They showed that the position and intensity of vibrational bands of As species sorbed on amorphous Al and Fe oxides are pH dependent.

Many published studies have related total concentration of elements measured using laboratory XRF, or using acid or *aqua regia* extraction and correlated it with Vis-NIR spectra. These studies can be mainly grouped into the two categories presented in [Sections 3.1.1 and 3.1.2](#).

3.1.1. Studies on contaminated sites

Key studies include [Malley and Williams \(1997\)](#) who used NIR spectroscopy to estimate heavy metal contaminants in freshwater sediment in Canada (number of samples (n) = 169). They successfully predicted Cu, Zn, Pb, Ni, Mn, and Fe. Their PLS coefficients indicated that heavy metal concentration predictions were attributed to protein, cellulose, and oil. [Kooistra et al. \(2001\)](#) evaluated 69 soil samples in the floodplains of the river Rhine in The Netherlands and found that Vis-NIR spectra can be used to predict Cd and Zn concentration quite well. Its success is attributed to the proxy of organic matter and clay content. [Kemper and Sommer \(2002\)](#) worked in an area west of Seville in Spain which was contaminated with mine tailings. This study independently validated Vis-NIR predictions for heavy metals and found good prediction ($R^2 > 0.8$) for elements such as As, Fe, Hg, Pb, S, and Sb. They also indicated that the success of the prediction is attributed to absorption features of iron and iron oxides. It should be noted that the existing literature commonly uses R^2 to assess the prediction quality of models. In terms of prediction it is not a suitable statistic as it measures, the fit of data to a straight line (i.e. any straight line) when in fact we wish to know how well the data follows the 45° line when plotting predicted against observed values. Despite this we have chosen to use R^2 to present our review results since it is frequently shown in published papers. A better statistics to assess prediction accuracy would be Lin's concordance correlation coefficient which measures how well observations follow the 45° line.

Further studies on contaminated sites examined soils from a mining region ([Siebielec et al., 2004](#)) and urban areas ([Wu et al., 2005a](#)). [Wu et al. \(2005b\)](#) reported good prediction for soil Hg concentration in agricultural soils in China ($n = 120$) using Vis-NIR spectra, its success was attributed to the correlation with goethite and clay minerals. [Bray et al. \(2009\)](#) evaluated Vis-NIR and MIR as a diagnostic tool for Cd, Cu, Pb, and Zn using an ordinal logistic regression technique for prediction of the probability of exceeding certain thresholds. They found that Zn and Cu can be well predicted with prediction accuracy (presented in terms of false positives and false negatives) greater than 80%; meanwhile Cd and Pb are not that well predicted with accuracies less than 70%.

Recent studies also investigated low concentration of heavy metals in agricultural soils. For example, [Dong et al. \(2011\)](#) used Vis-NIR and MIR to predict concentration of As, Cu, Zn, Pb, and Cr ($n = 111$). They found that organic matter and pH were closely correlated with As and the heavy metals. [Shi et al. \(2014\)](#) provided a comprehensive review of studies that predicted heavy metal content using Vis-NIR spectroscopy. While some studies show agreements that NIR spectroscopy can predict concentration of Cr, Cu, Hg, Ni, Pb, and Zn, the results are not consistent. For example, [Kooistra et al. \(2001\)](#) found good predictions for Cd and Zn ($R^2 \sim 0.8$), contrarily [Kemper and Sommer \(2002\)](#) found poorer predictions for these elements ($R^2 = 0.2$ (Cd) and 0.5 (Zn)),

while [Siebielec et al. \(2004\)](#) found reasonable predictions ($R^2 = 0.5$ and 0.7). There are also contradictory conclusions on the major absorption feature mechanisms ([Schwartz et al., 2011](#)).

A limitation to most studies is that the analysis is based on a small number of calibration data: 38% of the studies cited in [Table 2 of Shi et al. \(2014\)](#) have less than 100 samples. Additionally, most of the studies do not have independent validation data. Thus the results can be quite site-specific and are unlikely to be generalised. While there is a good amount of work that indicates that Vis-NIR can be used effectively for predicting heavy metal concentration, there is still a question whether the calibration functions are site-specific or of general applicability. Some insights might be gained when looking at geochemical survey studies which usually have a large number of samples across a large and diverse area and with background concentration of metals that is lower than contaminated sites.

3.1.2. Geochemical studies

[Soriano-Disla et al. \(2013\)](#) analysed MIR spectra of 4130 soil samples from the GEMAS European soil sampling programme (geochemical mapping of agricultural soils and grazing land of Europe) where the samples were analysed for 45 elements using *aqua regia* extraction. They selected 1000 samples randomly to develop PLS models, and validated their predictions against the rest of the data. They found that the following elements can be well predicted ($R^2 > 0.6$): Ca, Mg, Al, Fe, Ga, Co, Ni, Sc, Ti, Li, Sr, K, Cr, Th, Be, S, B, Rb, V, Y, Zn, Zr, Nb, Ce, Cs, Na, In, Bi, Cu, and Mn. Meanwhile the following elements cannot be well predicted ($R^2 < 0.5$): As, Ba, La, Tl, P, U, Sb, Mo, Pb, Se, Cd, Sn, Hg, Ag and W. The results are similar to the findings of [Middleton et al. \(2011\)](#) who analysed samples from glacial till sites in Finland.

[Reeves and Smith \(2009\)](#) analysed samples collected from two transects (east–west across the USA, and north–south from Manitoba (Canada) to El Paso, Texas (USA), $n = 453$ and 267 , respectively). The results showed that, as expected, MIR predicts better than NIR, except for Co, Cr and Ni. The authors concluded that extreme diversity in the samples found in the continental-scale survey resulted in poor results even for measures of soil carbon content, which have been shown to be most accurately determined in most studies.

In summary,

- Many studies have shown that Vis-NIR and MIR can be used to predict heavy metal content in the soil. Infrared spectroscopy was usually able to predict total heavy metal concentration but not the extractable (or bio-available) concentration.
- The mechanism is normally attributed to adsorption of metals to organic matter, Fe/Al oxides and clay minerals. The abundance of the metals is indirectly correlated to these factors.
- Laboratory studies showed that the adsorption behaviour is pH dependent.
- Many studies have a small sample size ($n < 100$), and do not have independent validation. Multivariate techniques such as PLS regression can easily overfit the data.
- Heavy metal concentrations are usually positively skewed, thus a few high concentration values might produce predictions resulting in higher R^2 values. Care must be taken in interpreting these results. The R^2 value is generally interpreted as the proportion of the variance explained, however it only evaluates the linear relationship between the variables and is insensitive to additive and proportional differences. In addition, R^2 values are quite sensitive to non-normal data ([Kowalski, 1972](#)).
- There are conflicting results on the ability of Vis-NIR to predict certain metals. This indicates that the results or calibrations are site-specific.

3.2. Organic contaminants

There is a very wide range of organic contaminants in soil; too many to cover in a short review so for this paper purpose we will focus on

petroleum-derived hydrocarbons which are the most important class of organic contaminants and also can highlight the possibilities of infrared spectroscopy for their prediction. The hydrocarbons present in crude oils and fuels can comprise hundreds of different organic compounds (Brownlow, 1979) that can be categorised into three main chemical groups: aliphatic (alkanes), aromatic, and hetero compounds containing N, S and O. In addition, a large component of petroleum is made up of cyclo-alkanes (naphthenes) along with combined structures (naphtheno-aromatics). In crude oils, the proportions of these groups can vary depending on the geological history of the oil. The main refinery products (diesel, kerosene, gasoline/petrol) are distillation fractions and differ chemically but also in their physical behaviour. Both diesel and gasoline comprise comparable amounts of straight chained and cyclic alkanes with relatively more aromatic material in the gasoline products. Diesels also contain significant aromatic compounds and phenanthrene has been used as a biomarker given its ease of measurement (Sadler and Connell, 2003), lower volatility and that it is insoluble (i.e. less likely to be washed out or lost to the atmosphere). In practical terms, the most important difference between these two commonly used fuel products is their boiling point (volatility) which is related to the range of chain lengths (C numbers) for each fraction. Diesel comprises C₁₁ to C₂₈ while gasoline comprises C₆ to C₁₀ molecules and as a consequence evaporates more rapidly. Current site assessment practices categorise the hydrocarbons extracted from contaminated soils into the following fractions: F1, F2, F3 and F4. These continuous intervals of increasing C number ranges (C₆–C₁₀, >C₁₀–C₁₆, >C₁₆–C₃₄ and >C₃₄–C₄₀ respectively) may differ in their volatility and soil impact but also have been assessed for their health risk on the basis of health screening levels (Friebel and Nadebaum, 2011).

Conventionally, hydrocarbon contamination has been determined by laborious solvent extraction of known amounts of soil followed by gas chromatographic (GC) separation and analysis usually with a flame ionisation detector (FID) or mass spectrometer (Sadler and Connell, 2003; Paíga et al., 2012). This provides a relative quantification of the range of hydrocarbon compounds persisting in the soil. Alternative extraction processes involve supercritical fluid extraction where CO₂ under high pressure is used as a solvent (Dawson et al., 2004). In later years (1990s onwards), BTEX (the aromatic hydrocarbons benzene, toluene, ethylbenzene and xylene) were the focus of analyses because these hydrocarbons are of environmental concern and were measurable in the field using newly developed field-portable GCs with photo-ionisation detectors (PID) highly sensitive to these compounds (Sadler and Connell, 2003; Paíga et al., 2012). Although soil moisture lowers the response to these organic analytes, the capacity to make determinations on-site after recovering soil air by probe or by taking soil for headspace analyses meant huge savings in time and sample processing. Alternative methods include a purge and trap system where inert gas (helium) is used to flush hydrocarbons to a GC instrument after warming the soil in a closed vessel (Dawson et al., 2004). Another standard method to determine TPH was by infrared analysis of oil-in-solvent soil extracts (using carbon free solvent). These TPH determinations which relied on a fixed-wavelength instrument responding to C–H bond stretching near 2930 cm⁻¹ (3420 nm) appeared to have limitations, mainly because of the insensitivity to aromatic components (Lambert et al., 2001). The full extent of hydrocarbon contamination may not therefore be properly captured by this type of analysis (Dawson et al., 2004) especially after degradation/weathering which can alter the proportions of aliphatic and aromatic material in the soil. Another disadvantage was that it required a C-free solvent. A comparative study between GC/FID and oil-in-solvent by infrared analysis conducted by Becker et al. (2002) also indicated that TPH determinations were 10 to 20% lower via the latter method but with a narrower coefficient of variation than when using GC/FID. This would also probably depend on the diversity of the extracted hydrocarbon compositions.

The benchmark still appears to be a laboratory extraction and GC analysis to obtain the full range of hydrocarbon peaks extracted from the soil for integration of their relative areas. Since 2011 Australian

soil testing groups have started adopting the Canadian approach where hydrocarbon contamination is reported on the basis of their volatile³ and semi-volatile ranges so that the F1, F2, F3 and F4 fractions (obtainable by GC separation/analysis) can be related back to specific health risk (Friebel and Nadebaum, 2011), as for example posed by direct contact.

Total petroleum hydrocarbons are a measure of the sum of extractable aliphatic and aromatic hydrocarbons over the C₆ to C₃₄ molecular size range. These are usually determined by FID peak areas after GC separation. These peaks may include unaltered but also altered compounds (residues) which can, in moderate to extreme cases, produce a hump consisting of an unresolved complex mixture of residues evident in the chromatogram at higher retentions (temperature). Furthermore, depending on the way TPH is carried out, the extent of contamination may be underestimated in terms of the unresolved diesel residues resulting from biodegradation (Lang et al., 2009).

The application of infrared spectroscopy to chemically monitor hydrocarbons in soil and biodegradation processes appears to be underdeveloped with some limited work in this area. Recognising the alteration of hydrocarbons in natural environments by GC and stable isotope methods has been considerably successful (Head et al., 2003; Sun et al., 2005) but these are complex and costly methods whereas spectroscopy could offer a rapid and reliable alternative.

Two patent applications have been published on the use of infrared spectroscopy for assessment of hydrocarbons in soil. Forrester et al. (2012) proposed a method for detecting contaminants based on near and mid-infrared spectroscopy and Ben-Dor et al. (2014) proposed a quantitative assessment of soil contaminants, particularly hydrocarbons, using NIR spectroscopy.

Since the mid-1990s several works in applied spectroscopy research have investigated the feasibility of using NIR and MIR for detecting contaminants in soil such as spilled diesel fuel and motor oil (Stallard et al., 1996), trichloroethylene (Vohra et al., 1996) and methylcyclopentadienyl manganese (a fuel additive) (Vreugdenhil and Butler, 1998). Buerck et al. (2001) devised a fibre-optic sensor system for the determination of aromatic hydrocarbons in groundwater or industrial wastewater. Meanwhile Vohra et al. (1996) also devised another remote sensor to detect trichloroethylene (TCE) in sand. Infrared techniques have been used to quantify PAH concentrations in soil via filtration and analysing the filtrate at certain wavelengths which were calibrated with a known amount of compounds (García Gómez et al., 2004).

Hydrocarbon compounds contain predominantly C–H and C–C bonds but there is plenty of information to be obtained from the infrared spectrum arising from C–H stretching and C–H bending, mainly in the MIR region (2500–25,000 nm). Fundamental vibrations can be observed in the 2500 to 6670 nm region, and furthermore, these compounds exhibit overtone and combination bands in the NIR region (Schwartz et al., 2011). Spectral reflectance characteristics of hydrocarbons in the Vis-NIR region were reported by Cloutis (1989) for applications to the remote sensing of terrestrial and extra-terrestrial targets. This possibility is realized by Hörig et al. (2001) using HyMap, an airborne hyperspectral scanner, for mapping and delineating oil-contaminated soils based on absorption features at 1730 nm and 2310 nm. Kühn et al. (2004) proposed a hydrocarbon index which measures the “depth” of the absorption around 1730 nm to be used in hyperspectral imagery to detect hydrocarbon contamination. This is particularly relevant where the addition of large quantities of organic matter has been part of the remediation process. A review on the use of Vis-NIR remote sensing of contamination areas is given by Slonecker et al. (2010).

Stallard et al. (1996) successfully applied NIR spectroscopy (1600–1900 nm) for the determination of motor oil contamination in a sandy loam. However, Zwanziger and Förster (1998) found challenges in a

³ The volatile TPH fraction refers to C₆ to C₁₀ aliphatic hydrocarbons (i.e. F1) plus BTEX and naphthalene (Schedule B1: Guideline on the Investigation Levels for Soil and Groundwater, National Environment Protection Council OPC50357-B/F2013L00768).

laboratory study and recommended that “spectra of contaminated soil and humus need cleverer spectral selection and pre-processing for better performance of calibration models.” Nevertheless, Malley et al. (1999) demonstrated the feasibility of using NIR to predict TPH in contaminated sites. However, the study only used 26 samples and acknowledged potential sources of error (variability in analytical laboratory results and the influence of moisture in field moist condition). The issue of soil moisture was again highlighted by Hazel et al. (1997) using MIR and this work showed the undesirable effect of moisture on spectra measurement. Another laboratory study by Blake and Gassman (2001) investigated the MIR spectra of the coatings of organic pollutants on the surface of a loamy soil.

Forrester et al. (2010) evaluated the use of Vis-NIR and MIR in three types of materials: adding (spiking) diesel and crude oil (as TPH) into reference minerals, adding oil into reference soils (at a number of TPH concentrations covering the range found in contaminated soils), and actual TPH concentrations (as determined by extraction/analysis) in contaminated soils. Results of PLS cross-validation for the spiked minerals showed that prediction errors (RMSE) with MIR were approximately 2000–4000 mg kg⁻¹ for a TPH range of 0–100,000 mg kg⁻¹ but got slightly higher error for NIR (4500–8000 mg kg⁻¹). A further study by Forrester et al. (2013) indicated that the aliphatic alkyl stretching vibration regions were the most sensitive to TPH: NIR frequencies at 2220 to 2440 nm, and MIR at 2170 to 3330 nm. The peak near 3663 nm (2730 cm⁻¹) was identified as potentially specific to TPH with very little overlap with soil organic matter and soil carbonate. They further demonstrated the applicability on actual contaminated soil samples (n = 205). They showed that the MIR calibration produced a RMSE < 1000 mg kg⁻¹ for the 0 to 15000 mg kg⁻¹ concentration range of TPH, and suggested that the accuracy of such calibrations could be regarded as sufficient for screening purposes.

Chakraborty et al. (2010) analysed TPH-contaminated samples from Louisiana using Vis-NIR and found that scanning samples under field-moist intact conditions showed the greatest accuracy. The poor performance of using air-dried samples is attributed to the loss of TPH during the drying process. Chakraborty et al. (2012b) further studied the spectra patterns by mixing samples from two soils with different clay contents, three levels of organic carbon, three petroleum types (crude oil, diesel and motor oil) with three or more levels of contamination per type. Their results showed that the spectral absorptions for each petroleum hydrocarbon overlapped with each other and could not be separated with any clustering technique. Nevertheless they were able to predict the petroleum amount using a wavelet-based multiple-linear regression method. This led Chakraborty et al. (2012a) to the use of Vis-NIR spectroscopy as a field proximal sensor where the areal extent of TPH contamination was mapped using geostatistics.

Sorak et al. (2012) presented a handheld NIR instrument and evaluated it for analysing oil contamination in a sandy soil. Soil samples were mixed with three types of oil in the concentration range from 0 to 7 dag kg⁻¹. They found good prediction with a PLS model, and in addition their principal component analysis of the spectra showed that the spectra of diesel, oil, and gasoline contaminated soils and uncontaminated soil could be discriminated. However, this laboratory study was based on a single soil whereas the study by Chakraborty et al. (2012b) conducted for soils with differing textures found that different petroleum types cannot be discriminated easily.

To evaluate the accuracy and precision of using NIR to measure TPH concentration, Schwartz et al. (2012) used soils mixed with a known amount of petroleum. NIR prediction was compared with measurement using solvent extraction from three certified laboratories. They found large variations between the results of the three laboratories, both internally (average deviation up to 20%) and between laboratories (average deviation up to 103%). Based on this, they concluded that the Vis-NIR spectroscopy method was as good as the commercial laboratories in terms of accuracy and could be a viable field-screening tool. Furthermore,

Schwartz et al. (2013) proposed an algorithm called the multipath approach for creating the best Vis-NIR prediction model for TPH.

As an illustration for this review, another experiment was conducted by the authors. The results for two test soils which had previously been contaminated with diesel and gasoline prior to spectroscopic analysis are presented in Figs. 3 to 5. Fig. 3 demonstrates our results on the mid-infrared absorbance of two types of soil (a loam with 9% carbon content, and a light clay with 0.7% carbon content) contaminated with diesel while Fig. 4 shows the spectra in the NIR region. Diesel spectra are characterised with the strong presence of alkyl groups in the 2700–3000 cm⁻¹ (3500 nm) region and its overtone in the 1730 and 2300 nm regions. The addition of diesel to soil showed up clearly in the 2700–3000 cm⁻¹ region for both soil types. In the NIR region, however, the addition of diesel to the soil does not present marked features on the spectra (Fig. 4). For the light clay with low carbon content, the 1730 nm region only shows slight changes with increasing diesel concentration. The region at 2300 nm shows a more prominent effect on the spectra. For the loam with a high C content, the interference of organic matter masked the effect of diesel in the spectra.

In the past TPH was mainly used to assess the degree of total hydrocarbon contamination where related analysis techniques could also be used to evaluate their biodegradation. A lot of attention was also given to BTEX and PAH compounds and monitoring their concentrations during remediation. This was enabled by the availability of methods to directly measure these aromatic compounds using specifically dedicated chromatographic instruments. As opposed to TPH, only a few studies have evaluated the possibility of using infrared spectroscopy on soil for quantifying the full range of hydrocarbons which include aliphatic, BTEX and PAH concentrations. Okparanma and Mouazen (2013b) evaluated the use of Vis-NIR spectroscopy for predicting the concentration of phenanthrene in 150 soil samples contaminated with a known amount of diesel. They showed good results with R² values of 0.75–0.83. The PLS coefficients showed that the model used infrared absorbances that relate to hydrocarbon derivatives from combinations or overtones of aromatic C–H functional groups or C–H stretching modes of saturated CH₂ and terminal CH₃ groups. In addition, colour, clay and organic carbon also explained the intensity of the Vis-NIR spectral signal of sorbed PAH in soil. Okparanma et al. (2014) analysed field-moist intact soil samples collected from three oil spill sites in Ogoniland in the Niger Delta province of Nigeria. Using sequential ultrasonic solvent extraction–gas chromatography as the reference chemical method, they predicted PAH concentration using Vis-NIR spectra. The results showed good prediction with a validation R² ranging from 0.77 to 0.89. Bray et al. (2009) used an ordinal logistic regression technique to evaluate the probability of exceeding a certain threshold for total PAH and Benzo[a]pyrene using Vis-NIR spectroscopy. Their models showed good accuracy (90%) at low contamination threshold but decreasing accuracy with increasing contamination thresholds defined by the Australian and New Zealand Environment and Conservation Council.

The condition of soil samples can affect the analysis of PAH. Narizzano et al. (2013) classified PAHs into two groups: PAHs lighter than pyrene which are seriously affected by drying temperature; and heavier PAHs that can be considered as non-volatile compounds. Thus lighter PAHs and especially BTEX need to be analysed in situ, and the effect of moisture on infrared spectra should be taken into account. Okparanma and Mouazen (2013c) evaluated the interaction effect of oil concentration, moisture and clay contents on the spectral characteristics of diesel-contaminated soils and the quality of calibration models.

In the soil environment, hydrocarbons are also subject to alteration (sometimes referred to as weathering) through preferential loss of the more volatile or soluble hydrocarbons. For example, lower molecular weight aromatics can be quite soluble and have been found in groundwater close to oil-fields (Zarella et al., 1967). Therefore, aromatic hydrocarbons such as benzene are likely to be transported in soil by ‘water washing’ if not volatilised where soil-adsorption is poor (Paíga et al., 2012). Fig. 5 shows the spectra of a gasoline sample (91 octane

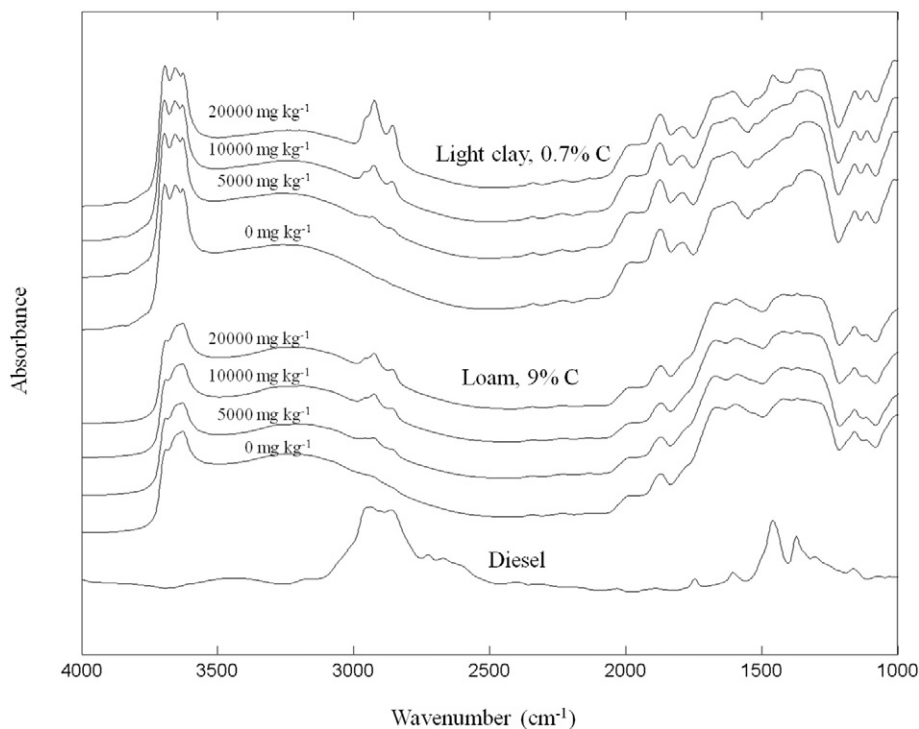


Fig. 3. Spectral absorbance in the mid-infrared region of diesel (in KBr), and a loam (C content = 9%), and a light clay (C content = 0.7%) contaminated with diesel at various concentrations.

rating) in KBr analysed within an hour and then after 12 h at 20 °C. There is a loss of absorbance features associated with the gasoline functional groups across the MIR range and disappearance of the characteristic absorbance at 2730 cm⁻¹ (aldehyde C–H bond). The spectra of gasoline contaminated soil do not show marked differences, although

subtle changes in the 3000–3500 cm⁻¹ range can be detected when the contaminated soil is analysed within an hour of gasoline addition.

Degradation by biological agents (microbes) also produce alteration effects on the remaining compositions by preferential oxidation where straight chained hydrocarbons are metabolised more easily before cyclic

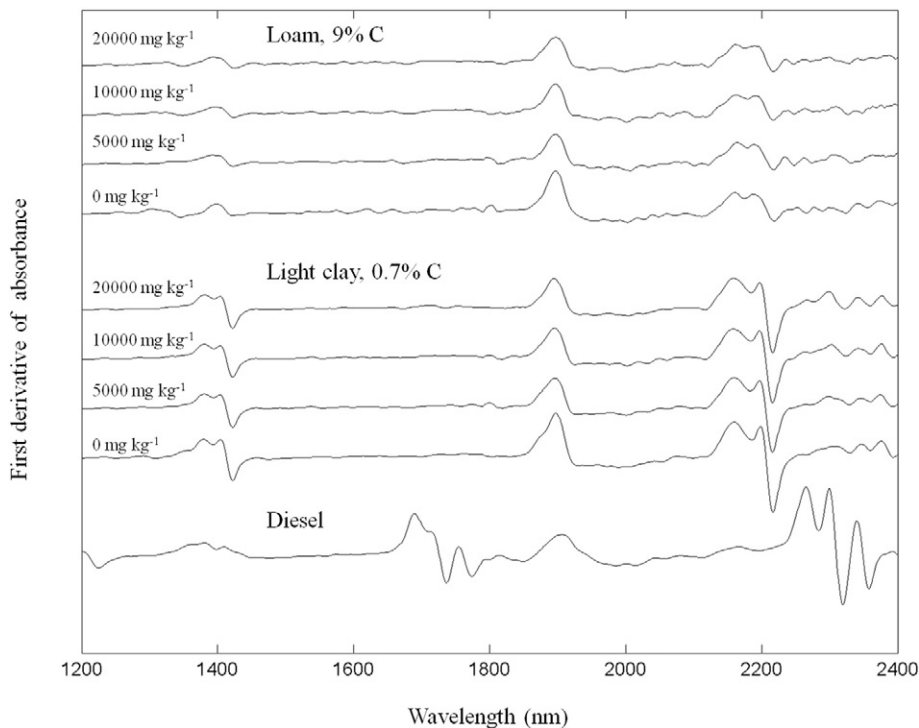


Fig. 4. First-derivative spectral absorbance in the near-infrared region of diesel (in KBr), and a loam (C content = 9%), and a light clay (C content = 0.7%) contaminated with diesel at various concentrations.

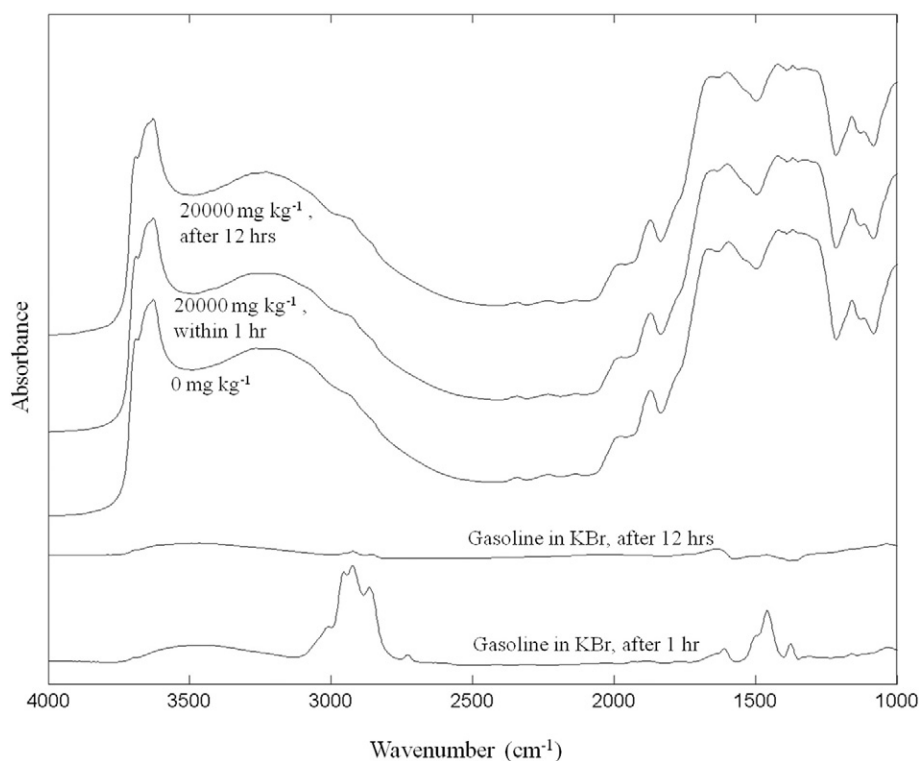


Fig. 5. Absorbance spectra of a gasoline sample in KBr and a loam, analysed within an hour and after 12 h at 20 °C.

and aromatic molecules are broken down (Das and Chandran, 2011; Head et al., 2003). Therefore, aliphatic hydrocarbons which are insoluble under normal conditions become more rapidly degraded by bacteria/fungi. This is probably the most important process when considering hydrocarbon-contaminated soil and their possible markers since compounds may not only be diminished through biodegradation but also result in chemical modification. Furthermore, the degradation of PAHs (including phenanthrene) may be augmented in the presence of aliphatic material (Swindell and Reid, 2006). A major product of hydrocarbon biodegradation under aerobic conditions may be mineralisation to CO_2 (ultimately a form of bioremediation clean-up) but it is also likely that parts of the remaining compounds become more oxygenated structures. Other than what has been presented by Bhat et al. (2011) or Wu et al. (2010) who studied biodegradation of transport fuels and phenanthrene respectively, it appears that nothing else has been published using infrared spectroscopy to monitor the transformation of hydrocarbons in soil. Their work indicated that alteration to the chemical composition of hydrocarbon components may occur in hours to days after contact with microbes which results in significant changes to the MIR spectral characteristics. The most significant information to be drawn was that the typical alkyl (CH_2 and terminal CH_3 methyl) absorbance peaks diminish over the period of the incubations and that new bands form consistent with oxidised units (carbonyl and OH groups).

3.3. Other applications

Near infrared spectroscopy also has been used for different types of contaminants and other soil-like materials. Sut et al. (2012) investigated the use of a portable NIR spectrometer for prediction of cyanide (CN^-) concentration in soils at a former manufactured gas plant in Germany. They concluded that the portable NIR instrument could be a reliable device for detecting cyanide concentrations $>2400 \text{ mg kg}^{-1}$ in the field. They further recommended that while NIR cannot replace traditional laboratory analyses due to high limits of detection, it could be used for the identification of contamination hot spots.

NIR spectrometers are also used for measurement of organic matter and heavy metal concentration on sewage sludge samples (Galvez-Sola et al., 2013), evaluation of eutrophication potential on aquatic sediments (Kleinebecker et al., 2013), and evaluating phosphorous contamination in agricultural soils (Bogrekci and Lee, 2005).

In summary,

- Vis-NIR and MIR spectroscopy should be able to readily predict organic contaminants due to the ability to detect chemical functional groups. However in soils, the spectra can be affected by soil mineral interactions and organic matter, creating peak shifts and disappearance of the peaks compared to the pure absorption spectra of the contaminants.
- MIR appears to be excellent in detecting hydrocarbon levels in soil. Absorbance near $2700\text{--}3000 \text{ cm}^{-1}$ can be assigned for the aliphatic content from diesel with great certainty despite normal concentrations of soil organic matter. Meanwhile there are more challenges using NIR spectroscopy for hydrocarbon because of the less defined overtone peaks and the complex interaction with soil organic materials.
- Most studies on oil contaminants have been performed in laboratories where the contaminants were added to one or several soil types at varying concentrations. In addition, many of the studies have a small number of observations used in calibration and do not have independent validation. Therefore the extrapolation of these results or PLS models to real-world data should be done with caution.
- Challenges for infrared spectroscopy for hydrocarbon contaminants include evaluating the effect of volatilisation losses and degradation of the petrol compounds during sample collection and analytical process. Thus in situ measurements are recommended.
- Soil NIR and MIR spectra are well-known to be affected by soil moisture. Studies have shown the need to measure TPH and PAH compounds under field conditions. There is no study which has looked into the moisture correction of the spectra, all studies just calibrated field observed spectra with laboratory-determined concentrations of TPH.
- Similar to heavy metals, TPH and hydrocarbon concentrations are usually positively skewed thus few high concentration values can inflate

the R^2 values (Bellon-Maurel et al., 2010). Salmeen et al. (1995) hypothesized TPH in soil has a log-normal distribution. Therefore careful interpretation of the prediction results is required. This also has implications for the sampling strategy.

- All of the calibration studies are empirical using PLS regression which relates the spectral absorbance to known concentration. Therefore the transferability of the model to new fields should be done cautiously. Models developed using fresh soils spiked to soil that have not been subject to real environmental conditions, physical and microbial, can limit their application. Although many studies quoted the importance of various bands in the Vis-NIR and MIR spectra as predictors, they are interpreted a posteriori, after the model was generated. This means that the bands are only being used to explain coefficients of the model. In addition, the uncertainty of prediction is never quantified.

4. PXRF for measuring soil contaminants

In the last twenty years or so, portable X-ray fluorescence (PXRF) analysers have been recognised as a possible way of measuring soil contaminants as they offer rapid, real-time, simultaneous multi-elemental analysis of soil samples in a solid condition with minimal or no sample treatments (West et al., 2013). The use of these devices has been well established in the laboratory but with recent advances in PXRF technology their potential use in the field has become very attractive for environmental surveying (Hou et al., 2004).

XRF spectrometers make use of the fact that every atom consists of a specific number of electrons positioned on energy shells (e.g. K, L and M-shell) around their nucleus. They operate on the principle that electrons from the inner energy shell of any atom leave their shell when excited with X-rays. These vacant positions are then filled almost immediately with electrons from a higher, outer, energy shell which results in the release of excess energy in the form of an X-ray photon (fluorescence) which is characteristic of the elements present in the soil sample (Potts, 2008). The corresponding wavelength of the XRF is therefore dependent on the energy level of the electrons in the inner shells of the atom. Furthermore, the fluorescence emission depends on the atom's principal inner shell involved in the excitation process (Hou et al., 2004). In theory, XRF detectors can therefore record the resulting X-ray spectrum of any element. However, generally, due to low-energy responses, elements of the periodic table that have an atomic number smaller than 12 (elements below magnesium) cannot be effectively measured.

Portable XRF spectrometers operate on the principle of energy dispersive X-ray fluorescence spectrometry whereby the amount of emitted fluorescence photons is directly measured by an X-ray detector

that simultaneously analyses their energy levels. The elemental composition of the soil sample is thus determined from the measured intensities of emitted fluorescence. The resulting XRF spectrum of the soil sample analysed then plots energy versus intensity as shown in Fig. 6. Most PXRF devices operate with an internal factory installed calibration generally based upon the Compton Normalization (CN) method that is unique to the relevant mode of operation (e.g. SOIL operational mode (Olympus InnovX-Systems, 2010)) and estimate the elemental concentration of the soil directly in mg kg^{-1} (ppm) or dag kg^{-1} (%).

Early PXRF devices employed sealed radioactive energy sources, but these were soon replaced by X-ray tube excitation sources which offer more versatility as they operate with higher and variable energy ranges for electron excitation. Extensive reviews on the development of PXRF devices over the past decades can be found in Hou et al. (2004), Potts and West (2008) and West et al. (2013).

In recent years the accuracy of PXRF devices has increased significantly with limits of detection low enough to measure soil contamination of several metals of interest (e.g. $<5 \text{ mg kg}^{-1}$ for Hg, Mo, Pb, Tl, Zn; $<10 \text{ mg kg}^{-1}$ for Ag, Cd, Co, Cr, Cu, Ni, V; $<20 \text{ mg kg}^{-1}$ for Ba, Sb, Sn (Olympus InnovX-Systems, 2010)) as well as non-metals of interest (e.g. $<5 \text{ mg kg}^{-1}$ for As, Se (Olympus InnovX-Systems, 2010)). There are, however, a number of factors which affect the detection limits and therefore the precision of the measurements. The length of the analysis time per sample is one of the factors that influence the performance of PXRF devices. Longer excitation times ($>1 \text{ min}$) generally improve the limit of detection as well as the precision of the measurement (Hou et al., 2004). In addition, the nature of the sample matrix can influence the measurement quite significantly as matrix effects can either attenuate the intensity of the characteristic X-rays or increase the intensity of scattered X-rays within the resulting fluorescence spectrum (Ge et al., 2005). Physical matrix effects that have an impact on the PXRF results are related to the soil's mineralogy, particle size, surface irregularities and moisture. As outlined in Markowicz (2008), soil samples with a large amount of fine particles tend to generate higher concentrations of the analyte when compared to samples containing larger particles despite equal concentrations of the analyte of interest. This effect is especially relevant for elements with small atomic numbers such as K, Ca, Ti, V, Cr and Mn. In addition, a homogenization of the sample matrix is recommended as it improves the overall measurement of the analyte. Furthermore, flat sample surfaces are required as irregularities influence the detected X-ray fluorescence intensities. Furthermore, soil moisture can affect the X-ray fluorescence spectra quite significantly. Its interference can cause a significant underestimation of the elemental composition of soils. It is therefore recommended to scan soil samples in air-dry condition (Kalnicky and Singhvi, 2001). But, as cited in Markowicz (2008) and Laiho and Perämäki (2005) moisture effects may be safely

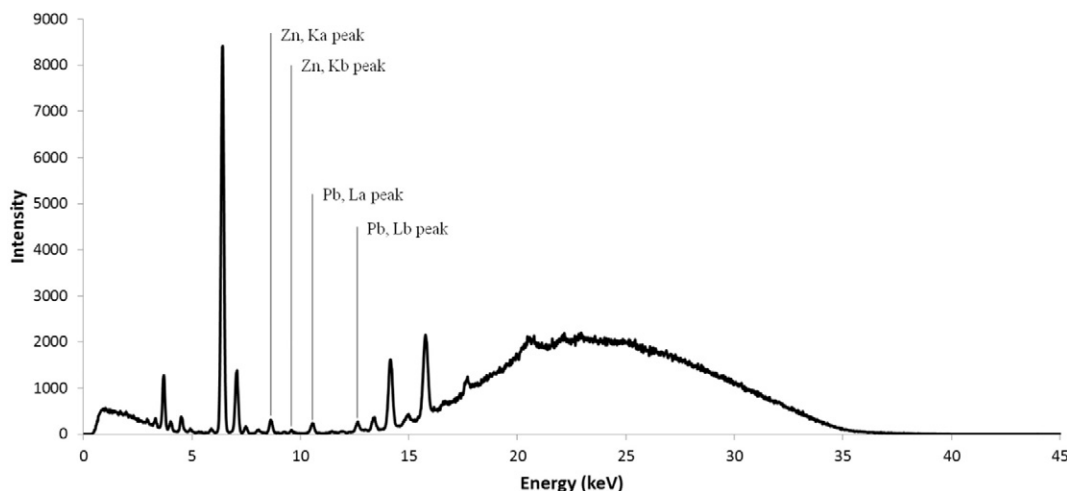


Fig. 6. PXRF spectrum of a sample of contaminated soil (Beam 2 in SOIL operational mode – 40 keV) (Delta Premium Portable Handheld XRF Analyzer, Olympus InnovX-Systems).

neglected for soil samples with a (gravimetric) water content, w , of less than 20 dag kg^{-1} (%). In this regard, Ge et al. (2005) performed a study investigating the influence of moisture ($w = 5\%, 15\%, 20\%, 25\%, 30\%, 35\%$) on the estimation of elemental concentrations with PXRF and proposed a method for correcting the effects caused by the presence of water in soil samples based on the principle that the intensity of scattered X-rays is directly proportional to the gravimetric water content ($w - \log(I_0/I_w)$ where I_0 represents the X-ray intensity of air-dried samples and I_w the X-ray intensity of field-moist samples). Recently, Weindorf et al. (2014) performed a study comparing PXRF readings (Ba, Ca, Cr, Fe, K, Mn, Pb, Rb, Sr, Ti, Zn and Zr) of frozen soil, and melted soil/water mixtures as well as moisture corrected soil to oven dry, ground soil samples. The authors found that PXRF readings of soil samples with moisture contents less than $w = 40\%$ acceptably compared to those of dry samples (e.g. with R^2 values ranging from 0.446 (Mn) to 0.930 (Sr)), and that moisture-corrected sample readings provided the best correlations to the dry, ground samples. Chemical matrix effects are related to interferences of the X-ray fluorescence spectrum of certain elements in the presence of certain other elements which can cause the absorption or enhancement of X-rays. For example, the characteristic X-ray intensities of Zn are strongly absorbed in the presence of high concentrations of Fe, whereas Cr levels are enhanced. Most of these chemical matrix effects, however, are accounted for and corrected by the PXRF software (Kalnicky and Singhvi, 2001). In addition, spectral interferences may result in limits of detection of certain elements (Markowicz, 2008). First, these are related to the energy resolution of the X-ray detector which may not be sufficient enough to fully separate the energy differences between two peaks whose spectra are smaller than the detector's resolution. Secondly, spectral interferences arise in cases where the energies or characteristic X-rays of two or more elements are very similar and therefore overlap. Furthermore, the thickness or density of the soil sample is also an important factor that influences the accuracy of the PXRF measurement as X-rays usually penetrate about 8 mm into the soil material during the excitation process (Ramsey, 2008). A sample thickness of about 15 mm is therefore recommended for PXRF scanning of bulk soil. In summary, to achieve best quantitative results, sample pre-treatments are required to eliminate the influence of moisture and to reduce effects of soil matrix heterogeneities. Sample pre-treatments may then involve drying, grinding, sieving (to less than 2 mm) and powderisation (to less than 200 μm) of the soil sample.

PXRF devices are used for in situ and ex situ measurements; in situ refers to the in-field use where the PXRF is placed directly onto the soil surface or soil samples are scanned in field condition through plastic bags; and ex situ refers to the scanning of pre-treated samples that were air-dried, ground and sieved and scanned through a plastic bag or powderised and pressed and scanned in a sample cup. Generally, based on their accuracy in situ measurements are regarded as qualitative/semi-quantitative results whereas ex situ measurements are seen as semi-quantitative/quantitative results. In this regard, Laiho and Perämäki (2005) conducted a study evaluating different sample preparation methods and their influence on PXRF measurements of soils from contaminated sites. They found that soil moisture and particle size are the main factors influencing the accuracy of the results. In their study they also tested the performance of several PXRF instruments and concluded that broadly, sample preparation strategies influenced the final results more than differences in elemental readings between PXRF instruments (Olympus Innov-X-Systems, USA; Metorex Inc., Finland; Niton Corp., USA).

Based on the interferences of the XRF spectrum outlined in the previous paragraphs, a sampling protocol for measuring contaminated soil samples by PXRF in situ and ex situ is proposed (Table 1).

Over the years, several studies have compared the performance of portable XRF techniques to provide fast and accurate analytical results to laboratory-based, conventional methods of analysis such as chromatographic separation and spectroscopic techniques (e.g. atomic absorption spectrophotometry (AAS) and inductively coupled plasma

atomic emission spectrometry (ICP-AES)) and found good agreement for a range of metals (e.g. see reviews by Kalnicky and Singhvi, 2001; Ramsey, 2008; West et al., 2013). Conventional, exclusively laboratory-based methods involve time-consuming sample preparations which result in a soil digest suitable for analytical measurement. These sample pre-treatments involve the use of concentrated, harmful acids (e.g. hydrochloric acid (HCl), nitric acid (HNO_3), hydrofluoric acid (HF)) for sample digestion and elemental extractions into a solution.

For example, Weindorf et al. (2008) tested the performance of PXRF compared to ICP for determining the elemental composition of compost products. They scanned samples in dried, ground condition, pressed into a sample cup. They found good correlation with metals (e.g. As, Cu and Zn with R^2 of 0.844, 0.946, 0.811, respectively), but concluded that ICP analysis provided better detection of elements at low levels ($<5 \text{ mg kg}^{-1}$) which is related to the limit of detection of the PXRF instruments. Following on, McWhirt et al. (2012) tested the applicability of PXRF for rapid determination of elemental concentrations in compost, and found that the method showed the most potential for Ca, Cr, Cu, Fe, K, Mn, P and Zn for dried samples. Radu and Diamond (2009) conducted a study monitoring soil pollution caused by silvermining dust in Ireland and achieved very good correlation of PXRF measurements with laboratory-based AAS results for Pb, As, Cu and Zn (R^2 of 0.995, 0.991, 0.959 and 0.843, respectively). However, in most cases PXRF values were slightly biased when compared to AAS results (slope <1 for Pb and Cu; slope >1 for As and Zn) which was attributed to the method of analysis of the PXRF instrument. In this study samples were scanned in oven-dry (50°C), ground ($<1 \text{ mm}$) condition through plastic bags. Weindorf et al. (2012a) studied the use of PXRF for environmental quality assessment of peri-urban agriculture and found that PXRF scanning resulted in quality results of heavy metal levels comparable to conventional laboratory techniques. For the majority of the trace elements studied, reasonably well correlation was found between PXRF and ICP data (As, Co, Cu, Fe, Mn, Pb and Zn with R^2 of 0.348, 0.356, 0.481, 0.454, 0.526, 0.955 and 0.859, respectively). Poor coefficients of determination for Ba and Cr (R^2 of 0.074 and 0.004) were attributed to some clearly identifiable outliers in the distribution.

Fig. 7 shows an example of estimating levels of heavy metal (Pb and Zn) contamination in soil using PXRF as compared to conventional laboratory measurements (ICP analysis after *aqua regia* extraction). This was an experiment conducted by the authors for this paper purpose. Samples were scanned through plastic bags in air-dry, ground condition. Overall, PXRF estimations of levels of Pb and Zn show good agreement with laboratory measurements.

PXRF devices have also been used successfully for in-field environmental screening to detect the occurrence of elevated concentrations of metals in the soil (Carr et al., 2008; Fujimori and Takigami, 2014; Kalnicky and Singhvi, 2001; Radu et al., 2013; Weindorf et al., 2013b). For example, Weindorf et al. (2013b) used a PXRF spectrometer to survey a site in Zlatna, Romania, with a history of mining/smelting operations. Based on a random sampling approach, 69 soil samples were scanned in field condition across multiple land-use types, resulting in the identification of more than 50% of the scanned sites exceeding the Romanian action limits for (heavy) metal contaminants (As $> 25 \text{ mg kg}^{-1}$, Cd $> 5 \text{ mg kg}^{-1}$, Co $> 50 \text{ mg kg}^{-1}$, Cu $> 200 \text{ mg kg}^{-1}$, Pb $> 100 \text{ mg kg}^{-1}$). The authors concluded that the use of georeferenced PXRF data offers a powerful tool for the in-situ assessment and areal identification of contaminated soil. Similarly, Radu et al. (2013) successfully mapped the distribution of the main anthropogenic metals (As, Cd, Co, Cr, Cu, Hg, Ni, Pb, V and Zn) at two sites in Ireland. Fujimori and Takigami (2014) investigated the distribution of heavy metals (Pb, Cu, Zn) in surface soils at an electronic waste recycling workshop in Manila, Philippines. They divided the site into five areas based on the distance from an entrance gate (y-axis) and found that pollution decreased to half of its maximum with distance from the pollution hot-spot.

Table 1

Protocol for the sampling and measuring of contaminated soil samples by PXRF (after Laiho and Perämäki, 2005).

In-situ analysis	Ex-situ analysis
<p><i>In-field</i></p> <ul style="list-style-type: none"> – Remove stones and plant fragments – Prepare a flat, even soil surface for scanning – Use at least 1 min of count time – Perform a minimum of 3 replicate measurements – Measure a blank sample, and standard reference samples every 10–20 samples 	<p><i>Plastic bag</i></p> <ul style="list-style-type: none"> – Air-dry the sample (30–40 °C) – Scan the sample (100–500 g) in ground condition (<2 mm) – Use a sample thickness of at least 15 mm – Use bags with thin plastic walls – Flatten the plastic bag so that the soil sample is distributed evenly on the surface – Place plastic bag on detector window so that the latter is fully covered – Use at least 1 min of count time – Perform a minimum of 3 replicate measurements – Measure a blank sample, and standard reference samples every 10–20 samples
<p><i>In-field (plastic bag)</i></p> <ul style="list-style-type: none"> – Sample about 500–1000 g for sample preparation – Pre-homogenize sample manually – Place sample in sample bag for scanning in field-condition or/and air-dry (100–500 g) – Use a sample thickness of at least 15 mm – Use bags with thin plastic walls – Flatten the plastic bag so that the soil sample is distributed evenly on the surface – Place plastic bag on detector window so that the latter is fully covered – Perform a minimum of 3 replicate measurements – Measure a blank sample, and standard reference samples every 10–20 samples 	<p><i>Sample cup</i></p> <ul style="list-style-type: none"> – Air-dry the sample (30–40 °C) – Ground the sample (<2 mm), followed by powderisation (<200 µm) – Scan sample in powder form, pressed in sample cups (at least 15 mm thick) – Extend the scanning time to >1 min – Prepare duplicate samples – Use >3 replicate measurements – Measure a blank sample, and standard reference samples every 10–20 samples – Prepare calibration samples to determine element-specific correlation curve (compare PXRF results to laboratory results)

In summary, these studies concluded that PXRF is a promising method to be used as a rapid in-field analytical technique as it has the ability to result in viable direct, in-situ, non-destructive measurement of the soil's elemental concentration. But some regulatory limits of the metals of interest may not fall above the limit of detection of the PXRF device. The in-field areal assessment of contaminated sites, however, requires an appropriate field-sampling design to account for spatial variability.

PXRF is now widely accepted for field screening of elemental concentrations of environmental samples (Hou et al., 2004). In this regard, some official methodologies now recognise PXRF technology for use in in-situ measurement of soil contaminants such as the United States Environmental Protection Agency (US EPA) Method 6200 which in turn also recognises the instrument's internal factory-installed calibrations for elemental detection (Radu et al., 2013; Weindorf et al., 2014). Furthermore, the latest USDA-NRCS soil survey and laboratory method manual now also includes a referenced method for PXRF use (USDA-NRCS, 2014). The reliability of PXRF as an analytical method to detect a range of heavy metals in contaminated soil in-situ has been investigated by Wu et al. (2012). Pairwise comparison between PXRF and ICP-AES found that PXRF is highly applicable to measure Pb, Zn, Ni and Cu, with detected elemental concentrations well in the excess of pollution threshold limits and correlation of R^2 of 0.668, 0.655, 0.728, and 0.409, respectively. However, in this study somewhat a poor correlation was found for Hg, Cd, Cr and As. Soils were homogenized and scanned in field condition through plastic bags.

As discussed, PXRF is suitable for determining the concentration of a range of elemental soil contaminants, mainly heavy metals, accurately. But recent studies have also shown that PXRF can be employed to calibrate for a range of soil properties (McLaren et al., 2012; Weindorf et al., 2009, 2012b, 2012c, 2013a; Zhu and Weindorf, 2009; Zhu et al., 2011) and in addition can also be used to investigate the degree of soil weathering (Che et al., 2012). In a recent study, for example, Wang et al. (2013) also presented the complementary use of PXRF and Fourier transform NIR spectroscopy to predict soil texture.

These studies use the elemental concentrations measured by the PXRF device to predict certain soil properties generally applying multiple linear regression models. There is the opportunity however to infer soil properties directly from the raw XRF spectra employing chemometric approaches which will be discussed further in Section 5.

To summarise, the following lists the advantages and disadvantages for PXRF measurements of contaminated soil:

Advantages:

- Rapid, immediate, non-destructive, simultaneous measurement of metal contamination of soils in-situ
- Detection limit of PXRF devices is small enough to measure significant contamination of a range of metals/elements of interest (As, Ba, Co, Cr, Cu, Hg Mo, Ni, Pb, Sn, V and Zn)
- PXRF devices have a relatively wide dynamic range, i.e. they are able to detect elemental concentrations from the low ppm

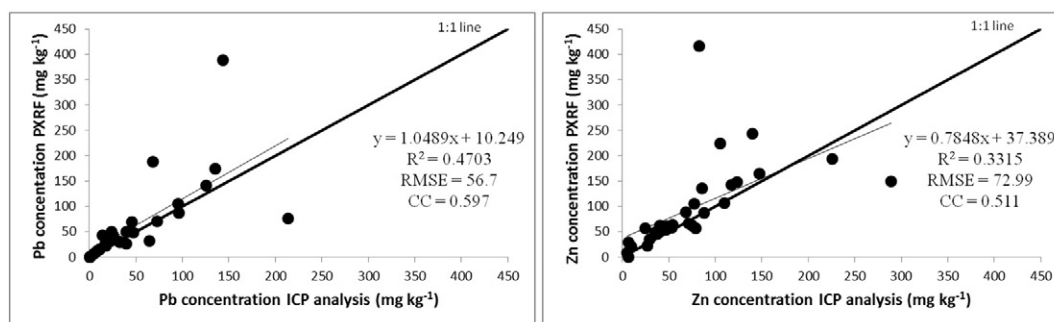


Fig. 7. Goodness of fit of Pb and Zn concentrations estimated using PXRF as compared to ICP analysis (R^2 – R squared coefficient of determination, RMSE – root mean squared error, CC – concordance correlation coefficient which measures the agreement between PXRF and ICP analysed samples, i.e. the closeness to the 1:1 line).

- (mg kg⁻¹) to the high dag kg⁻¹ (%) range with no sample pre-treatments
- In-situ environmental areal based screening of soil contamination
 - Good analytical precision.

Disadvantages:

- Limits of detection varies for the range of elements measured by PXRF
- Possible high degree of uncertainty of in-situ measurements because of sample heterogeneity and influence of soil moisture.

5. Taking IR and PXRF to the field and the role of spectral libraries

The timely characterisation of soil contaminants as well as assessing their magnitude and spatial distribution mitigates risks to human health and environmental quality. As generally mentioned in Sections 3 and 4, both IR and PXRF can potentially overcome much the cost and time-consuming efforts related to traditional laboratory analyses of many soil contaminants. This gain in efficiency further improves efforts for soil contaminant mapping because sampling can occur more intensely and/or frequently. Nevertheless, some aspect of timeliness is still unaddressed with these efficiencies because after sampling, the samples still need to be processed and scanned in the laboratory. A logical step therefore, peripherally discussed in Sections 3 and 4, is to take the instruments out into the field, where predictions and interpolations (mapping) can be made on the spot, and in real time. This is discussed in detail in this section.

A first consideration for field operation of IR and PXRF instruments is one of portability. A widespread phenomenon with many computers, telephones (and related devices), and analytical instruments is that as the technology improves and is refined, there has been a tendency to move towards smaller and thus more portable devices (Eren, 2003). For XRF instrumentation, portability is a relatively new extension of the technology, where it was originally developed for the mining industry for low-cost mineral prospecting. As discussed in Section 4 the availability of an efficient radioisotope source, coupled with highly sensitive detectors and associated electronics, has seen portable XRF technology gain wide acceptance as a viable analytical approach for mining and environmental applications including assessment of contaminated soils – principally those contaminated with heavy metals (Kalnicky and Singhvi, 2001; Weindorf et al., 2012a, 2013b; Fujimori and Takigami, 2014). Portability of IR devices by comparison is more advanced than PXRF. Investigations using field-based infrared however have largely been concerned with agronomically important soil variables such as soil carbon content, soil texture and to some extent nitrogen content, soil colour, and clay mineralogy (Ben-Dor et al., 2008; Waiser et al., 2007; Morgan et al., 2009; Viscarra Rossel et al., 2009; Stevens et al., 2008; Gras et al., 2014). These studies have been predominantly concerned with the use of visible and near infrared spectrometers. Comparatively, the use of portable mid-infrared spectrometers is more limited (Reeves, 2010; Merry and Janik, 2001).

Despite promising results, in-situ soil conditions such as soil moisture, structure, stoniness, coarse organic residues, smearing and small scale heterogeneity, mottles and redox features affect the overall performance of both field-based IR and PXRF (Ge et al., 2005; Stenberg, 2010). XRF signals are also highly matrix dependent, meaning that without proper calibration, predictions may display systematic biases (Hou et al., 2004). Most of these factors or in situ soil conditions can largely be avoided in the laboratory via drying, screening, grinding and homogenising the soil samples to be analysed. In the field however they confound predictions and are sources of error. Further sources of error generally avoided in the laboratory are contamination of the probe by dust (Baumgardner et al., 1985), variable changes in distance between sensor and soil (Hou et al., 2004), and even changes in

background solar illumination (Baumgardner et al., 1985). Soil moisture unarguably affects the accuracy of soil predictions of soil properties (Kalnicky and Singhvi, 2001; Ge et al., 2005; Minasny et al., 2011). There is however some discussion of the extent to which factors such as soil aggregation and heterogeneity affect the reliability of in situ derived soil prediction for Vis-NIR. For example, Waiser et al. (2007) found that clay contents predicted from Vis-NIR scans of dried in situ soil are more accurate than predictions from scans of field-moist in situ soil. Scans of dried and ground soil resulted in the most accurate predictions of clay content (Waiser et al., 2007). In this case the effect of soil moisture appears to be more attenuating than soil structure or aggregation. A study by Fontan et al. (2010) had similar findings when comparing results derived by scans of dried ground soil and dried soil clods. In contrasting results, the effect of soil smearing resulted in less accurate predictions of clay in Waiser et al. (2007), yet had no effect when predicting soil carbon in Morgan et al. (2009). In consideration of these studies, soil moisture undoubtedly poses a problem for soil property prediction in situ for either instrument. It is also apparent that other in situ soil conditions can also reduce the accuracy of predictions. However, some further investigation is required specifically concerned with soil contaminants and how in situ soil conditions affect the predictive performance from both instruments. Alternatively, one may compensate the less accurate in situ predictions with larger sampling intensities. The maxim that applies here is as follows: given spatial heterogeneity many moderately accurate predictions are equal to, or better than few, very accurate measurements.

Efforts for correcting for the effect of soil moisture have received some attention in both the Vis-NIR and PXRF literature. External parameter orthogonalisation (EPO) (Roger et al., 2003) has been used as a Vis-NIR pre-processing step to remove the effect of soil moisture from spectra (e.g. Minasny et al., 2011). The algorithm consists of building the subspace where soil moisture is manifested using an experimental design consisting of collected spectra from the same sample with and without soil moisture i.e. field condition and air dried condition respectively. The spectral influence of soil moisture is then removed from the total spectral space by orthogonal projection (Gobrecht et al., 2014). Minasny et al. (2011) demonstrated an improved calibration and validation in soil carbon concentration using EPO-transformed spectra. EPO transformation of spectra from Ge et al. (2014) applied the EPO transformation so that lab-based Vis-NIR models can be used to predict properties of field moist spectra.

PXRF is not a suitable corrective step because soil moisture is manifested differently compared to that for Vis-NIR spectra. As detailed by Ge et al. (2005), for PXRF, the intensity of characteristic X-rays of analytes (elemental concentrations) decreases as the water content of the sample increases owing to the stronger X-ray absorption by water than by air in fractures and pores in samples. Secondly, the intensity of scattered primary X-rays from the source is increased as the water content of the sample increases. The result of these mechanisms is that net peak areas of characteristic X-rays from the analyte elements are reduced, consequently leading to lower precision, detection limit and accuracy. Under the assumption that the matrix component of the soil sample is constant, Ge et al. (2005) stipulated that the reduction in characteristic X-ray intensity of an analyte is directly proportional to the concentration of moisture in the soil. Subsequently, these authors proposed a simple corrective step that was shown to work effectively for wet samples up to 20 dag kg⁻¹ (%) gravimetric water content. For field situations the main points taken from the above studies is that Vis-NIR spectra can be effectively corrected for soil moisture, and similarly for PXRF up to 20% gravimetric water content; soil containing 20 dag kg⁻¹ (%) or more clay are often wetter than this. More importantly however, because soil moisture needs to be known in order to correct the X-ray intensities (for PXRF), a soil moisture probe will be required in the field. Further to this is that field operators will need to consider the suitability of soil sampling when soil is very wet i.e. immediately after a heavy rainfall event, or flooding occurrence.

Generating estimates of soil contaminants in the field appears an attainable goal. Yet there are further practicalities to consider that come in the form of how estimates are made (from each instrument), and to some degree, an apparatus is also needed in addition to the Vis-NIR and PXRF instruments to facilitate whole-profile or down-hole measurements – which becomes more-or-less an engineering problem. Prediction of soil contaminants, principally heavy-metal concentrations, with PXRF can be made directly with the device as software and internal factory installed calibrations will be already pre-installed onto the device; although some additional calibration with some measured standards may be necessary to account for specific soil matrix effects. Alternatively, the procedure would be to pre-calibrate the instrument for a range of target soil properties, then in the field apply those calibrated models to generate predictions on the spot. Although soil properties (other than elemental concentrations) may be inferred from empirical calibration models using the elemental concentrations as predictors (Zhu et al., 2011), yet this has not been performed in the field. For Vis-NIR predictions of soil properties, methods are purely empirical. Some methods are process-based, for example soil colour and clay mineralogy which can be predicted using known specific absorption features (Clark and Roush, 1984). Yet the main empirical method is one of classical chemometrics where measured soil properties are calibrated against corresponding soil spectra as detailed in Section 3. In the instance of chemometric approaches for either PXRF or Vis-NIR, a significant amount of pre-field work is necessary to be undertaken before in situ predictions can be made.

The first step is the collection of a spectral library/ies that contain the ideal duo of target soil properties/contaminants and corresponding soil spectra. Once in possession of these, the routine of spectral pre-processing, followed by fitting of any number of calibration model types (Soriano-Disla et al., 2014) is performed. Naturally some form of both internal and external validations of the calibration models is performed to assess their fitness for predicting target soil contaminants. In essence, it is then just a matter of saving the calibration model/s to a computer, then taking the computer out into the field. In the field, a sample is collected, then scanned (ideally multiple times to increase signal-to-noise ratio). The resulting spectra are pre-processed following the same methods used for the calibration of model spectra. The processed sample spectra are then fed into the prediction models, after which predictions of target soil properties are given.

This straightforward description is the basis and framework for an in situ soil spectral inference engine (SSIE). The contextual use of this is described further in Section 7. The SSIE could range in complexity – very plain as described above – or could contain multiple functions and sub-processes to deal with the foreseeable issues expected for using the devices in the field, namely soil moisture, soil structure, aggregation and heterogeneity. Before considering these issues, a comprehensive SSIE will contain, just for one spectral library, a possible multiple number of versions of differently pre-processed spectra together with associated (and possibly multiple) calibration models (Gras et al., 2014). It is up to the discretion of the operator to eliminate redundancy from the SSIE by removing those processes and models that do not contribute a lot in terms of accurate predictions, as determined from validation procedures. For Vis-NIR, calibration models will be based on EPO transformed spectra. Subsequently, an experimental transformation matrix is necessary in the SSIE in order to remove the soil moisture effects from in situ scanned samples. For PXRF, information on measured soil moisture will be entered into the SSIE as an additional input to apply a correction factor as proposed in Ge et al. (2005), for example. Effects of soil structure, aggregation, and heterogeneity will need to be accounted for by building separate models for ground soil and for soil in its natural state (e.g. Minasny et al., 2009a). In situ soil will more than likely be scanned in its natural state, so it might be more suitable to just concentrate on building calibrations using soil that has not undergone any physical alteration. However availability of this type of data is a possible limitation to this ideal approach.

Given the availability of more than one spectral library, an operator may be able to construct separate models based on either spectral library or may construct some form of *portmanteau* model where possible (Sankey et al., 2008; Minasny et al., 2009b). In any case, an important construct for the SSIE is a sub-process to assess whether a given spectral library is able to generate reliable predictions. In general, one may quickly assess this using principal component analysis. The principal components of the spectral library will fill a certain reduced or transformed spectral space. New spectra (from soil scanned in the field) will then be projected onto these principal components. If the new sample does not fall within the spectral space or hull of the spectral library, it will be indicative that the library and sample are unrelated, and therefore the models associated with that particular library may be unsuitable to apply. In any case, a useful feature of a SSIE, one that has not yet been addressed to date is to attribute quantifications of uncertainty to the soil contaminant predictions. Where sample spectra are not in the same 'spectral space' as the spectral library, higher levels of prediction uncertainty will be expected. Where coverage of soil samples by library occurs, and where multiple calibration models are in play, a logical decision would be to opt to use the model that performs best for the target soil property based on performance of external model validations. Empirical estimates of uncertainty (Tranter et al., 2009) may be applied based on the calibration models from spectral library estimates in order to derive upper and lower prediction limits (based on some level of confidence). The opportunity currently exists however for further investigation to examine and construct the specifics of how to quantify uncertainties of soil spectral predictions within SSIE or similar framework.

A suitable apparatus to take into the field, additional to the sensing devices, becomes important when identification and assessment of soil contaminants require measurement of the whole-soil profile support, as opposed to just the top soil horizons or some fixed depth. Even if measurement was required to go only to 1 m depth, some capacity for soil extraction or sensor probe insertion is necessary. The point-and-shoot device for PXRF is unsuitable for insertion down a soil core-hole unless a large enough pit is excavated, which is generally not desirable or allowable. Some retrofitting of an implement capable of attaching to the device and being inserted in to a hole may be necessary. Use of commercially manufactured Vis-NIR probes capable of whole-soil profile characterisation has previously been investigated (e.g. Lund, 2010). Their use is not widespread and limited to agricultural applications only. Retrofitting of existing handheld Vis-NIR sensors for down soil profile characterisation had also been demonstrated (Ben-Dor et al., 2008). Given the infancy of the technology for sub-soil characterisation with these PXRF and Vis-NIR devices, it is likely that in situ measurement of soil contamination will involve firstly a mechanical means of soil extraction such as either a percussion or pneumatic corer; followed by extraction of the soil core onto a suitable workbench. The soil core can then be scanned with either handheld instrument at depths determined suitable for the project at hand.

6. Spatial analysis for sampling, predicting and mapping contamination

The use of heuristic methods for sampling, predicting and mapping soil contamination is insufficient when environmental and health issues are at stake. Such methods are likely not to capture the complex spatial distribution of the contaminants, do not allow formal integration of relevant secondary information and do not provide an uncertainty measure associated with the estimates. In this context, geostatistics is a well-established scientific discipline that provides flexible spatial analysis methods to accurately delineate contaminated areas and to quantify uncertainties about the contaminant content (D'Or et al., 2009).

Tables 2 and 3 present examples of studies where geostatistics has been used to map soil contamination at spatial extents up to 5 km² which would encompass the range of areas that could be assessed for

Table 2
Case studies using geostatistics to map soil contamination with organics.

	Study	Country	Land cover	Depth	Sampling method	Analysis method	Distr.	N	Area (Km ²)	Mean	Var.	Skew	Median	Trans	Model	c0	c1	d1	c2	d2
2-Methylnaphthalene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	168	129,600	3	–	–	–	–	–	–	–	–
Acenaphthylene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	47	15,376	4	–	–	–	–	–	–	–	–
Anthracene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	17	3600	5	–	–	–	–	–	–	–	–
Benzo[a]anthracene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	4	225	5	–	–	–	–	–	–	–	–
Benzo[a]pyrene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	2	25	5	–	Log	Spherical	2	2	40	0	0
Benzo[b]fluoranthene + benzo[k]fluoranthene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	17	4225	6	–	–	–	–	–	–	–	–
Benzo[ghi]perylene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	7	256	4	–	–	–	–	–	–	–	–
Chrysene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	5	256	5	–	–	–	–	–	–	–	–
Fluoranthene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	19	6084	7	–	–	–	–	–	–	–	–
Fluorene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	25	4225	4	–	–	–	–	–	–	–	–
Ideno[1,2,3-cd]pyrene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	6	324	5	–	–	–	–	–	–	–	–
Naphthalene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	374	937,024	5	–	–	–	–	–	–	–	–
Phenanthrene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	50	21,316	5	–	–	–	–	–	–	–	–
Pyrene	Carlon et al. (2001)	Italy	Industrial	0–700	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	15	3136	6	–	–	–	–	–	–	–	–
TPAH	Carlon et al. (2001)	Italy	Industrial	0–400	Not specified	Gas chromatography–mass spectrometry extracted using acetone–ethane mixture	Highly skewed	71	3	755	2,643,876	4	–	Log	Spherical	0.21	1.1	89	0	0

Table 3

Case studies using geostatistics to map soil contamination with heavy metals.

Study	Country	Land cover	Depth	Sampling method	Analysis method	Distribution	N	Area (km ²)	Mean	Variance	Skew	Median	Transform	Model	c0	c1	d1	c2	d2
As Burgos et al. (2006)	Spain	Mine	0–15	Grid (20 × 50 m), 12 subplots (7 × 8 m)	ICP-OES Aqua Regia digestion	–	48	0.001	211	10,609.0	–	–	Log	Spherical	0.028	0.06	18.8	0	0
Ferreira da Silva et al. (2004)	Portugal	Mine	0–15	Grid (100 × 100 m)	ICP-ES Aqua Regia	Lognormal	106	1.4	820	1,610,361.0	2.8	273	None	Spherical	47,844	2,303,588	750	0	241
Ferreira da Silva et al. (2004)	Portugal	Mine	0–15	Grid (100 × 100 m)	ICP-ES Aqua Regia	Lognormal	106	1.4	820	1,610,361.0	2.8	273	Box–Cox	Spherical	0.03	0.39	1500	0	480
Yang et al. (2009)	China	Agriculture urban	0–20	Irregular grid, with 5 subsamples at each point	GFAAS (graphite furnace atomic absorption spectroscopy) Aqua Regia digestion	Normal	100	0.0004	6.16	2.25	–0.116	–	None	Spherical	0.66	2.35	11.63	0	0
Bourennane et al. (2006)	France	Agriculture wastewater irrigation plane	0–20	Square grid	Not specified Aqua Regia digestion	–	50	0.15	3.98	4.29	1.53	–	–	Spherical	1.37	2.00	80.00	0	0
Burgos et al. (2006)	Spain	Mine	0–15	Grid (20 × 50 m), 12 subplots (7 × 8 m)	ICP-OES Aqua Regia	–	48	0.001	4.44	1.35	–	–	None	Linear	0.717	1.77	20.2	0	0
Yang et al. (2009)	China	Agriculture urban	0–20	Irregular grid, with 5 subsamples at each point	GFAAS (graphite furnace atomic absorption spectroscopy) Aqua Regia digestion	Lognormal	100	0.0004	0.15	0.002	0.488	–	Log	Spherical	0.0007	0.0016	3.28	0	0
Zupan et al. (2000)	Slovenia	Industrial, forest	0–5	Systematic sampling design, two grids—one general and other lowland where main sources of pollution are	AAS ethylenediamine-tetraacetic acid (EDTA)	Lognormal	119	5	2.1	0.000	–	0.9	–	Spherical	0.76	1.72	9.5	0	0
Zupan et al. (2000)	Slovenia	Industrial forest	0–5	Systematic sampling design, two grids—one general and other lowland where main sources of pollution are	AAS Aqua Regia digestion	Lognormal	119	5	2.5	15.1	–	1.1	–	Spherical	0.3	2.18	9.50	0	0

(continued on next page)

Table 3 (continued)

Study	Country	Land cover	Depth	Sampling method	Analysis method	Distribution	N	Area (km ²)	Mean	Variance	Skew	Median	Transform	Model	c0	c1	d1	c2	d2	
Cr	Assadian et al. 1998 (Mexico)	Mexico	Agriculture (alfalfa)	0–30	Parallel transects along canal	ICP nitric acid and hydrogen peroxide digestion	–	79	0.018	10.00	22.1	–	–	None	Linear	0.01	16.6	160	0.00	0.00
	Assadian et al. 1998 (Mexico)	U.S.A.	Agriculture (alfalfa)	0–30	Parallel transects along canal	ICP, nitric acid and hydrogen peroxide digestion	–	55	0.036	14.60	11.6	–	–	None	Linear	0.51	1.46	160.00	0.00	0.00
	Bourennane et al. (2006)	France	Agriculture wastewater irrigation plane	0–20	Square grid	Not specified Aqua Regia digestion	–	50	0.15	76.03	694.32	0.94	–	Unspecified	Spherical	312	247	80.00	0	0
	Yang et al. (2009)	China	Agriculture urban	0–20	Irregular grid, with 5 subsamples at each point	Not specified Aqua Regia digestion	Lognormal	100	0.0004	57.77	76.91	0.762	–	Log	Spherical	39.3	77	7.76	0	0
	Bourennane et al. (2006)	France	Agriculture-wastewater irrigation plane	0–20	Square grid	Not specified Aqua Regia digestion	–	50	0.15	173.40	4188.7	0.43	–	–	Spherical	1774	1888.00	105.00	0	0
	Burgos et al. (2006)	Spain	Mine	0–15	Grid (20 × 50 m), 12 subplots (7 × 8 m)	ICP-OES, Aqua Regia	–	48	0.001	119.00	707.6	–	–	Log	Linear	0.004	0.012	21.2	0	0
	Yang et al. (2009)	China	Agriculture urban	0–20	Irregular grid, with 5 subsamples at each point	FAAS, Aqua Regia digestion	Normal	100	0.0004	21.22	11.7	0.352	–	None	Spherical	6.86	11.5	7.37	0	0
	Zupan et al. (2000)	Slovenia	Industrial forest	0–5	Systematic sampling design, two grids—one general and other lowland where main sources of pollution are	AAS ethylenediamine-tetraacetic acid (EDTA)	Lognormal	119	5	8.6	54.8	–	6.2	–	Exponential	0.46	0.74	4.7	0	0
	Zupan et al. (2000)	Slovenia	Industrial forest	0–5	Systematic sampling design, two grids—one general and other lowland where main sources of pollution are	AAS Aqua Regia digestion	Lognormal	119	5	24.8	240.2	–	20.9	–	Exponential	0.24	0.36	3.30	0	0

Pb	Assadian et al. 1998 (Mexico)	Mexico	Agriculture (alfalfa)	0–30	Parallel transects along canal	ICP nitric acid and hydrogen peroxide digestion	–	79	0.018	6.50	39.7	–	–	ln	Linear	0.0001	0.26	160	0	0
	Assadian et al. 1998 (Mexico)	U.S.A.	Agriculture (alfalfa)	0–30	Parallel transects along canal	ICP nitric acid and hydrogen peroxide digestion	–	55	0.036	9.00	1.96	–	–	ln	Linear	0.0001	0.049	160	0	0
	Bourennane et al. (2006)	France	Agriculture wastewater irrigation plane	0–20	Square grid	Not specified Aqua Regia digestion	–	50	0.15	321.58	17,223.9	0.64	–	–	Spherical	4176	10,555.00	100.00	0	0
	Burgos et al. (2006)	Spain	Mine	0–15	Grid (20 × 50 m), 12 subplots (7 × 8 m)	ICP-OES Aqua Regia	–	48	0.001	471.00	46,656.0	–	–	Log	Linear	0.026	0.047	18.6	0	0
	Ferreira da Silva et al. (2004)	Portugal	Mine	0–15	Grid (100 × 100 m)	ICP-ES Aqua Regia	Lognormal	106	1.4	403.00	602,176.0	5.2	173	None	Spherical	151,161	503,871	400	0	121
	Ferreira da Silva et al. (2004)	Portugal	Mine	0–15	Grid (100 × 100 m)	ICP-ES Aqua Regia	Lognormal	106	1.4	403.00	602,176.0	5.2	173	Box–Cox	Spherical	0.15	0.5	1000	0	300
	Yang et al. (2009)	China	Agriculture urban	0–20	Irregular grid, with 5 subsamples at each point	GFAAS (graphite furnace atomic absorption spectroscopy) Aqua Regia digestion	Lognormal	100	0.0004	18.8	15.4	0.79	–	Log	Spherical	8.48	15.41	3.42	0	0
Hg	Yang et al. (2009)	China	Agriculture urban	0–20	Irregular grid, with 5 subsamples at each point	Atomic fluorometry	Lognormal	100	0.0004	0.08	0.004	0.692	–	Log	Spherical	0.0021	0.0035	9.31	0	0
Ni	Assadian et al. 1998 (Mexico)	Mexico	Agriculture (alfalfa)	0–30	Parallel transects along canal	ICP nitric acid and hydrogen peroxide digestion	–	79	0.018	7.00	10.9	–	–	None	Linear	0.01	8.21	160.00	0.00	0.00
	Assadian et al. 1998 (Mexico)	U.S.A.	Agriculture (alfalfa)	0–30	Parallel transects along canal	ICP nitric acid and hydrogen peroxide digestion	–	55	0.036	11.50	6.25	–	–	None	Spherical	0.12	0.37	100.00	0.00	0.00
	Bourennane et al. (2006)	France	Agriculture wastewater irrigation plane	0–20	Square grid	Not specified, Aqua Regia digestion	–	50	0.15	18.97	22.2	0.47	–	–	Spherical	9.41	10.5	80.00	0	0
	Yang et al. (2009)	China	Agriculture urban	0–20	Irregular grid, with 5 subsamples at each point	FAAS Aqua Regia digestion	Normal	100	0.0004	25.04	21.1	–0.185	–	None	Linear	10.71	21.00	10.08	0	0

(continued on next page)

Table 3 (continued)

Study	Country	Land cover	Depth	Sampling method	Analysis method	Distribution	N	Area (km ²)	Mean	Variance	Skew	Median	Transform	Model	c0	c1	d1	c2	d2
Zn Assadian et al. 1998 (Mexico)	Mexico	Agriculture (alfalfa)	0–30	Parallel transects along canal	ICP nitric acid and hydrogen peroxide digestion	–	79	0.018	28.70	392.0	–	–	ln	Linear	0.0001	0.173	160	0	0
Assadian et al. 1998 (Mexico)	U.S.A.	Agriculture (alfalfa)	0–30	Parallel transects along canal	ICP nitric acid and hydrogen peroxide digestion	–	55	0.036	40.20	70.6	–	–	ln	Spherical	0.006	0.01	80	0	0
Atteia et al. (1994)	Switzerland	agriculture	0–25	Square grid and nesting	Direct current plasma spectrometry nitric acid digestion	Lognormal	366	14.5	78.50	1482.1	2.74	74	Log	Double spherical	0.00	0.02	316.00	0.01	1437
Bourennane et al. (2006)	France	Agriculture wastewater irrigation plane	0–20	Square grid	Not specified Aqua Regia digestion	–	50	0.15	684.07	54,321.6	0.55	–	–	Spherical	17,378	27,702	90.00	0	0
Burgos et al. (2006)	Spain	Mine	0–15	Grid (20 × 50 m), 12 subplots (7 × 8 m)	ICP-OES Aqua Regia	–	48	0.001	381.00	18,496.0	–	–	Log	Linear	0.019	0.025	16.6	0	0
Yang et al. (2009)	China	Agriculture urban	0–20	Irregular grid, with 5 subsamples at each point	FAAS Aqua Regia digestion	Normal	100	0.0004	69.96	51.7	0.073	–	None	Gaussian	37.14	51.8	8.92	0	0
Zupan et al. (2000)	Slovenia	Industrial forest	0–5	Systematic sampling design, two grids—one general and other lowland where main sources of pollution are	AAS ethylenediamine-tetraacetic acid (EDTA)	Lognormal	119	5	113	40,925.3	–	41	–	Exponential	0.13	2.57	7.4	0	0
Zupan et al. (2000)	Slovenia	Industrial forest	0–5	Systematic sampling design, two grids—one general and other lowland where main sources of pollution are	AAS Aqua Regia digestion	Lognormal	119	5	337	216,876.5	–	165	–	Exponential	0.016	1.6	4.50	0	0

contamination, the upper end being large urban developments. Table 2 presents results for two variograms for organics based on one study with two different contaminants. Table 3 presents results for 36 variograms based on six contaminants (As, Cr, Pb, Hg, Ni, Zn).

The aim of many of these papers was generally one or more of the following: the use of geostatistical techniques to locate additional sampling locations, to provide a map delineating contaminated areas or to provide the probability of exceeding a critical threshold (for this review, we focus on case studies including heavy metals and organics). Hence geostatistical contamination assessment relies on reproducing the distribution of the data available to provide an estimate of the contamination concentration at locations between samples which is further complemented with a measure of uncertainty.

In this context, choosing the appropriate sampling design is foremost important. Most of the studies in Tables 2 and 3 used a systematic design (grid-based or transect) but one problem with such designs is that they do not use other information sources to improve design. For example, the use of finer grids where contamination may be expected and a small subset of close samples is useful to improve estimates of the nugget semivariance (C_0). Furthermore, the number of sampling locations has to ensure data representativeness needed for spatial analysis.

Contaminant data are commonly described by highly skewed distributions which are more challenging to map and analyse. This is shown for many of the studies in Tables 2 and 3 where the data is skewed (>0) and/or follows a log-normal distribution. Often higher values cannot be discarded as outliers or measurement errors since it might indicate a hotspot. Besides locating the hotspot, spatial analysis can be used to delineate its extent both in area and in depth. Hence the geostatistical method used should avoid underestimating higher values (smoothing effect) and account for their influence in predicting contamination spread. This might happen when using a log or other similar transformations prior to geostatistical analysis. Most of studies in Tables 2 and 3 used this approach.

An alternative approach is to use robust geostatistics (Lark, 2000, 2002; Saby et al., 2006; Marchant et al., 2011b) or to use untransformed data directly to predict and map contamination. Preferable prediction methods for untransformed data include non-linear kriging techniques such as indicator kriging (Journel, 1983). The indicator approach implies a non-linear transformation of the data: each observation is transformed into a set of k indicator values corresponding to k threshold values. Indicator kriging is then applied to the set of indicators and estimated values are assembled to form a conditional cumulative distribution function (ccdf) for each location (Goovaerts, 1997). The ccdf provides a model of local uncertainty for the unsampled locations. Besides the advantages of using indicator kriging to provide contamination maps and evaluate misclassification costs (Cattle et al., 2002), the method also allows the incorporation of soft information (Goovaerts et al., 1997; Brus et al., 2002; Saito and Goovaerts, 2002). Other applications include providing additional sampling locations or optimised sampling designs to improve contamination assessment (van Groenigen et al., 2000; Barabás et al., 2001; Van Meirvenne and Goovaerts, 2001; Chu et al., 2010).

For decision-making, a critical question is to ascertain which level of risk or probability threshold determines remediating a polluted area. Difficulties arise when dealing with intermediate probabilities and one may wish to minimise misclassification especially a false negative situation when a hazardous area may be declared safe. To improve reliability in delineating contaminated areas, another option is to use a geostatistical simulation approach. This approach is used to generate equally likely realizations of the attribute from their joint distribution taking into account the spatial variation of observed data at sampled locations. Also, these sets of realizations can be used as inputs to another model for further prediction (for example, a risk assessment model) providing an uncertainty measure for the model responses (Goovaerts, 1997). Additionally, the post-processing of realizations

provides a measure of spatial (or multi-location) uncertainty or the probability that a given threshold is exceeded jointly at several locations (Goovaerts, 1997). According to Goovaerts (2001), the simulation-based approach has several advantages over kriging such as providing a model of spatial uncertainty (or the probability that a given threshold is exceeded jointly at several locations) and a set of realizations (i.e. equiprobable simulated images) to study the propagation of uncertainty through GIS operations or transfer functions. Also, the ccdf for supports larger than the measurement support can be numerically approximated by the cumulative distribution of block simulated values that are obtained by averaging values that have been simulated (Journel, 1992; Kyriakidis, 1997).

Simulation methods differ depending on the approach chosen to model the probability distribution. The most commonly used approaches are Sequential Gaussian Simulation (SGS) (Almeida and Journel, 1994) (multiGaussian approach) and Sequential Indicator Simulation (SIS) (Goovaerts, 1997) (indicator approach). Examples of applications using these simulation methods are presented by Hooker and Nathanail (2006) and Zhang et al. (2009) (SGS); and by Juang et al. (2004), Gay and Korre (2006) and Lin et al. (2011) (SIS). Also Demougeot-Renard et al. (2004) and Verstraete and Van Meirvenne (2008) used a Gaussian simulation approach to refine sampling strategies and improve assessment accuracy. Another simulation method is applied by Franco et al. (2006) to build a hazard map based on the joint dispersion of heavy metals. The predictions are provided by direct sequential simulation (Soares, 2001) which does not require data transformation as in SGS and SIS.

Recently, model-based geostatistics has also been used to predict and map soil contamination. Within the linear mixed model framework, the spatial variation of the contaminant is quantified by fixed and random effects. The fixed-effects component describes the relationship between the expected concentration of the contaminant and covariates. These covariates are regarded as potential predictors of the contaminant and can thus incorporate physical knowledge about contaminant dispersion in the environment. The random-effects component describes spatially-correlated random variation, i.e., the variation not fully explained by the physical trend model incorporated in the fixed effects. Spatial prediction is then carried using the empirical best linear unbiased predictor (E-BLUP) based on the linear mixed model (Lark, 2012). This formulation allows the estimation of the variogram parameters using the residual maximum likelihood (REML) estimator (Lark et al., 2006) that will minimise the bias in the estimates (Marchant and Lark, 2007). Villanneau et al. (2011) used this approach to determine the spatial pattern of persistent organic pollutant variation.

The spatial copula method has been proposed as an alternative solution to deal with non-Gaussian soil data. Copulas describe the dependence structure of the multivariate variable separate from its marginal distributions. The copula method was adopted by Marchant et al. (2011a) for spatial prediction of soil Cd concentration in France; in this case the Gaussian marginal distribution was replaced by a generalised extreme value (GEV) distribution while the dependence structure model was still Gaussian. This model-based method allows for the prediction of the whole distribution of the property conditional on observed data at unobserved sites. This conditional distribution can be used to ascertain if the probability that the concentration of Cd in the soil at a particular site exceeds a regulatory threshold or the 95% confidence limit of soil Cd concentration at a site.

It is evident that spatial analysis methods have been progressing to improve sampling designs and increase prediction accuracy. As mentioned before (Section 2) real-time sampling supported by proximal sensing will also contribute to a better performance of spatial models in terms of their accuracy. Moreover, sampling using field spectroscopic methods (Vis-NIR and PXRF) can provide real-time measurements of contamination concentrations to support on-the-go spatial prediction and mapping thus directing sampling efforts to areas with higher uncertainty.

Within the geostatistical framework it is possible to integrate Vis-NIR and PXRF measurements. Odlare et al. (2005) and Cobo et al. (2010) used variograms produced with near- and mid-infrared reflectance spectroscopy data to study spatial variation in soil properties. More advanced applications to assess soil contamination using proximal sensing data can be provided using co-kriging or co-simulation. These geostatistical methods allow the incorporation of secondary data to improve predictions provided that the variables are spatially correlated. Liu et al. (2006) and Sollitto et al. (2010) presented an application of co-kriging to map heavy metal contamination.

At the time of writing no scientific papers have presented results on the use of combined IR and PXRF as secondary data to improve sampling in contaminated sites or to predict contamination.

Regardless of the method used for contamination mapping it is always necessary to validate model performance and prediction accuracy. Goovaerts (2001) performs a cross-validation to assess the goodness and precision of models of local uncertainty (using, for example, accuracy plots). This approach is followed by Barabás et al. (2001) and Van Meirvenne and Goovaerts (2001). Validation of spatial uncertainty as provided by simulation is not as straightforward. Ensuring the reproduction of the histogram and variogram might not be enough to establish the goodness of the predictions (Goovaerts, 2001).

Recently the squared standardized prediction error (SSPE) was proposed to validate the spatial model (Lark, 2009). This validates the prediction variance rather than the prediction value as performed by other measures such as root-mean-square error and mean error.

Finally, at one particular site there could be multiple contaminants being mapped which have different levels of spatial variability. For example, for each individual study where multiple contaminants were mapped, the semivariogram parameters (especially the very short-range and structural variances – C_0 and C_1) were quite different (i.e. there was no relationship between the mean and variance) meaning that given the same sample size and configuration, the final maps for each contaminant would have different levels of precision. Essentially, this means that for one site we may not have one design that is optimal for all contaminants.

7. Prospective methodology for soil contamination assessment supported by field spectroscopy and advanced spatial analysis

Decision-based models for soil contamination assessment can be optimised by improving the quality and quantity of data. Our review confirms that improved sampling design aided by portable field spectrometers and advanced spatial analysis can be combined into a new approach for soil contamination assessment. The motivation for this new approach is to provide more and accurate data to minimise lack of statistical representativeness, increase the spatial quantification, reduce the time needed to evaluate the presence and extent of contamination, add focus to the sampling plan, and overall to reduce the project costs. The suggested steps to build this approach are pictured on the right-hand side of Fig. 1.

Real-time sampling supported by field Vis-IR and PXRF can be a part of an adaptive sampling design whereby a first stage will aim to delineate contaminated areas and a second stage will focus in areas where the degree of contamination appears to be higher. A possible third stage would include the collection of samples to validate Vis-IR and PXRF measurements or to choose the locations where it is necessary to extract soil cores to verify contamination in the soil profile.

Vis-IR and PXRF measurements will provide estimates for contaminant concentration. But how can we assess the quality of Vis-IR and PXRF measurements to thereby provide reliable estimates? And will, or should, these measurements be processed separately or combined to provide more accurate estimates?

7.1. The synergistic use of Vis-IR and PXRF – data fusion

Often in applied science, practitioners will favour one instrument or method over another, largely through individual experience or local factors. Ab initio since both field Vis-NIR and PXRF produce estimates of a wide range of contaminants and these estimates are probably less precise than conventional laboratory-based assays, it seems prudent to combine the estimates produced from the two methods rather than favour one or the other. Indeed, this idea of gathering information from various sources seems to be one of the major propositions of chemometrics (Kowalski, 1975; Lavine, 2005).

Considering the field situation, we have the possibility of using both instruments (Vis-NIR and PXRF) on the same sample (volume of soil) within a small number of minutes. Both methods are affected by soil moisture but there are potentially correction methods for both (Section 5).

In this mode, this is an example of field proximal soil sensing originally suggested by Viscarra Rossel and McBratney (1998), discussed in detail in Viscarra Rossel et al. (2010) and motivated by the need for high-resolution spatial and temporal soil information. This demand usually occurs in high-value operations where soil is an integral component of management, and as such contaminated site assessment and remediation are a prime example.

Joint use of the two instruments with a priori calibration permits to a large degree the adaptive choice of sampling locations in the field. The estimates from the current and previous sites will suggest the location of a subsequent site.

Sampling would begin either by expert knowledge (preliminary stage of the investigation), or the use of external covariates such as air photographs, airborne NIR, or mobile field proximal soil sensing using, for example, electromagnetic induction or gamma radiometry to allow stratification (Miklos et al., 2010). If no prior information is available some kind of equal-area stratification could be used (Walvoort et al., 2010).

At a particular site, Vis-NIR and PXRF spectra at each location will be the subject of prior calibration. One approach may be to concatenate the spectra from the two instruments and to calibrate based on the sequentially joined spectra. Viscarra Rossel et al. (2006) showed that concatenated UV, visible, NIR, and MIR spectra produced better predictions than any single part of the spectral range. However the NIR wavelengths are far distant from those of XRF in the electromagnetic spectrum, so simple concatenation seems less natural.

Another approach is to combine or fuse the predictions produced from the two methods. One analytical method may be more precise, for say heavy metals (PXRF), and the other for organic compounds (Vis-NIR). This however enhances the advantage of conjunctive use in the field – a wide range of potential contaminants can be detected at every site. Such an approach could be styled sensor fusion (Crowley and Demazeau, 1993), i.e., the combination of sensory data or data derived from various sources such that the resulting information is in some sense better than would be possible were these sources are used individually. In this case, better can mean more complete, dependable or accurate. Alternatively, this process can be regarded as a form of enhanced soil inference (McBratney et al., 2002, 2006) with the calibrations for the various contaminants for the two instruments being considered as spectral pedotransfer functions.

Formal methods that can be used to process information from two sensors each with associated uncertainty are the so-called model-averaging procedures. Vis-IR and PXRF predictions could be considered as an ensemble of outcomes (here, two) which one wishes to combine into a single outcome. This type of situation is common in atmospheric, hydrologic and econometric research (Bates and Granger, 1969; Wagener and Gupta, 2005) where multiple forecasts of a given process are derived from a number of competing predictive instruments (in those cases they are actually models). Each contributor model will have its own strengths and weaknesses. Rather than selecting the single

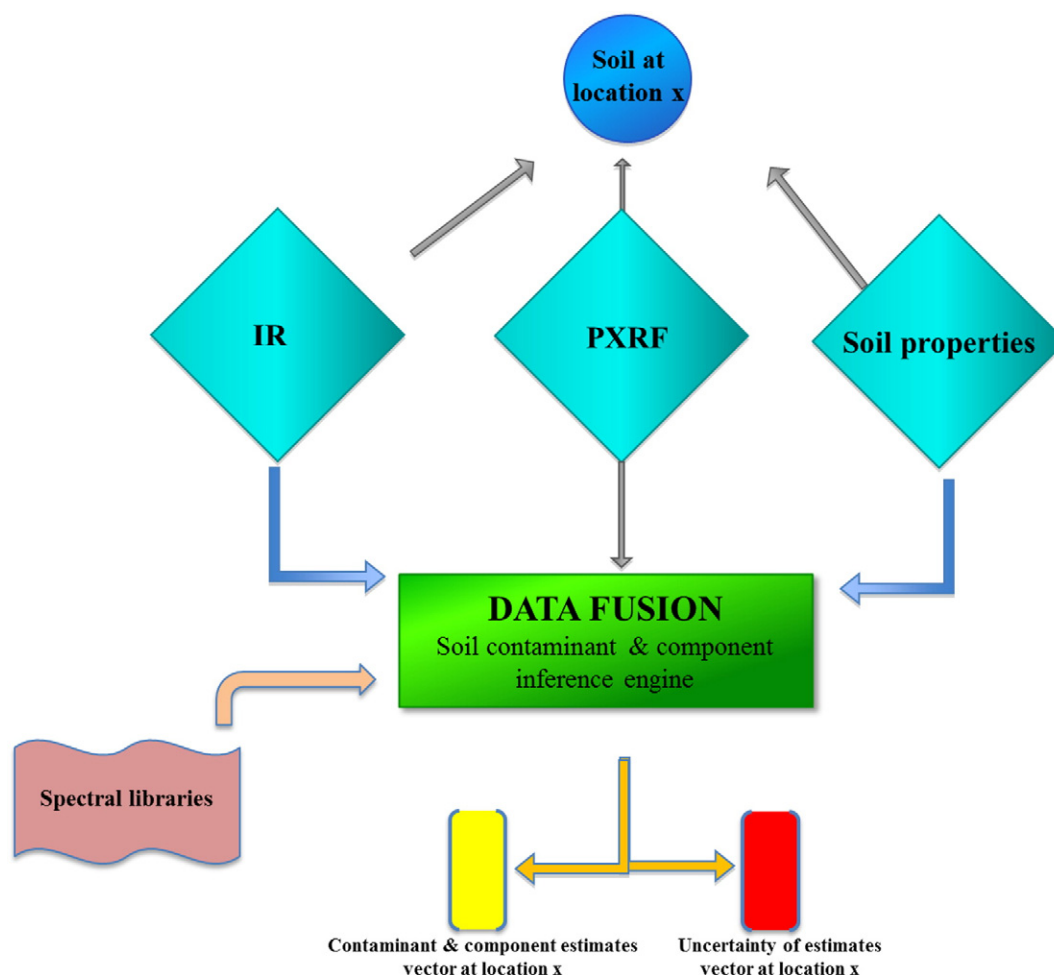


Fig. 8. In-field inference engine scheme.

best-performing model for a given situation or scenario (which is the traditional approach), combining model outcomes is a natural generalisation of this (Diks and Vrugt, 2010). Ideally, the new combined outcome is at least as good as any of the individual outcomes.

Combining different model outcomes in this way is termed model ensemble or averaging (Rojas et al., 2008). Diks and Vrugt (2010) thoroughly described, applied, and compared a number of different model-averaging approaches with reference to point forecasting for hydrologic modelling applications. In the current context there will be two contributor models, which are the Vis-IR and PXRF spectral calibration equations for a given contaminant. In some cases this number might be increased by including a set of a priori calibration equations from each instrument for a particular contaminant. Each calibration equation has a weighting attributed to it. For most model-averaging approaches, the weights from the competing calibration equations sum to unity. Given this, the difference between model-averaging methods depends on how the weights are estimated. The simplest option is to apply equal weighting to the two methods. This will generally be undesirable because predicted contaminant concentrations will be unlikely to be equally certain from the Vis-NIR and PXRF spectra.

A better choice is that proposed by Bates and Granger (1969), which is to weight each predicted contaminant concentration in proportion to the reciprocal of its associated variance, and therefore calibration equations must provide this variance, i.e., calibration equations alone are not enough; they must also have a methodology for providing uncertainty estimates (Tranter et al., 2010; McBratney et al., 2011b). Heuvelink and Bierkens (1992) provide probably the first demonstration of model averaging in soil science using the inverse variance approach

for combining soil map predictions. Other model-averaging approaches include information criterion averaging (Buckland et al., 1997), which is less computationally demanding than the Bayesian and Mallows model-averaging methods described in Hoeting et al. (1999) and Hjort and Claeskens (2003) respectively. Interestingly, a far simpler but perhaps equally efficient model-averaging approach, used in Diks and Vrugt (2010), is the Granger–Ramanathan (GM) averaging (Granger and Ramanathan, 1984). In this approach the weights (which do not necessarily sum to unity) are solved using ordinary-least-squares regression, where the predictor variables are the different predicted concentrations, and the target variable is the associated actual observations. Some actual laboratory assay will be required although weights from a prior survey may be valid. Malone et al. (2014) have tested these various approaches for soil map prediction of soil carbon and showed the GM estimator to be a robust one.

Using such fusion or inferencing approaches, it should be possible to estimate a range of contaminants using combined real-time Vis-IR and PXRF measurements.

Finally, these estimates can be brought together via a data fusion framework called a soil inference engine (McBratney et al., 2002) allowing for contaminants to be estimated simultaneously, at different depths, and with known uncertainty.

7.2. In-field soil inference engine

The “inference engine” (Fig. 8) consists of a set of algorithms to process spectral data into a contamination estimate and also an estimate of other soil properties that ultimately are needed to develop a

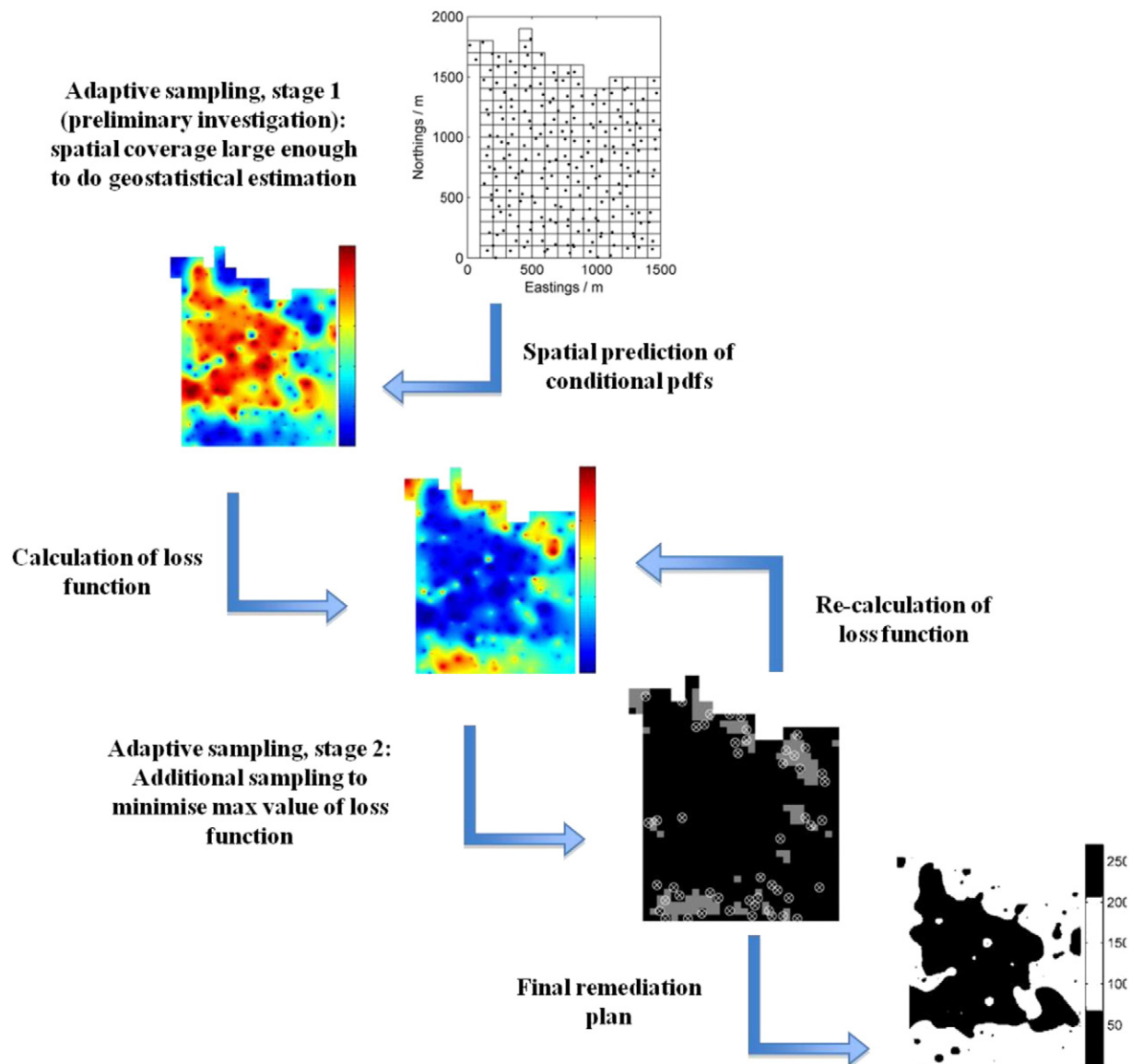


Fig. 9. Field sampling and spatial prediction scheme.

remediation plan (such as clay and organic matter content). The spectral measurements are acquired on-site in pre-defined sampling locations. Contamination estimates are then used in a spatial prediction model to map contamination degree and extent.

7.3. Spatial analysis – adaptive sampling and mapping

Initially, Vis-IR and PXRF provide a data set of point estimates for the contaminant concentrations. However dense, this dataset needs to be treated to define homogeneous contamination areas (along with an estimate of the associated uncertainty) to further sustain a remediation plan.

The proposed spatial analysis framework (Fig. 9) will provide at each location an estimate of the (posterior) probability density function (pdf) of contamination concentration. The initial sampling locations are chosen to ensure the necessary spatial coverage needed for the spatial model.

With the ability to make estimates in real-time in the field, it should be possible to adapt the sampling to the next location which maximises the reduction in uncertainty integrated over the study area (Marchant et al., 2013), and so on, until some overall quality criterion or financial constraint has been reached.

Hence, based on the contamination pdf, a loss function can be applied to evaluate the prediction uncertainty. This loss function incorporates distinctive issues such as the uncertainty associated with the estimate of the contamination, prediction variance associated with mapping, costs considering equivalent laboratory analysis and remediation costs. Based on the uncertainty output, it is possible to evaluate if, and where, more field spectral data should be acquired. This will allow optimised target sampling based on the uncertainty of prediction and location of hotspots.

We believe that this approach can provide comprehensive information on the type, extent and level of contamination and assess the adequacy and completeness of all information available to be used in making decisions on remediation. Ideally this would be implemented as on-the-go software to allow on-site real-time interpretation.

To summarise, given the advances in spatial analysis and field-deployable sensing it is possible to build a more efficient methodology for soil contamination assessment based on adaptive sampling supported by on-the-go Vis-NIR and PXRF measurements. Data fusion of Vis-IR and PXRF measurements is achievable and accurate measurements can be provided (with spectral libraries and field validation). Adaptive sampling can be combined with spatial analysis to map contaminants and evaluate the extent of contamination using an optimisation technique

that accounts for the environmental/health contaminant levels and the associated costs (a loss function). A conceptual framework is proposed here to be used on-site to provide near real-time contamination assessments. This is achieved by implementing an inference engine that deals with the uncertainty of the contamination estimates and predictions. It also has the potential to deal with multiple contaminants and to provide 3D maps. The result will be an optimised map of contaminated regions within a site with a limited requirement of laboratory assay. Prospectively, the proposed approach could be implemented as on-the-go field deployable software to provide a reliable and fast initial on-site contamination assessment.

8. Conclusions

- a. There is a large number of contaminated soil sites worldwide (order of magnitude 10^7) and the rate of remediation is relatively meagre, perhaps of the order of 0.5% per annum. The true rate of creation of contaminated sites is largely unknown (but could be of the same order as the remediation rate).
- b. Soil contaminants at any site potentially include a wide range of natural and synthetic inorganic and organic compounds and minerals thus making analytical costs very large. This presents a financial impediment to discovery and remediation therefore increasing public health and environmental risks.
- c. Laboratory Vis-IR shows promise for assessing a range of key soil properties including texture, CEC, pH, total C, carbonates, and particularly organic compounds and some efficacy for metals and minerals.
- d. PXRF is suitable for a range of metals (atomic mass > 12) and can also be calibrated for a range of soil properties, particularly clay content.
- e. Field-deployable portable instruments for Vis-NIR and XRF are now widely available. Field deployment of Vis-NIR and XRF will require special calibration approaches particularly to deal with soil moisture. Special sampling geometries and probes may need to be developed to observe potentially contaminated soil at depth.
- f. Soil contaminants generally show highly skewed distributions and a large degree of spatial variation across sites making their spatial analysis and prediction problematic, especially based on the inevitably small number of observations due to large analytical costs.
- g. There is a strong possibility that a new field soil proximal sensing modality namely conjoint use of Vis-NIR and PXRF can enhance the overall contaminated-site assessment approach.
- h. In addition to further work on calibration of the two methods for a wide range of contaminants singly or jointly, work is required especially to calibrate under field conditions, particularly considering soil moisture, and to devise inferencing procedures and appropriately optimised field sampling strategies.

Acknowledgments

The authors acknowledge Mr Phil Mulvey (Chief Executive Officer and Senior Principal and Hydrogeologist at the Centre for Contaminant Geoscience Pty Ltd – Environmental Earth Sciences International) for providing up-to-date soil contamination and remediation data. Author A. Horta was funded by the Richard Claude Mankin scholarship provided by the University of Sydney.

References

Alloway, B.J. (Ed.), 1995. *Heavy Metals in Soils*, 3rd edition Springer, Netherlands <http://dx.doi.org/10.1007/978-94-007-4470-7>.

Almeida, A.S., Journel, A.G., 1994. Joint simulation of multiple variables with a Markov-type coregionalization model. *Math. Geol.* 26 (5), 565–588.

Assadian, N.W., Esparza, L.C., Fenn, L.B., Ali, A.S., Miyamoto, S., Figueroa, U.V., Warrick, A.W., 1998. Spatial variability of heavy metals in irrigated alfalfa fields in the upper Rio Grande River basin. *Agric. Water Manag.* 36 (2), 141–156.

Atteia, O., Dubois, J.P., Webster, R., 1994. Geostatistical analysis of soil contamination in the Swiss Jura. *Environ. Pollut.* 86 (3), 315–327.

Barabás, N., Goovaerts, P., Adriaens, P., 2001. Geostatistical assessment and validation of uncertainty for three-dimensional dioxin data from sediments in an estuarine river. *Environ. Sci. Technol.* 35 (16), 3294–3301.

Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. *Oper. Res. Q.* 20, 451–468.

Baumgardner, M.F., Silva, L.F., Biehl, L.L., Stoner, E.R., 1985. Reflectance properties of soils. *Adv. Agron.* 38, 1–44.

Becker, R., Koch, K., Wachholz, S., Win, T., 2002. Quantification of total petrol hydrocarbons (TPH) in soil by IR-spectrometry and gas chromatography – conclusions from three proficiency testing rounds. *Accred. Qual. Assur.* 7, 286–289.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., 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.

Ben-Dor, E., Heller, D., Chudnovsky, A., 2008. A novel method of classifying soil profiles in the field using optical means. *Soil Sci. Soc. Am. J.* 72 (4), 1113–1123.

Ben-Dor, E., Eshel, G., Schwartz, G., 2014. Quantitative assessment of soil contaminants, particularly hydrocarbons, using reflectance spectroscopy. U.S. Provisional Patent Application No. 61/659,494.

Bhat, M.M., Kumar, D., Narain, K., Shukla, R., Yunus, M., 2011. Mycoremediation of hydrocarbon contaminated soil-FTIR based analysis. *World Appl. Sci. J.* 15, 1547–1552.

Bierkens, M.F.P., 1997. Using stratification and residual kriging to map soil pollution in urban areas. In: Baafi, E.Y., Schofield, N.A. (Eds.), *Geostatistics Wollongong '96*. vol. 2. Kluwer Academic Publishers, The Netherlands, pp. 996–1007.

Blake, T.A., Gassman, P.L., 2001. Detection of Soil Surface Contaminants by Infrared Reflection Spectroscopy. , pp. 239–261.

Bogrekcı, I., Lee, W.S., 2005. Spectral phosphorus mapping using diffuse reflectance of soils and grass. *Biosyst. Eng.* 91, 305–312.

Bourennane, H., Dère, C., Lamy, I., Cornu, S., Baize, D., van Oort, F., King, D., 2006. Enhancing spatial estimates of metal pollutants in raw wastewater irrigated fields using a topsoil organic carbon map predicted from aerial photography. *Sci. Total Environ.* 361 (1–3), 229–248.

Bray, J.G.P., Rossel, R.V., McBratney, A.B., 2009. Diagnostic screening of urban soil contaminants using diffuse reflectance spectroscopy. *Aust. J. Soil Res.* 47 (4), 433–442.

Brevik, E., Burgess, L. (Eds.), 2012. *Soils and Human Health*. CRC Press, Boca Raton.

Brownlow, A.H., 1979. *Geochemistry*. Prentice-Hall N.J.

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. *J. Environ. Qual.* 31 (6), 1875–1884.

Buckland, S.T., Burnham, K.P., Augustin, N.H., 1997. Model selection: an integral part of inference. *Biometrics* 53, 603–618.

Buerck, J., Roth, S., Kraemer, K., Scholz, M., Klaas, N., 2001. Application of a fiber-optic NIR-EFA sensor system for in situ monitoring of aromatic hydrocarbons in contaminated groundwater. *J. Hazard. Mater.* 83 (1–2), 11–28.

Burgos, P., Madejón, E., Pérez-de-Mora, A., Cabrera, F., 2006. Spatial variability of the chemical characteristics of a trace-element-contaminated soil before and after remediation. *Geoderma* 130 (1–2), 157–175.

Burrough, P.A., 2001. GIS and geostatistics: essential partners for spatial analysis. *Environ. Ecol. Stat.* 8, 361–377.

Cai, Q.-Y., Mo, C.-H., Wu, Q.-T., Katsoyiannis, A., Zeng, Q.-Y., 2008. The status of soil contamination by semivolatile organic chemicals (SVOCs) in China: a review. *Sci. Total Environ.* 389, 209–224.

Carlson, C., Critto, A., Marcomini, A., Nathanail, P., 2001. Risk based characterisation of contaminated industrial site using multivariate and geostatistical tools. *Environ. Pollut.* 111 (3), 417–427.

Carr, R., Zhang, C., Moles, N., Harder, M., 2008. Identification and mapping of heavy metal pollution in soils of a sports ground in Galway City, Ireland, using a portable XRF analyser and GIS. *Environ. Geochem. Health* 30 (1), 45–52.

Cattle, J.A., McBratney, A.B., Minasny, B., 2002. Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. *J. Environ. Qual.* 31, 1576–1588.

Chakraborty, S., Weindorf, D.C., Morgan, C.L.S., Ge, Y., Galbraith, J.M., Li, B., Kahlon, C.S., 2010. Rapid identification of oil-contaminated soils using visible near-infrared diffuse reflectance spectroscopy. *J. Environ. Qual.* 39 (4), 1378–1387.

Chakraborty, S., Weindorf, D.C., Zhu, Y., Li, B., Morgan, C.L.S., Ge, Y., Galbraith, J., 2012a. Assessing spatial variability of soil petroleum contamination using visible near-infrared diffuse reflectance spectroscopy. *J. Environ. Monit.* 14 (11), 2886–2892.

Chakraborty, S., Weindorf, D.C., Zhu, Y., Li, B., Morgan, C.L.S., Ge, Y., Galbraith, J., 2012b. Spectral reflectance variability from soil physicochemical properties in oil contaminated soils. *Geoderma* 177–178, 80–89.

Che, V.B., Fontijn, K., Ernst, G.G.J., Kervyn, M., Elburg, M., Van Ranst, E., Suh, C.E., 2012. Evaluating the degree of weathering in landslide-prone soils in the humid tropics: the case of Limbe, SW Cameroon. *Geoderma* 170, 378–389.

Chen, J., 2007. Rapid urbanization in China: a real challenge to soil protection and food security. *Catena* 69, 1–15.

Chen, M., Ma, L.Q., Harris, W.G., 2002. Arsenic concentrations in Florida surface soils: influence of soil type and properties. *Soil Sci. Soc. Am. J.* 66, 632–640.

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 hyperspectral remote sensing: a case study of the Rodalquilar mining area, SE Spain. *Remote Sens. Environ.* 112 (7), 3222–3233.

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

- Chu, H.J., Lin, Y.P., Jang, C.S., Chang, T.K., 2010. Delineating the hazard zone of multiple soil pollutants by multivariate indicator kriging and conditioned Latin hypercube sampling. *Geoderma* 158 (3), 242–251.
- Clark, R.N., Roush, T.L., 1984. Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. *J. Geophys. Res.* 89 (NB7), 6329–6340.
- Clements, L., Palaia, T., Davis, J., 2009. Characterisation of sites impacted by petroleum hydrocarbons: national guideline document. CRC Care Technical Report No. 11. CRC for Contamination Assessment and Remediation of the Environment, Adelaide, Australia.
- Cloutis, E.A., 1989. Spectral reflectance properties of hydrocarbons: remote-sensing implications. *Science* 245 (4914), 165–168.
- Cobo, J.G., Dercon, G., Yekeye, T., Chapungu, L., Kadzere, C., Murwira, A., Cadisch, G., 2010. Integration of mid-infrared spectroscopy and geostatistics in the assessment of soil spatial variability at landscape level. *Geoderma* 158 (3), 398–411.
- Crowley, J.L., Demazeau, Y., 1993. Principles and techniques for sensor data fusion. *Signal Process.* 32, 5–27.
- Crumblin, D.M., 2001. Using the triad approach to improve the cost effectiveness of hazardous waste site clean-ups. EPA 542-R-01-016. US EPA, Washington, DC.
- Das, N., Chandran, P., 2011. Microbial degradation of petroleum hydrocarbon contaminants: an overview. *Biotechnol. Res. Int.* 1–13.
- Dawson, J.J.C., Maciel, H., Paton, G.I., Semple, K.T., 2004. Analysis of organic pollutants in environmental samples. In: Smith, K.A., Cresser, M.S. (Eds.), *Soil and Environmental Analysis: Modern Instrument Techniques*. Marcel Dekker, New York, pp. 515–542.
- De Groot, J., Brus, D.J., Bierkens, M.F., Knotters, M., 2006. *Sampling for Natural Resource Monitoring*. Springer.
- Demougeot-Renard, H., de Fouquet, C., Renard, P., 2004. Forecasting the number of soil samples required to reduce remediation cost uncertainty. *J. Environ. Qual.* 33 (5), 1694–1702.
- Diks, C.G.H., Vrugt, J.A., 2010. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stoch. Env. Res. Risk A.* 24, 809–820.
- Dong, Y.W., Yang, S.Q., Xu, C.Y., Li, Y.Z., Bai, W., Fan, Z.N., Wang, Y.N., Li, Q.Z., 2011. Determination of soil parameters in apple-growing regions by near- and mid-infrared spectroscopy. *Pedosphere* 21 (5), 591–602.
- D'Oro, D., Demougeot-Renard, H., Garcia, M., 2009. An integrated geostatistical approach for contaminated site and soil characterisation. *Math. Geosci.* 41, 307–322.
- Dupuy, N., Douay, F., 2001. Infrared and chemometrics study of the interaction between heavy metals and organic matter in soils. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 57 (5), 1037–1047.
- EC, 2005. *Soil Atlas of Europe*. European Soil Bureau Network, Office for Official Publications of the European Communities, Luxembourg, p. 128.
- Elzinga, E.J., Vissenberg, H.A., Van Den Berg, B., Van Grinsven, J.J.M., Swartjes, F.A., 1997. *Freundlich-adsorptievergelijkingen voor cadmium, koper en zink op basis van literatuurgegevens*. RIVM Rep. 711501001.
- Englund, E.J., Heravi, N., 1994. Phased sampling for soil remediation. *Environ. Ecol. Stat.* 1 (3), 247–263.
- EPA Victoria, 2009. *Industrial Waste Resource Guidelines – Soil sampling*. <http://www.epa.vic.gov.au/our-work/publications/publication/2009/july/iwrg621> (last accessed 27th August 2014).
- Eren, H., 2003. *Electronic Portable Instruments: Design and Applications*. CRC Press, Boca Raton.
- Ferguson, C.C., 1999. Assessing risks from contaminated sites: policy and practice in 16 European countries. *Land Contam. Assess.* 7, 33–55.
- Ferreira da Silva, E., Zhang, C., Serrano Pinto, L.S., Patinha, C., Reis, P., 2004. Hazard assessment on arsenic and lead in soils of Castromil gold mining area, Portugal. *Appl. Geochem.* 19 (6), 887–898.
- Fontan, J.M., Calvache, S., Lopez-Bellido, R.J., Lopez-Bellido, L., 2010. Soil carbon measurement in clods and sieved samples in a Mediterranean vertisol by visible and near-infrared reflectance spectroscopy. *Geoderma* 156 (3–4), 93–98.
- Forrester, S., Janik, L., McLaughlin, M., Gilkes, R., 2010. An infrared spectroscopic test for total petroleum hydrocarbon (TPH) contamination in soils. *Proceedings of the 19th World Congress of Soil Science: Soil solutions for a changing world*, Brisbane, Australia, 1–6 August 2010. International Union of Soil Sciences (IUSS), c/o Institut für Bodenforschung, Universität für Bodenkultur, pp. 13–16.
- Forrester, S.F., Janik, L.J., McLaughlin, M.J., 2012. Method of contaminant prediction. U.S. Provisional Patent Application No. 61/245,346.
- Forrester, S.T., Janik, L.J., McLaughlin, M.J., Soriano-Disla, J.M., Stewart, R., Dearman, B., 2013. Total petroleum hydrocarbon concentration prediction in soils using diffuse reflectance infrared spectroscopy. *Soil Sci. Soc. Am. J.* 77 (2), 450–460.
- Franco, C., Soares, A., Delgado, J., 2006. Geostatistical modelling of heavy metal contamination in the topsoil of Guadamar river margins (S Spain) using a stochastic simulation technique. *Geoderma* 136 (3), 852–864.
- Friebe, E., Nadebaum, P., 2011. HSLs for petroleum hydrocarbons in soil and groundwater, part 1: technical development document. Technical Report No. 10. CRC for Contamination Assessment and Remediation of the Environment, Adelaide, Australia.
- Fujimori, T., Takigami, H., 2014. Pollution distribution of heavy metals in surface soil at an informal electronic-waste recycling site. *Environ. Geochem. Health* 36 (1), 159–168.
- Galvez-Sola, L., Morales, J., Mayoral, A.M., Paredes, C., Bustamante, M.A., Marhuenda-Egea, F.C., Barber, J.X., Moral, R., 2013. Estimation of parameters in sewage sludge by near-infrared reflectance spectroscopy (NIRS) using several regression tools. *Talanta* 110, 81–88.
- García Gómez, R.S., Pandiyan, T., Aguilar Iris, V.E., Luna-Pabello, V., Duran De Bazúa, C., 2004. Spectroscopic determination of poly-aromatic compounds in petroleum contaminated soils. *Water Air Soil Pollut.* 158 (1), 137–151.
- Gay, J.R., Korre, A., 2006. A spatially-evaluated methodology for assessing risk to a population from contaminated land. *Environ. Pollut.* 142 (2), 227–234.
- Ge, L., Lai, W., Lin, Y., 2005. Influence of and correction for moisture in rocks, soils and sediments on in situ XRF analysis. *X-Ray Spectrom.* 34 (1), 28–34.
- Ge, Y., Morgan, C.L., Ackerson, J.P., 2014. VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. *Geoderma* 221, 61–69.
- Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. *Anal. Chim. Acta.* 185, 1–17.
- Gilbert, R.O., 1987. *Statistical methods for environmental pollution monitoring*. Chapter 5. Stratified Random Sampling. John Wiley, NY.
- Gobrecht, A., Roger, J.M., Bellon-Maurel, V., 2014. Chapter four – major issues of diffuse reflectance NIR spectroscopy in the specific context of soil carbon content estimation: a review. *Adv. Agron.* 123, 145–175.
- Goldberg, S., Johnston, C.T., 2001. Mechanisms of arsenic adsorption on amorphous oxides evaluated using macroscopic measurements, vibrational spectroscopy, and surface complexation modeling. *J. Colloid Interface Sci.* 234 (1), 204–216.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. *Geoderma* 103 (1), 3–26.
- Goovaerts, P., Webster, R., Dubois, J.P., 1997. Assessing the risk of soil contamination in the Swiss Jura using indicator geostatistics. *Environ. Ecol. Stat.* 4 (1), 49–64.
- Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasts. *J. Forecast.* 3, 197–204.
- Gras, J.-P., Barthes, B.G., Mahaut, B., Trupin, S., 2014. Best practices for obtaining and processing field visible and near infrared (VNIR) spectra of topsoils. *Geoderma* 214, 126–134.
- Hazel, G., Bucholtz, F., Aggarwal, I.D., Nau, G., Ewing, K.J., 1997. Multivariate analysis of Mid-IR FT-IR spectra of hydrocarbon-contaminated wet soils. *Appl. Spectrosc.* 51 (7), 984–989.
- He, Y., Tang, L., Wu, X., Hou, X., Lee, Y.-I., 2007. Spectroscopy: the best way toward green analytical chemistry? *Appl. Spectrosc. Rev.* 42, 119–138.
- Head, I.M., Jones, D.M., Larter, S.R., 2003. Biological activity in the deep subsurface and the origin of heavy oil. *Nature* 426, 344–352.
- Heuvelink, G.B.M., Bierkens, M.F.P., 1992. Combining soil maps with interpolations from point observations to predict quantitative soil properties. *Geoderma* 55, 1–15.
- Hjort, N.L., Claeskens, G., 2003. Frequentist model average estimators. *J. Am. Stat. Assoc.* 98 (464), 879–899.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial (with discussion). *Stat. Sci.* 14, 382–401 ([Corrected version.] Correction (15), 193–195).
- Hooker, P.J., Nathanail, C.P., 2006. Risk-based characterisation of lead in urban soils. *Chem. Geol.* 226 (3), 340–351.
- Hörig, B., Kühn, F., Oschütz, F., Lehmann, F., 2001. HyMap hyperspectral remote sensing to detect hydrocarbons. *Int. J. Remote Sens.* 22, 1413–1422.
- Horn, A.L., Reiher, W., Düring, R.A., Gäth, S., 2006. Efficiency of pedotransfer functions describing cadmium sorption in soils. *Water Air Soil Pollut.* 170 (1–4), 229–247.
- Hou, X., He, Y., Jones, B.T., 2004. Recent advances in portable X-ray fluorescence spectrometry. *Appl. Spectrosc. Rev.* 39 (1), 1–25.
- 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). *Environ. Pollut.* 124, 247–256.
- Journel, A.G., 1983. Non-parametric estimation of spatial distributions. *Math. Geol.* 15 (3), 445–468.
- Journel, A.G., 1992. Geostatistics: roadblocks and challenges. In: Soares, A. (Ed.), *Geostatistics Troia 92*. Kluwer Academic Publishers, Dordrecht, pp. 213–224.
- Juang, K.W., Chen, Y.S., Lee, D.Y., 2004. Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils. *Environ. Pollut.* 127 (2), 229–238.
- Juang, K.W., Lee, D.Y., Teng, Y.L., 2005. Adaptive sampling based on the cumulative distribution function of order statistics to delineate heavy-metal contaminated soils using kriging. *Environ. Pollut.* 138 (2), 268–277.
- Juang, K.W., Liao, W.J., Liu, T.L., Tsui, L., Lee, D.Y., 2008. Additional sampling based on regulation threshold and kriging variance to reduce the probability of false delineation in a contaminated site. *Sci. Total Environ.* 389 (1), 20–28.
- Kalnicky, D.J., Singhvi, R., 2001. Field portable XRF analysis of environmental samples. *J. Hazard. Mater.* 83 (1–2), 93–122.
- Kemper, T., Sommer, S., 2002. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environ. Sci. Technol.* 36 (12), 2742–2747.
- Kleinebecker, T., Poelen, M.D.M., Smolders, A.J.P., Lamers, L.P.M., Hölzel, N., 2013. Fast and inexpensive detection of total and extractable element concentrations in aquatic sediments using Near-Infrared Reflectance Spectroscopy (NIRS). *PLoS ONE* 8 (7).
- Kooistra, L., Wehrens, R., Leuven, R.S.E.W., Buydens, L.M.C., 2001. Possibilities of visible-near-infrared spectroscopy for the assessment of soil contamination in river floodplains. *Anal. Chim. Acta.* 446 (1–2), 97–105.
- Kowalski, C.J., 1972. On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *Appl. Stat.* 1–12.
- Kowalski, B.R., 1975. Chemometrics: views and propositions. *J. Chem. Inf. Comput. Sci.* 15, 201–203.
- Kühn, F., Oppermann, K., Hörig, B., 2004. Hydrocarbon index – an algorithm for hyperspectral detection of hydrocarbons. *Int. J. Remote Sens.* 25, 2467–2473.
- Kyriakidis, P.C., 1997. Selecting panels for remediation in contaminated soils via stochastic imaging. In: Baafi, E.Y., Schofield, N.A. (Eds.), *Geostatistics Wollongong '96*. vol. 2. Kluwer Academic Publishers, The Netherlands, pp. 973–983.
- Laiho, J.V.P., Perämäki, P., 2005. Evaluation of portable X-ray fluorescence (PXRF) sample preparation methods. *Geol. Surv. Finland Spec. Pap.* 38, 73–82.

- Lambert, P., Fingas, M., Goldthorp, M., 2001. An evaluation of field total petroleum hydrocarbon (TPH) systems. *J. Hazard. Mater.* 83, 65–81.
- Lang, D.A., Bastow, T.P., Van Aarssen, B.G., Warton, B., Davis, G.B., Johnston, C.D., 2009. Polar compounds from the dissolution of weathered diesel. *Ground Water Monit. Remediat.* 29, 85–93.
- Lark, R.M., 2000. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *Eur. J. Soil Sci.* 51 (4), 717–728.
- Lark, R.M., 2002. Modelling complex soil properties as contaminated regionalized variables. *Geoderma* 106 (3), 173–190.
- Lark, R.M., 2009. Kriging a soil variable with a simple nonstationary variance model. *J. Agric. Biol. Environ. Stat.* 14 (3), 301–321.
- Lark, R.M., 2012. Towards soil geostatistics. *Spat. Stat.* 1, 92–99.
- Lark, R.M., Cullis, B.R., Welham, S.J., 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. *Eur. J. Soil Sci.* 57 (6), 787–799.
- Lavine, B.K., 2005. *Chemometrics and Cheminformatics*. American Chemical Society.
- Lewandowski, I., Schmidt, U., Londo, M., Faaij, A., 2006. The economic value of the phytoremediation function – assessed by the example of cadmium remediation by willow (*Salix* spp.). *Agric. Syst.* 89, 68–89.
- Lin, Y.P., Chu, H.J., Huang, Y.L., Cheng, B.Y., Chang, T.K., 2011. Modeling spatial uncertainty of heavy metal content in soil by conditional Latin hypercube sampling and geostatistical simulation. *Environ. Earth Sci.* 62 (2), 299–311.
- 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. *Environ. Pollut.* 141 (2), 257–264.
- Liu, Y.L., Wen, C., Liu, X.J., 2013. China's food security soiled by contamination. *Science* 339, 1382–1383.
- Lund, E.D., 2010. Can field-based spectroscopic sensors measure soil carbon in a regulated carbon trading program? Proceedings of the 19th World Congress of Soil Science: Soil Solutions for a Changing World, Brisbane, Australia, pp. 33–36.
- Luo, X.-S., Yu, S., Zhu, Y.-G., Li, X.-D., 2012. Trace metal contamination in urban soils of China. *Sci. Total Environ.* 421–422, 17–30.
- Malherbe, L., 2002. Designing a contaminated soil sampling strategy for human health risk assessment. *Accred. Qual. Assur.* 7 (5), 189–194.
- Malley, D.F., Williams, P.C., 1997. Use of near-infrared reflectance spectroscopy in prediction of heavy metals in freshwater sediment by their association with organic matter. *Environ. Sci. Technol.* 31 (12), 3461–3467.
- Malley, D.F., Hunter, K.N., Webster, G.R.B., 1999. Analysis of diesel fuel contamination in soils by near-infrared reflectance spectrometry and solid phase microextraction-gas chromatography. *Soil Sediment Contam.* 8 (4), 481–489.
- Malone, B.P., Minasny, B., Odgers, N.P., McBratney, A.B., 2014. Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* 232–234, 34–44.
- Marchant, B.P., Lark, R.M., 2007. Optimized sample schemes for geostatistical surveys. *Math. Geol.* 39 (1), 113–134.
- Marchant, B.P., Saby, N.P.A., Jolivet, C.C., Arrouays, D., Lark, R.M., 2011a. Spatial prediction of soil properties with copulas. *Geoderma* 162, 327–334.
- Marchant, B.P., Tye, A.M., Rawlins, B.G., 2011b. The assessment of point-source and diffuse soil metal pollution using robust geostatistical methods: a case study in Swansea (Wales, UK). *Eur. J. Soil Sci.* 62, 346–358.
- Marchant, B.P., McBratney, A.B., Lark, R.M., Minasny, B., 2013. Optimized multi-phase sampling for soil remediation surveys. *Spat. Stat.* 4, 1–13.
- Markowicz, A.A., 2008. Quantification and correction procedures. In: Potts, P.J., West, M. (Eds.), *Portable X-ray Fluorescence Spectrometry: Capabilities for In Situ Analysis*. The Royal Society of Chemistry, Cambridge, UK, pp. 13–38.
- Markus, J., McBratney, A.B., 2001. A review of the contamination of soil with lead: II. Spatial distribution and risk assessment of soil lead. *Environ. Int.* 27, 399–411.
- McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, W.R., 2002. From pedotransfer functions to soil inference systems. *Geoderma* 109, 41–73.
- McBratney, A.B., Minasny, B., Viscarra Rossel, R.A., 2006. Spectral soil analysis and inference systems: a powerful combination for solving the soil data crisis. *Geoderma* 136, 272–278.
- McBratney, A.B., Minasny, B., Whelan, B.M., 2011a. Defining proximal soil sensing. The Second Global Workshop on Proximal Soil Sensing – IUSS Working Group on Proximal Soil Sensing, Montreal, Quebec, Canada.
- McBratney, A.B., Minasny, B., Tranter, G., 2011b. Necessary metadata for pedotransfer functions. *Geoderma* 160, 627–629.
- McLaren, T.I., Guppy, C.N., Tighe, M.K., Forster, N., Grave, P., Lisle, L.M., Bennett, J.W., 2012. Rapid, nondestructive total elemental analysis of vertisol soils using portable X-ray fluorescence. *Soil Sci. Soc. Am. J.* 76 (4), 1436–1445.
- McWhirt, A., Weindorf, D.C., Zhu, Y., 2012. Rapid analysis of elemental concentrations in compost via portable X-ray fluorescence spectrometry. *Compost Sci. Util.* 20 (3), 185–193.
- Merry, R.H., Janik, L.J., 2001. Mid Infrared Spectroscopy for Rapid and Cheap Analysis of Soils. Urrbrae, South Australia, (<http://www.regional.org.au/au/asa/2001/3/c/merry.htm>).
- Middleton, M., Närhi, P., Kuosmanen, V., Sutinen, R., 2011. Quantification of glacial till chemical composition by reflectance spectroscopy. *Appl. Geochem.* 26 (12), 2215–2225.
- Miklos, M., Short, M.G., McBratney, A.B., Minasny, B., 2010. Mapping and comparing the distribution of soil carbon under cropping and grazing management practices in Narrabri, north-west New South Wales. *Soil Res.* 48, 248–257.
- Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemom. Intell. Lab. Syst.* 94 (1), 72–79.
- Minasny, B., McBratney, A.B., Pichon, L., Sun, W., Short, M.G., 2009a. Evaluating near infrared spectroscopy for field prediction of soil properties. *Aust. J. Soil Res.* 47 (7), 664–673.
- Minasny, B., Tranter, G., McBratney, A.B., Brough, D.M., Murphy, B.W., 2009b. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma* 153 (1–2), 155–162.
- Minasny, B., McBratney, A.B., Bellon-Maurel, V., Roger, J.-M., Gobrecht, A., Ferrand, L., Joalland, S., 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma* 167–68, 118–124.
- Ministry of Housing, Spatial Planning and the Environment, 2010. Into Dutch Soils. www.deltares.nl/xmlpages/tan/files?p_file_id=14053 (last accessed 27th August 2014).
- Morgan, C.L.S., Waiser, T.H., Brown, D.J., Hallmark, C.T., 2009. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. *Geoderma* 151 (3–4), 249–256.
- Motelay-Massei, A., Ollivon, D., Garban, B., Teil, M.J., Blanchard, M., Chevreuil, M., 2004. Distribution and spatial trends of PAHs and PCBs in soils in the Seine River basin, France. *Chemosphere* 55 (4), 555–565.
- Mulligan, C.N., Yong, R.N., Gibbs, B.F., 2001. Remediation technologies for metal-contaminated soils and groundwater: an evaluation. *Eng. Geol.* 60, 193–207.
- Narizzano, R., Riso, F., Magherini, A., Magi, E., Giampieri, M., Devia, C., Venturini, G., Carlini, E., 2013. Effect of drying conditions during sample pre-treatment on the determination of polycyclic aromatic hydrocarbons in soils. *Anal. Bioanal. Chem.* 405 (2–3), 1115–1121.
- National Environment Protection (Assessment of Site Contamination) Measure as amended in 2013, National Environment Protection Council, Canberra, Australia. <http://www.scew.gov.au/nepms/assessment-site-contamination> (last accessed 27th August 2014).
- NSW, E.P.A., 1995. Sampling Design Guidelines. <http://www.epa.nsw.gov.au/resources/clm/95059samppgdline.pdf> (last accessed 27th August 2014).
- NSW EPA, 2013. Regulatory impact statement – proposed contaminated land management regulation. <http://www.epa.nsw.gov.au/clm/130403risclm.htm> (last accessed 27th August 2014).
- Odlare, M., Svensson, K., Pell, M., 2005. Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field. *Geoderma* 126 (3), 193–202.
- Okparanma, R.N., Mouazen, A.M., 2013a. Determination of total petroleum hydrocarbon (TPH) and polycyclic aromatic hydrocarbon (PAH) in soils: a review of spectroscopic and nonspectroscopic techniques. *Appl. Spectrosc. Rev.* 48 (6), 458–486.
- Okparanma, R.N., Mouazen, A.M., 2013b. Visible and near-infrared spectroscopy analysis of a polycyclic aromatic hydrocarbon in soils. *Sci. World J.* 2013.
- Okparanma, R.N., Mouazen, A.M., 2013c. Combined effects of oil concentration, clay and moisture contents on diffuse reflectance spectra of diesel-contaminated soils. *Water Air Soil Pollut.* 224 (5).
- Okparanma, R.N., Coulon, F., Mouazen, A.M., 2014. Analysis of petroleum-contaminated soils by diffuse reflectance spectroscopy and sequential ultrasonic solvent extraction-gas chromatography. *Environ. Pollut.* 184, 298–305.
- Olympus InnovX-Systems, 2010. User Manual. DeltaTM Family. Handheld XRF Analyzers, US.
- Paiga, P., Mendes, L., Albergaria, J.T., Delerue-Matos, C.M., 2012. Determination of total petroleum hydrocarbons in soil from different locations using infrared spectrophotometry and gas chromatography. *Chem. Pap.* 66, 711–721.
- Panagos, P., Van Liedekerke, M., Yigini, Y., Montanarella, L., 2013. Contaminated sites in Europe: review of the current situation based on data collected through a European network. *J. Environ. Public Health* 2013.
- Piccolo, A., Stevenson, F.J., 1982. Infrared spectra of Cu²⁺ + Pb²⁺ and Ca²⁺ complexes of soil humic substances. *Geoderma* 27 (3), 195–208.
- Potts, Philip, 2008. Introduction, analytical instrumentation and application overview. In: Potts, Philip J., West, Margaret (Eds.), *Portable X-ray Fluorescence Spectrometry: Capabilities for In Situ Analysis*. Royal Society of Chemistry, Cambridge, UK, pp. 1–12.
- Potts, P.J., West, M., 2008. *Portable X-ray Fluorescence Spectrometry: Capabilities for In Situ Analysis*. The Royal Society of Chemistry, Cambridge, UK.
- Radu, T., Diamond, D., 2009. Comparison of soil pollution concentrations determined using AAS and portable XRF techniques. *J. Hazard. Mater.* 171 (1–3), 1168–1171.
- Radu, T., Gallagher, S., Byrne, B., Harris, P., Coveney, S., McCarron, S., McCarthy, T., Diamond, D., 2013. Portable X-ray fluorescence as a rapid technique for surveying elemental distributions in soil. *Spectrosc. Lett.* 46 (7), 516–526.
- Ramsay, M., 2008. Contaminated land: cost-effective investigation within sampling constraints. In: Potts, P.J., West, M. (Eds.), *Portable X-ray Fluorescence Spectrometry: Capabilities for In Situ Analysis*. The Royal Society of Chemistry, Cambridge, UK, pp. 39–55.
- Rathod, P.H., Rossiter, D.G., Noomen, M.F., van der Meer, F.D., 2013. Proximal spectral sensing to monitor phytoremediation of metal-contaminated soils. *Int. J. Phytoremediation* 15 (5), 405–426.
- Reeves III, J.B., 2010. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: where are we and what needs to be done? *Geoderma* 158 (1–2), 3–14.
- Reeves III, J.B., Smith, D.B., 2009. The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America. *Appl. Geochem.* 24 (8), 1472–1481.
- Regan, H.M., Sample, B.E., Ferson, S., 2002. Comparison of deterministic and probabilistic calculation of ecological soil screening levels. *Environ. Toxicol. Chem.* 21, 882–890.
- Richter, N., Jarmer, T., Chabrilat, S., Oyonarte, C., Hostert, P., Kaufmann, H., 2009. Free iron oxide determination in Mediterranean soils using diffuse reflectance spectroscopy. *Soil Sci. Soc. Am. J.* 73 (1), 72–81.
- Rodrigues, S.M., Pereira, M.E., Ferreira da Silva, E., Hursthouse, A.S., Duarte, A.C., 2009. A review of regulatory decisions for environmental protection: part II—the case-study of contaminated land management in Portugal. *Environ. Int.* 35, 214–225.

- Roger, J.M., Chauchard, F., Bellon-Maurel, V., 2003. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemom. Intell. Lab. Syst.* 66 (2), 191–204.
- Rojas, R., Feyen, L., Dassargues, A., 2008. Conceptual model uncertainty in groundwater modeling: combining generalized likelihood uncertainty estimation and Bayesian model averaging. *Water Resour. Res.* 44 (12).
- Saby, N., Arrouays, D., Bouillon, L., Jolivet, C., Pochot, A., 2006. Geostatistical assessment of Pb in soil around Paris, France. *Sci. Total Environ.* 367 (1), 212–221.
- Sadler, R., Connell, D., 2003. Analytical methods for the determination of total petroleum hydrocarbons. In: Langley, A., Gilbey, M., Kennedy, B. (Eds.), *Soil Proceedings of the Fifth National Workshop on the Assessment of Site Contamination*, pp. 133–150.
- Saito, H., Goovaerts, P., 2002. Accounting for measurement error in uncertainty modeling and decision-making using indicator kriging and p-field simulation: application to a dioxin contaminated site. *Environmetrics* 13 (5–6), 555–567.
- Salmeen, I.T., Kim, B.R., Briggs, L.M., 1995. Case of lognormally distributed TPH in contaminated soil. *J. Environ. Eng.* 121 (9), 664–667.
- Sankey, J.B., Brown, D.J., Bernard, M.L., Lawrence, R.L., 2008. Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. *Geoderma* 148 (2), 149–158.
- Schultz, D.S., 1997. Electroosmosis technology for soil remediation: laboratory results, field trial, and economic modeling. *J. Hazard. Mater.* 55, 81–91.
- Schwartz, G., Eshel, G., Ben-Dor, E., 2011. Reflectance spectroscopy as a tool for monitoring contaminated soils. In: Pascucci, Simone (Ed.), *Soil Contamination*. InTech, pp. 67–90.
- Schwartz, G., Ben-Dor, E., Eshel, G., 2012. Quantitative analysis of total petroleum hydrocarbons in soils: comparison between reflectance spectroscopy and solvent extraction by 3 certified laboratories. *Appl. Environ. Soil Sci.* 2012.
- Schwartz, G., Ben-Dor, E., Eshel, G., 2013. Quantitative assessment of hydrocarbon contamination in soil using reflectance spectroscopy: a “multipath” approach. *Appl. Spectrosc.* 67 (11), 1323–1331.
- Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for the determination of soil properties. *Soil Sci. Soc. Am. J.* 66, 988–998.
- Shi, T., Chen, Y., Liu, Y., Wu, G., 2014. Visible and near-infrared reflectance spectroscopy—an alternative for monitoring soil contamination by heavy metals. *J. Hazard. Mater.* 265, 166–176.
- Siebielec, G., McCarty, G.W., Stuczynski, T.I., Reeves III, J.B., 2004. Near- and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content. *J. Environ. Qual.* 33 (6), 2056–2069.
- Slonecker, T., Fisher, G.B., Aiello, D.P., Haack, B., 2010. Visible and infrared remote imaging of hazardous waste: a review. *Remote Sens.* 2, 2474–2508.
- Soares, A., 2001. Direct sequential simulation and cosimulation. *Math. Geol.* 33 (8), 911–926.
- Sollitto, D., Romić, M., Castrignanò, A., Romić, D., Bakic, H., 2010. Assessing heavy metal contamination in soils of the Zagreb region (Northwest Croatia) using multivariate geostatistics. *Catena* 80 (3), 182–194.
- Sorak, D., Herberholz, L., Iwaszek, S., Altinpinar, S., Pfeifer, F., Siesler, H.W., 2012. New developments and applications of handheld Raman, mid-infrared, and near-infrared spectrometers. *Appl. Spectrosc. Rev.* 47 (2), 83–115.
- Soriano-Disla, J.M., Janik, L., McLaughlin, M.J., Forrester, S., Kirby, J., Reimann, C., Albanese, S., Andersson, M., Arnoldussen, A., Baritz, R., Batista, M.J., Bel-Lan, A., Birke, M., Cicchella, D., Demetriades, A., Dinelli, E., De Vivo, B., De Vos, W., Dohrmann, R., Duris, M., Dusza-Dobek, A., Eggen, O.A., Eklund, M., Ernsten, V., Filzmoser, P., Finne, T.E., Flight, D., Fuchs, M., Fugedi, U., Gilucis, A., Gosar, M., Gregorauskiene, V., Gulan, A., Halamid, J., Haslinger, E., Hayoz, P., Hobbiger, G., Hoffmann, R., Hoogewerf, J., Hrvatic, H., Husnjak, S., Johnson, C.C., Jordan, G., Kivisilla, J., Klos, V., Krone, F., Kweck, P., Kuti, L., Ladenberger, A., Lima, A., Locutura, J., Lucivjansky, P., Mackovych, D., Malyuk, B.I., Maquill, R., Meuli, R.G., Miosic, N., Mol, G., Négrel, P., O'Connor, P., Oorts, K., Ottesen, R.T., Pasieczna, A., Petersell, V., Pfeleiderer, S., Poňavič, M., Prazeres, C., Rauch, U., Salpateur, I., Schedl, A., Scheib, A., Schoeters, I., Sefcik, P., Sellersjö, E., Skopljak, F., Slaninka, I., Šorša, A., Srnkota, R., Stafilov, T., Tarvainen, T., Trendavilov, V., Valera, P., Verougstraete, V., Vidojevid, D., Zissimos, A.M., Zomeni, Z., 2013. The use of diffuse reflectance mid-infrared spectroscopy for the prediction of the concentration of chemical elements estimated by X-ray fluorescence in agricultural and grazing European soils. *Appl. Geochem.* 29, 135–143.
- Soriano-Disla, J.M., Janik, L.J., Rossel, R.A.V., Macdonald, L.M., McLaughlin, M.J., 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Appl. Spectrosc. Rev.* 49 (2), 139–186.
- Stallard, B.R., Garcia, M.J., Kaushik, S., 1996. Near-IR reflectance spectroscopy for the determination of motor oil contamination in sandy loam. *Appl. Spectrosc.* 50 (3), 334–338.
- State of the Environment Committee, 2011. Australia State of the Environment 2011. Independent report to the Australian Government Minister for Sustainability, Environment, Water, Population and Communities (Canberra, Australia). <http://www.environment.gov.au/science/soe/2011> (last accessed 27th August 2014).
- Stenberg, B., 2010. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma* 158 (1–2), 15–22.
- Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and Near Infrared Spectroscopy in Soil Science, pp. 163–215.
- Stevens, A., van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., Ben-Dor, E., 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma* 144 (1–2), 395–404.
- Stuart, B., 2004. *Infrared Spectroscopy*. Wiley Online Library.
- Sun, Y., Chen, Z.h., Xu, S., Cai, P., 2005. Stable carbon and hydrogen isotopic fractionation of individual n-alkanes accompanying biodegradation: evidence from a group of progressively biodegraded oils. *Org. Geochem.* 36, 225–238.
- Sut, M., Fischer, T., Repmann, F., Raab, T., Dimitrova, T., 2012. Feasibility of field portable near infrared (NIR) spectroscopy to determine cyanide concentrations in soil. *Water Air Soil Pollut.* 223 (8), 5495–5504.
- Swindell, A.L., Reid, B.J., 2006. Influence of diesel concentration on the fate of phenanthrene in soil. *Environ. Pollut.* 140, 79–86.
- Tranter, G., McBratney, A.B., Minasny, B., 2009. Using distance metrics to determine the appropriate domain of pedotransfer function predictions. *Geoderma* 149 (3–4), 421–425.
- Tranter, G., Minasny, B., McBratney, A.B., 2010. Estimating pedotransfer function prediction limits using fuzzy k-means with extragrades. *Soil Sci. Soc. Am. J.* 74, 1967–1975.
- US EPA, 2002. Guidance for Choosing a Sampling Design for Environmental Data Collection (EPA QA/G-55). Office of Environmental Information, Washington DC, USA (<http://www.epa.gov/quality/> (last accessed 27th August 2014)).
- USDA-NRCS, 2014. Soil survey field and laboratory methods manual. In: Burt, R., Survey Staff, Soil (Eds.), *Soil Survey Investigations Report No. 51, Version 2.0*. U.S. Department of Agriculture, Natural Resources Conservation Service.
- van der Perk, M., 2006. *Soil and Water Contamination: From Molecular to Catchment Scale*. CRC Press, Boca Raton.
- Van Groenigen, J.W., Pieters, G., Stein, A., 2000. Optimizing spatial sampling for multivariate contamination in urban areas. *Environmetrics* 11 (2), 227–244.
- Van Liedekerke, M., Prokop, G., Rabl-Berger, S., Kibblewhite, M., Louwagie, G., 2014. Progress in the management of contaminated sites in Europe. Reference Report by the Joint Research Centre of the European Commission (<http://www.eea.europa.eu/data-and-maps/indicators/progress-in-management-of-contaminated-sites> (last accessed 27th August 2014)).
- Van Meirvenne, M., Goovaerts, P., 2001. Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. *Geoderma* 102 (1), 75–100.
- Verstraete, S., Van Meirvenne, M., 2008. A multi-stage sampling strategy for the delineation of soil pollution in a contaminated brownfield. *Environ. Pollut.* 154 (2), 184–191.
- Villanneau, E.J., Saby, N., Marchant, B.P., Jolivet, C.C., Bouillon, L., Caria, G., Arrouays, D., 2011. Which persistent organic pollutants can we map in soil using a large spacing systematic soil monitoring design? A case study in Northern France. *Sci. Total Environ.* 409 (19), 3719–3731.
- Viscarra Rossel, R.A., McBratney, A.B., 1998. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. *Geoderma* 85 (1), 19–39.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131 (1–2), 59–75.
- Viscarra Rossel, R.A., Cattle, S.R., Ortega, A., Fouad, Y., 2009. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma* 150 (3–4), 253–266.
- Viscarra Rossel, R.A., McBratney, A.B., Minasny, B., 2010. *Proximal Soil Sensing*. Springer, The Netherlands.
- Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C., 2011. Proximal soil sensing: an effective approach for soil measurements in space and time. *Adv. Agron.* 113, 237–282.
- Vohland, M., Bossung, C., Fründ, H.C., 2009. A spectroscopic approach to assess trace – heavy metal contents in contaminated floodplain soils via spectrally active soil components. *J. Plant Nutr. Soil Sci.* 172 (2), 201–209.
- Vohra, S.T., Bucholtz, F., Nau, G.M., Ewing, K.J., Aggarwal, I.D., 1996. Remote detection of trichloroethylene in soil by a fiber-optic infrared reflectance probe. *Appl. Spectrosc.* 50 (8), 985–990.
- Vreugdenhil, A.J., Butler, I.S., 1998. Investigation of MMT adsorption on soils by Diffuse Reflectance Infrared Spectroscopy (DRIFTS) and Headspace Analysis Gas-phase Infrared Spectroscopy (HAGIS). *Appl. Organomet. Chem.* 12 (2), 121–128.
- Wagner, T., Gupta, H.V., 2005. Model identification for hydrological forecasting under uncertainty. *Stoch. Env. Res. Risk A* 19, 378–387.
- Waiser, T.H., Morgan, C.L.S., Brown, D.J., Hallmark, C.T., 2007. In situ characterization of soil clay content with visible near-infrared diffuse reflectance spectroscopy. *Soil Sci. Soc. Am. J.* 71 (2), 389–396.
- Walvoort, D.J.J., Brus, D.J., de Gruijter, J.J., 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. *Comput. Geosci.* 36, 1261–1267.
- Wang, S.-q., Li, W.-d., Li, J., Liu, X.-s., 2013. Prediction of soil texture using FT-NIR spectroscopy and PXRf spectroscopy with data fusion. *Soil Sci.* 178 (11), 626–638.
- Weindorf, D.C., Sarkar, R., Dia, M., Wang, H., Chang, Q., Haggard, B., McWhirt, A., Wooten, A., 2008. Correlation of X-ray fluorescence spectrometry and inductively coupled plasma atomic emission spectroscopy for elemental determination in composted products. *Compost Sci. Util.* 16 (2), 79–82.
- Weindorf, D.C., Zhu, Y., Ferrell, R., Rolong, N., Barnett, T., Allen, B.L., Herrero, J., Hudnall, W., 2009. Evaluation of portable X-ray fluorescence for gypsum quantification in soils. *Soil Sci.* 174 (10), 556–562.
- Weindorf, D.C., Zhu, Y., Chakraborty, S., Bakr, N., Huang, B., 2012a. Use of portable X-ray fluorescence spectrometry for environmental quality assessment of peri-urban agriculture. *Environ. Monit. Assess.* 184 (1), 217–227.
- Weindorf, D.C., Zhu, Y., Haggard, B., Lofton, J., Chakraborty, S., Bakr, N., Zhang, W., Weindorf, W.C., Legoria, M., 2012b. Enhanced pedon horizonation using portable X-ray fluorescence spectrometry. *Soil Sci. Soc. Am. J.* 76 (2), 522–531.
- Weindorf, D.C., Zhu, Y., McDaniel, P., Valerio, M., Lynn, L., Michaelson, G., Clark, M., Ping, C.L., 2012c. Characterizing soils via portable X-ray fluorescence spectrometer: 2. Spodic and Albic horizons. *Geoderma* 189–190, 268–277.

- Weindorf, D.C., Herrero, J., Castañeda, C., Bakr, N., Swanhart, S., 2013a. Direct soil gypsum quantification via portable X-ray fluorescence spectrometry. *Soil Sci. Soc. Am. J.* 77 (6), 2071–2077.
- Weindorf, D.C., Paulette, L., Man, T., 2013b. In-situ assessment of metal contamination via portable X-ray fluorescence spectroscopy: Zlatna, Romania. *Environ. Pollut.* 182, 92–100.
- Weindorf, D.C., Bakr, N., Zhu, Y., McWhirt, A., Ping, C.L., Michaelson, G., Nelson, C., Shook, K., Nuss, S., 2014. Influence of ice on soil elemental characterization via portable X-ray fluorescence spectrometry. *Pedosphere* 24 (1), 1–12.
- West, M., Ellis, A.T., Potts, P.J., Strel, C., Vanhoof, C., Wegrzynek, D., Wobrauschek, P., 2013. 2013 atomic spectrometry update — a review of advances in X-ray fluorescence spectrometry. *J. Anal. At. Spectrom.* 28 (10), 1544–1590.
- Wu, Y., Chen, J., Wu, X., Tian, Q., Ji, J., Qin, Z., 2005a. Possibilities of reflectance spectroscopy for the assessment of contaminant elements in suburban soils. *Appl. Geochem.* 20 (6), 1051–1059.
- Wu, Y.Z., Chen, J., Ji, J.F., Tian, Q.J., Wu, X.M., 2005b. Feasibility of reflectance spectroscopy for the assessment of soil mercury contamination. *Environ. Sci. Technol.* 39 (3), 873–878.
- Wu, Y., Chen, J., Ji, J., Gong, P., Liao, Q., Tian, Q., Ma, H., 2007. A mechanism study of reflectance spectroscopy for investigating heavy metals in soils. *Soil Sci. Soc. Am. J.* 71 (3), 918–926.
- Wu, M.L., Nie, M.Q., Wang, X.C., Su, J.M., Cao, W., 2010. Analysis of phenanthrene biodegradation by using FTIR, UV and GC–MS. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 75, 1047–1050.
- Wu, C.M., Tsai, H.T., Yang, K.H., Wen, J.C., 2012. How reliable is X-ray fluorescence (XRF) measurement for different metals in soil contamination? *Environ. Forensic* 13 (2), 110–121.
- Yang, P., Mao, R., Shao, H., Gao, Y., 2009. An investigation on the distribution of eight hazardous heavy metals in the suburban farmland of China. *J. Hazard. Mater.* 167 (1–3), 1246–1251.
- Zarella, W.M., Mousseau, R.J., Coggeshall, N.D., Norris, M.S., Schryer, G.J., 1967. Analysis and significance of hydrocarbons in subsurface brines. *Geochim. Cosmochim. Acta* 31, 1155–1166.
- Zhang, H., Huang, G.H., Zeng, G.M., 2009. Health risks from arsenic-contaminated soil in Flin Flon–Creighton, Canada: integrating geostatistical simulation and dose–response model. *Environ. Pollut.* 157 (8), 2413–2420.
- Zhu, Y., Weindorf, D.C., 2009. Determination of soil calcium using field portable X-ray fluorescence. *Soil Sci.* 174 (3), 151–155.
- Zhu, Y., Weindorf, D.C., Zhang, W., 2011. Characterizing soils using a portable X-ray fluorescence spectrometer: 1. Soil texture. *Geoderma* 167–68, 167–177.
- Zupan, M., Einax, J., Kraft, J., Lobnik, F., Hudnik, V., 2000. Chemometric characterization of soil and plant pollution: part 1: multivariate data analysis and geostatistical determination of relationship and spatial structure of inorganic contaminants in soil. *Environ. Sci. Pollut. Res.* 7 (2), 89–96.
- Zwanziger, H.W., Förster, H., 1998. Near infrared spectroscopy of fuel contaminated sand and soil. I. preliminary results and calibration study. *J. Near Infrared Spectrosc.* 6 (1–4), 189–197.