**Environmental Pollution Journal**

**journal homepage: [www.elsevier.com/locate/envpol](http://www.elsevier.com/locate/envpol)**

**Application of Machine Learning in Geo-Environmental Engineering using satellite imagery: A review**

**Amir TavallaieNejada, b, Maria de Lurdes Dinisa,b**

aCERENA-Polo FEUP - Centre for Natural Resources and the Environment

bFEUP - Faculty of Engineering, University of Porto

Rua Dr. Roberto Frias 4200-465, Porto, Portugal

**Abstract:**

As one of the most important components of nature and the primary source of human food supply, the soil needs to conserve and remove the contaminants. In order for humans and other species to survive on the Earth, the soil must be cleaned of contaminants and preserved from continuing pollution. It is important to identify soil contaminants sources, detect soil contaminants pathways, and rehabilitate them. One approach uses satellite imagery to identify contaminated areas, contaminant sources, and transmission methods, suggest ways to prevent contamination, and allows to decide by the best rehabilitation method. Due to the advancement of technology in the field of satellite imagery and the launch of specialized satellites in the area of identifying soil pollution, more and more research is being done in this field. This study discusses an overview of soil pollution detection methods through satellite images and techniques used to detect the pollution so far. The types of machine learning methods used in satellite imagery detection will also be reviewed due to the advancement of machine learning in all areas. Finally, the types of satellites, the current limitations of spaceborne sensors for soil contamination monitoring, and how to estimate the efficiency of the methods will be discussed.

1. **Introduction**

Due to the fast-economic progress and urbanization in the last few decades, environmental pollution has become an increasingly serious issue. Exorbitant contaminants in the environment severely threaten human health by entering the food chain and relocating into drinking water sources [1].

Due to the fast-economic progress and urbanization in the last few decades, environmental pollution has become an increasingly serious issue. Excessive contaminants in the environment pose a severe threat to human health by means of entering in the food chain and migrating into drinking water sources [1]. Because of their high metal-scavenging potential, soils are the main sink for released pollutants into the environment [2]. In turn, pollutants degrade the chemical and microbiological quality of soil, [3], [4] and subsequently, because of their continual nature and long biological half-lives, create a risk to humans through their potential direct contact with contaminated soils or by transfer from soil to crop [5], [6]. The most common soil contaminants are agrochemicals (fertilizers, pesticides, and herbicides), natural gas, petroleum hydrocarbons, PTEs, acid mine drainage, and heavy metals. Regarding to sever and increasing industrial activities worldwide, significant contents of natural gas, petroleum hydrocarbons, and PTEs are considerably being diffused into all soils including the agricultural soils [1], [7]. Leaking oil and gas pipelines and petroleum hydrocarbons seepage are problematic in many areas. In the case of large undiscovered leakage, the extensive volume of explosive gases can occur in the soil [7]. The main influences of these contaminants in soil are microbiological alterations, neomineralization (e.g., calcite, pyrite), bleaching (discoloration of red soils), electrochemical alterations, and radiometric anomalies [8], [9]. According to Noomen et al. [10] depending on gas seepage length and soil-type hydrocarbons cause a vast range of alterations in the soil. Available oxygen of the soil is displaced by hydrocarbon gases [11] and is reduced by methanotrophic bacteria, [12] which affect vegetation growth and development [13], [14].

Both underground and open-cast mining activities are associated with many environmental issues such as acid mine drainage [15], PTEs release into the environment [16], and generation of a large amount of Potentially Toxic Elements (PTEs) [17]. Mining and related activities are among the industries that may release toxic elements into the environment. Accumulation of these elements in the upper soil horizons and their transfer into the food chain have adverse effects on crop yield, food quality and soil microbial groups [17]. Chromium (Cr), as one of the most harmful toxic elements, causes severe health implications such as skin and mucous membrane ulceration, allergic and eczematous skin reactions, allergic asthmatic reactions, perforation of the nasal septum and bronchial carcinomas [18]. Therefore, continuous environmental monitoring should be conducted to identify and track toxic elements, including Cr, at their early release stages. Acid mine drainage leads to accumulation of Fe minerals, which cause progressive increases in the pH [19]; pH is a principal factor in distribution and mobilization of PTE across the soil profile. Therefore, in acidic soils several PTEs including Cd, Zn, Co, Cu, and Ni are simply mobilized and easily obtainable for plant uptake [20]. PTE pollution in soils also happens because of anthropogenic activities and affects the physical and chemical features of the soil ecosystem [21], [22]. Although soil PTEs and their negative effects have become a global concern, their concentrations in some areas are still extremely high. Concentrations of PTEs in soils close to mining areas from different countries are given in Table 1.

Table 1. PTEs content (mg/kg) in mine area soils from various countries compared with global reference values and other reference soils [3]

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Location | Soil type | No. | As | Mn | Zn | Cu | Cr | Ni | Pb | Cd | Ref. |
| Oltu, Turkey |  | 19 | - | - | 35.9 | 23.4 | 135.6 | 59.8 | 34.2 | 0.03 | [23] |
| Ptolemais, Greece | Reclaimed soil | 101 | 12.3 | - | - | - | 17.5 | 10.1 | - | - | [24] |
| Douro, Portugal | Waste-impacted soil | 3 | 38.3 | 139 | 57 | 36.5 | 92.3 | 21.4 | 30.8 | 0.2 | [25] |
| Smolnica, Poland | Reclaimed soil | - | - | - | 142 | 18 | - | - | 39.9 | 1.65 | [26] |
| Pokrok, Czech Republic | Reclaimed soil | 103 | 4.48 | 599 | 25.2 | 13.7 | - | - | 18.4 | 0.27 | [17] |
| Xuzhou, China | Reclaimed soil | - | - | - | 130.9 | 52.3 | 73.4 | - | 47.4 | 3.2 | [27] |
| Guizhou, China | Mine-impacted soil | - | - | - | 135.8 | - | - | - | 42.4 |  | [28] |
| Jiangsu, China | Agricultural soils | 122 | 8.06 | - | - | 31.95 | 73.96 | - | 28.17 | 0.28 | [21] |
| Sonepur, India | Opencast mine-impacted soil | 32 | - | 3.96 | 947 | 27 | 98 | 34 | 27.3 | 0.012 | [29] |
| Ledo, India | Mine-impacted soil | - | - | - | - | - | 112 | 87.5 | 183 | 2.6 | [30] |
| Suncheon, South Korea | Mine-impacted soil | 22 | 226 | - | - | 49.3 | - | - | 88.1 | - | [31] |
| Waal, Netherlands | Floodplain soils | 69 | - | - | 80.97 | - | - | - | - | 0.68 | [32] |
| Minimum |  |  | 0.5 | 86 | 1.5 | 0.5 | 17.5 | 4.3 | 0.5 | 0.02 | [3] |
| Maximum |  |  | 38.3 | 3700 | 296 | 110.4 | 523 | 390 | 433 | 4.48 | [3] |
| Reference soil, China |  |  | 11.2 | - | 74 | 23 | 61 | 27 | 27 | 0.097 | [33] |
| Reference soil, USA |  |  | 5.2 | 380 | 55 | 21 | 41 | 15 | 17 | - | [34] |
| Soil quality for agricultural soil, Canada |  |  | 12 | - | 200 | 63 | 64 | 50 | 70 | 1.4 | [35] |

Moreover, Heavy metals are metals with an atomic mass of over 55.8 g/mol or a density of over 5 g cm-3 [36], arsenic, mercury, zinc, lead, cadmium, chromium, copper, manganese, nickel and vanadium are among the most important harmful rare elements found in the biosphere [37]. Sphalerite is the main ore of zinc which is found in three forms including sphalerite, wurtzite and matraite [38]. The average level of zinc present in the lithosphere was estimated to be 80 mg kg−1 and the natural value of zinc in soils was stated as 10–300 mg kg−1 (Kabata-Pendias, 2010). The concentration of lead in soils is between 1 and 200 mg kg−1, on average is 15 mg kg−1 and its critical limit is 50 mg kg−1 [39], cadmium and lead are two metals resulting from combustion activities and transportation and are linked to zinc resulting from abrasion of tires. Cadmium and lead are considered to be the most mobile and least mobile elements in soil with typical values of 0.06–1.1 and 2– 300 mg kg−1, respectively [37]. The maximum allowable limit for cadmium in agricultural products has been reported to be 0.1 mg kg−1 and in non-agricultural crop, this value should not exceed the allowable limit [40]. It should keep in mind that, heavy metal-induced stress in rice crops can be characterized by spatiotemporal continuity and stability, as is the salty soil condition [41]. Other stressors (e.g., drought, pests, diseases, and mismanagement) are typically more spatially and temporally transient [42], [43].

Research on soil contamination and soil changes due to contamination is getting more attention, particularly in the soil ecosystems reclamation and sustainable use. For the sake of economic, environmental, and health perspectives, monitoring and analyzing these alterations are necessary for residents, decision makers, and environmental-observers. Conventional methods for determining the soil condition in vast areas require field sampling, chemical analyses in a laboratory and geostatistical interpolation, which are time-consuming and expensive [44]. For instance, according to Shi et al., [1] because of limited funds, investments on PTE decontamination have relatively lagged behind or have even been totally ignored, despite the prevalence of PTEs in soils. Moreover, monitoring of contaminated sites is typically conducted by the respective industries such as the petroleum industry using temperature, pressure, and flow changes methods along the pipeline, which have inherent risks and rely on the accuracy of the experts [45]. In addition, these methods are inefficient at detecting small soil changes and therefore, benefits of detecting small changes, controlling them before they become large and able to cause greater impacts, will be neglected [9]. However, any significant alteration of soil condition needs to be carefully evaluated using available high-tech sensors and techniques for early detection of soil status due to soil contamination. These sensors including proximal and remote sensing techniques, which differ in their radiometric, temporal, spectral, and spatial resolution and as a result, in their monitoring ability [46]. The spatial resolution may be millimeters (drone-based cameras), 0.5–2 m (airborne and some hyperspectral sensors such as AISA, HySPEX, APEX), 2–10 m (some satellite sensors such as WorldView-2, RapidEye, and Sentinel-2), 10–30 m (some satellite sensors such as Spot and Landsat), and up to 250–1000 m or greater (MODIS and NOAA AVHRR) depending on the sensor’s platform [47]. In this review article, satellite monitoring techniques is presented.

1. **Predictors and underlying mechanisms**

Recently, researchers have explored the use of innovative tools that make the detection of soil contaminants easier and faster, thus enabling higher resolution prediction of contamination levels [48]. An emerging method is known as visible and infrared reflectance spectroscopy (VIRS), which involves in the field measurement of contaminants from either a handheld portable device, unmanned aerial vehicles (UAVs), or even satellites, for fast remote sensing of large spatial areas [47], [49]. The visible reflectance spectrum (VIS, 380e750 nm), near-infrared spectrum (NIR, 750e1300 nm), short wave inferred spectrum (SWIR, 1300e2500 nm), mid-infrared spectrum (MIR, 0.25e2.5 mm) and long-wave infrared spectrum (LWI, 8e12 mm) have all been applied for VIRS based soil monitoring [50]. The use of this sensing technique can accelerate soil pollution mapping at high resolution with less expense and time than other soil sampling approaches. In this review article, satellite monitoring techniques is presented.

The use of (visible and infrared reflectance spectroscopy) VIRS relies on the fact that atoms and molecules absorb and emit electromagnetic radiation because of electro transition and molecular vibration [51]. Identification and quantification of different chemicals can be achieved based on emission and absorption spectra. In soil contamination monitoring, VIRS captures reflectance energy from the land surface with the reflectance spectra informing us of the soil composition [52]. Certain organic soil contaminants, such as polycyclic aromatic hydrocarbons (PAH) and petroleum hydrocarbons (collectively termed total petroleum hydrocarbons (TPH)), are often detectable in visible and infrared reflectance spectra [53], [54]. In the case of heavy metals, direct monitoring can only be achieved at concentrations that rarely occur in the field (e.g., 4000 mg/kg in the case of Cd) [55]–[57]. Fortunately, interactions between trace levels of heavy metals and more abundant soil components (e.g. clay, organic matter and Fe oxides) provides an opportunity to detect them indirectly [58], [59]. Another way of detecting trace levels of metals is to monitor vegetation spectra because of the influence contaminants exert on plant physiology [50]. Specific mechanisms for predicting soil pollutants are introduced in this section.

In [60] VNIR used as Visible/near-infrared reflectance. In Remote sensing, VNIR spectroscopy have been used as rapid and practical approaches for determining the distribution of heavy metals [44]. Many researches have confirmed that high concentrations of heavy metals represent electromagnetic spectrum features within the VNIR region [52] and these spectral characteristics can be employed to estimate heavy metal concentrations [61]. Iron oxide, clay and organic matter contents of the soil affect spectral intensity [52], [62], [63], which can have an impact on the predictability of soil properties [64]. It is suggested the utilization of VNIR ranges (400–2500 nm) of electromagnetic spectrum for the detection of Pb, Cr, V, Ti, Cu, Zn and Mn in the soil. They demonstrated the positive relationships between the digital values of a satellite image within 400–510 nm and Pb, Zn and Cr between the ranges of 580–625, 630–690, 770–895 nm and other heavy metals. The linear relationships provided for Zn, Cr and Pb demonstrated an R2 of 0.95, 0.85 and 0.83, respectively.

* 1. **Molecular vibration**

In the case of organic compounds, stretching and vibrations of aliphatic (alkyl) compounds and certain functional groups can often be observed in near-infrared spectrum (NIR, 750e1300 nm) and mid-infrared spectrum (MIR, 0.25e2.5 mm) [54], [65]. The first overtone of TPH is observed in the wavelength range of 1600-1820 nm, and the second at 1100-1500 nm. Observation of the second overtone is more difficult if TPH concentrations are relatively low [66]. In the case of PAHs, the first overtone of C-H stretching and deformation of C-H combination, and the second overtone of C-H stretching in aromatic C-H are observed at wavelengths of 1675 nm, 1417 nm and 1097 nm, respectively [67]. In MIR region, the peaks around 1630-1580 , 1930-1840 and 2060-1930 are associated with aromatic functions [68], [69]. The concentration of TPH in soil samples collected from oil contaminated sites can be determined by Vis-NIR spectrophotometry, with absorption peaks around 1712 nm,1758 nm and 2207 nm [54]. The 1712 nm and 1758 nm peaks are in the first overtone region, which are attributed to the stretching of terminal CH3 and saturated CH2 in alkyl [70]; the 2207 nm peak is associated with either amide (C=O) or the stretch and bending caused by crude oil [71]. Okparanma et al. [72] demonstrated that PAHs in soil are detectable at a wavelength of 1670 nm, which was attributed to aromatic C-H. The calibration R2 value for their PAH prediction model was 0.89, and the PRD reached 3.12. Observed spectra for organic contaminants may overlap with soil organic matter (SOM), but the presence of SOM would not normally influence TPH detection [69]. This is because TPH consists of medium length chains, whereas SOM mainly composes of long -CH2 chains, and relatively low amounts of -CH3 (Forrester et al., 2013). For example, it has been found that spiking TPH contaminated soils with SOM has little effect on observed NIR absorption spectra, but it may affect the MIR region (especially 1980, 1870, and 1790 peaks) [69]. Forrester et al. [65] noted several characteristic absorption peaks in the spectrum of TPH contaminated soil with the presence of SOM, which were attributed to the vibrational overtone of terminal methyl in the MIR region. The presence of such peaks can fortuitously aid TPH detection.

* 1. **Soil properties**
     1. **Soil organic matter**

Soil organic matter (SOM) derives from the breakdown of plant and animal debris. Many studies have shown that the combination of molecular vibration and overtones in SOM, including O-H, C-H, C=O groups, can be identified in Vis-NIR spectra [32]. Because humic and fulvic acids in SOM bind with heavy metal cations, through COOH, OH, and C=O interactions [73], correlation between SOM and heavy metals levels has been observed [74]. Several studies have exploited SOM spectral bands to predict heavy metal concentrations in soil. For example, at an agricultural site contaminated by polluted irrigation water, it was found that Cd levels were positively correlated with SOM. Measurement of 410, 581-626, and 670-690 nm wavelengths were found to be effective for predicting Cd levels [75]. Chakraborty et al. [76] used VIS-NIR spectroscopy to determine As concentrations using the absorption feature associated with O-H and C-H bonds in SOM at a wavelength of around 1290-1310 nm.

* + 1. **Fe-oxides**

Iron oxides and hydroxides are widely found in the earth’s surface, especially iron oxyhydroxide (goethite), which forms from weathered iron-rich minerals [52], [56]. Because Fe-oxides are characterized by high surface charge, large surface area and strong adsorption capacity, they play a crucial role in the fate and transport of heavy metals in the subsurface [77]. For this reason, concentrations of soil heavy metals often correlate to those of Fe-oxides [56]. VIRS detection is possible because various peaks, including 565, 435, 500 nm and bands between 650 and 760 nm (Fig. 1), have been associated with Fe-oxides, with significant correlations identified with soil heavy metals [57]. Kemper and Sommer [78] found that As closely correlated with the reflectance of Fe oxide related bands at ~550 nm wavelength. Wu et al. [79] reported that Ni concentrations can exhibit a negative correlation with iron oxides, especially in the 480-580 nm wavelength region. Chakraborty et al. [76] reported that As concentrations had a strong correlation with Fe oxides, meaning that high levels of regression fitness with diffuse reflectance data could be achieved.

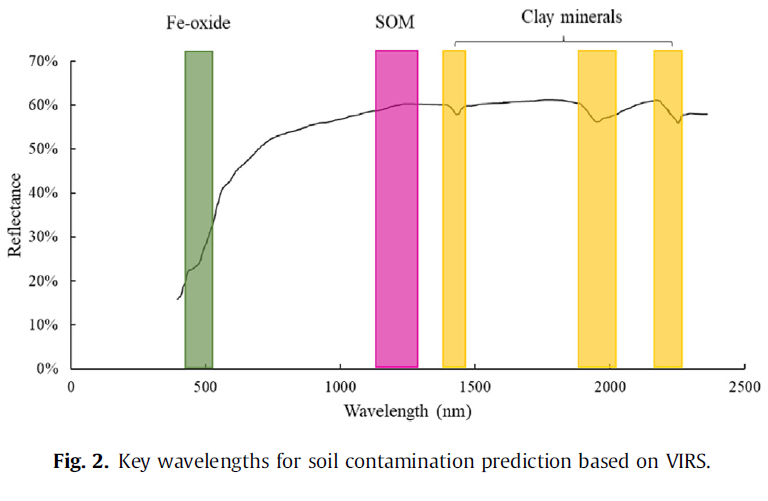


Fig 1. Key wavelengths for soil contamination prediction based on VIRS.

* + 1. **Clay minerals**

Hydroxyl absorption associated with molecular water can be detected at 1400 nm, 1900 nm and 2200 nm, which is associated with clay minerals [59], [78], [80]. Bands at 538 nm wavelength correspond with the Si-O and Si-O-Al bonds in clay minerals [21]. This is important for soil contamination surveys because the cation exchange capacity (CEC) of clays minerals are often high, meaning that heavy metals cations can easily replace clay mineral cations. Heavy metals tend to sorb to clays by van der Waals forces and hydrogen bonds [81]. Concentrations of heavy metals in mine tailings can correlate with bands at 1400 nm, 1900 nm and 2200 nm [78]. Choe et al. [31] found that As levels had a statistically significant (p < 0.006) correlation with reflectance at 2200 nm. The calibration R2 value was 0.56. Song et al. [21] found that Cu displayed the highest correlation at 538 nm, which was related to Si-O bands, with an R2 value of 0.551 (p < 0.001). A positive correlation between Hg concentration and adsorption at 2210 nm was reported by Wu et al. [58].

* 1. **Vegetation**

On sites covered by dense vegetation, optical remote sensing remains ineffective for detecting oil leakages directly, because light penetration is strongly limited by the foliage and the spectral signature of soils is thus not accessible. The only information about soil composition can be provided indirectly by vegetation through its optical properties [82]–[84]. Consequently, unfavorable growing conditions in soils result in modifications of vegetation health and optical properties that can be tracked using hyperspectral remote sensing [10], [85], [86]. Therefore, since crude oil and petroleum products affect vegetation health, they can be detected and quantified indirectly using optical imagery [87]–[90]. To achieve this, several conditions must be fulfilled: (1) The contamination must affect the biophysical and biochemical parameters of vegetation, (2) alterations in these parameters must modify the spectral signature of vegetation and (3) the specifications of imaging sensors (*e.g.* the spatial and spectral resolutions) must make it possible to track these alterations.

Wavelengths around 540, 690, 730, and 780 nm are closely associated with chlorophyll-a/-b contents in plant leaves and pigment composition [91]. Leaf anatomical features, including mass per area and structure differences (i.e., cell morphology and parenchyma structure) can present significant correlation with NIR peaks [92]. By combining VIS, NIR and shortwave infrared, the water content in vegetation can be monitored (Fig 2.) [93]. Because pigments, anatomical features, and plant water content relate to plant health [1], vegetation reflectance can be used for assessing soil contamination levels [94]. Changes to the physicochemical and biological properties of soils also cause an effect on vegetation reflectance [87], [95], [96]. Shi et al. [51] explored the reflectance of rice plants to predict soil As concentrations. It was found that 768, 939, 953, 1132, and 1145 nm wavelengths correlated to As levels, while 768, 939 and 953 nm wavelengths were related to the leaf area index and chlorophyll density, and 1132 and 1145 nm wavelengths were associated with the cellular structure, which could be used for indirect measurement of As levels. A partial least squares regression (PLSR) model was developed with an R2 of 0.77 [97]. Two-band and three-band vegetation indices have been used to predict As levels by leaner and polymeric regression models. The three-band index (-)/(-) is the more effective of these [50]. It should be noted that environmental factors unrelated to soil contaminant levels (e.g., nutrient availability) may affect the health of plants and should be considered when relying on vegetation reflectance data [87]. Moreover, the sensitivity to contaminant exposure is different for different plant species [9], [87].

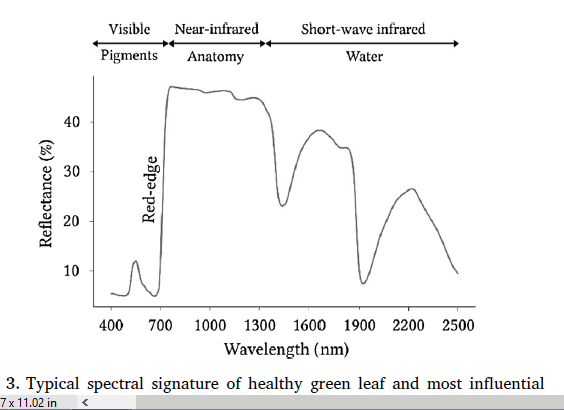


Fig 2. Typical signature of healthy green leaf and most influential parameters in the different spectral regions

In most cases, studies aimed to assess the impact of crude oil and petroleum products on the environment using multi- (Landsat) or hyperspectral (Hyperion) satellite imagery at 30-m spatial resolution a very high spatial and spectral resolutions are needed to achieve efficient discrimination of oil and other stressors (Table 2) [98].

Table 2. Studies aiming to detect and quantify crude oil and petroleum products using multi- and hyperspectral airborne and satellite images. (Refl.: Reflectance; VI: Vegetation Indices; CR: *Continuum* Removal; RF: Random Forest; REP: Red-Edge Position; comp.: Comparison; RTM: Radiative Transfer Model.) [98].

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Vegetation type** | **Target** | **Sensor name** | **Sensor type (spatial**  **resolution)** | **Bands (spectral**  **domain)** | **Method** | **Ref.** |
| **Multispectral** | | | | | | |
| Mangrove | Crude oil leakage | Landsat-8 | Satellite (30 m) | 9 (435–2294 nm) | VI+Mean comp. | [99] |
| Crops, grassland & trees | Crude oil leakage | Landsat-8 | Satellite (30 m) | 9 (435–2294 nm) | VI+RF classification | [100] |
| Mangrove | Crude oil leakage | Landsat-5 & -7 | Satellite (30 m) | 6 (450–2350 nm) | VI+Simple regression | [101] |
| Mangrove | Crude oil leakage | Landsat-5 & -7 | Satellite (30 m) | 6 (450–2350 nm) | VI+Mean comp. | [101] |
| **Hyperspectral** | | | | | | |
| Tropical forest | Crude oil leakage | Hyperion | Satellite (30 m) | 242 (400–2500 nm) | VI+Threshold | [86] |
| Plain & rainforest | Crude oil leakage | Hyperion | Satellite (30 m) | 242 (400–2500 nm) | CR+Mean comp. | [102] |
| Plain & rainforest | Crude oil leakage | Hyperion | Satellite (30 m) | 242 (400–2500 nm) | Refl. & VI+Mean comp. | [102] |

Experiments carried out under controlled conditions are a necessary first step, since they help determining the response of vegetation specifically induced by crude oil and petroleum products. These experiments must be representative of realistic field conditions (i.e. species, TPH concentrations) and serve as basis for developing classification or regression methods that are suitable for use on images [98].

Fig 3. Three main steps of imagery detection of soil contaminants through the vegetation [98].

The upscaling of methods is the most important difficulty in this approach, so it is crucial to address it progressively; for example, from leaf to canopy scales and finally on images. The validation of the methods in the field is an intermediate – and critical – step prior to imagery application. Then, the methods should be progressively applied to imagery; first, on selected sites with known species’ sensitivities, and thereafter at large scale (Fig 4.) [98].

In [55] to monitor rice under different stressors accurately at the region scale, the following method was implemented. (i) Ground measurement data were used to derive the vegetation indices (VIs) sensitive to rice biophysical parameters (i.e., chlorophyll and LAI). (ii) Sensitive VIs were applied to screen the unstressed rice and stressed rice. (iii) Spatio-temporal analyses were performed to detect rice under stable stress and abrupt stress. (iv) The spatial distribution of rice under different stressors was mapped based on the GIS analysis method [55].

* + 1. **Vegetation spectral signatures**

The spectral reflectance of vegetation is mainly a function of tissue optical properties, canopy biophysical properties, soil reflectance, viewing geometry and illumination circumstances [103], [104]. Reflected electromagnetic radiation can deliver informative data about the plants’ condition and enable researchers to remotely analyse the plants’ physiological and chemical characteristics [105], [106]. According to [107] regardless of plants species, the spectral assignments of leaves are similar in the optical spectral ranges of the electromagnetic spectrum, visible–near infrared–shortwave infrared (VIS–NIR–SWIR). Figure 4 shows the typical spectral response features of green vegetation in detail. According to [108], dominant factors controlling vegetation reflectance can be summarized to photosynthetic pigments (e.g. chlorophyll, carotenoids and xanthophylls), cell structure, leaf water content and biochemical concentrations (e.g. lignin, cellulose, starch and protein). There is a strong interaction of light with pigment concentrations in the VIS, strong reflectance and transmittance by the leaf structure in the NIR and strong absorption features of water in the SWIR [107], which all cause lower reflectance [85]. There are two main absorption features in blue (0.45 μm) and in red (0.67 μm) bands that are associated with the two main leaf pigments (chlorophyll a and b). These strong absorption bands induce a reflectance peak in the yellow-green (0.55 μm) band [109]. Sanches et al. [9] mentioned that during leaf senescence, chlorophyll decreases, which leads to a reflectance rise within the red wavelengths. Hence, the spectral response in the VIS range is influenced by carotenoids that absorb blue light and reflect green and red light, which causes yellowing of green leaves. Progressive senescence also degrades carotenoids that induces an escalation of leaf reflectance within the blue wavelengths [110]. According to [85], a decline in chlorophyll content shows the following spectral responses: (1) a reduction in the infrared shoulder height, (2) a reduction in the maximum absorption and (3) a shift in the red-edge (RE) position towards shorter wavelengths. Hoffer [109] and Horler et al. [111], mentioned that cellulose and leaf pigments are transparent to NIR wavelengths and thus in this region, leaf structure explains the optical properties [45]. And, also in [112] stated NIR reflectance is highly affected by cell size, cell layers and mesophyll thickness. In fact, in this range, there is a typical reflectance plateau in the leaf spectrum, and its level depends on the internal leaf structure. The other optical domain in the SWIR is characterized by the vegetation’s water light absorption, for which there is a negative correlation between reflectance and leaf water content in this region. It can be concluded that for all three of the aforementioned spectral domains, factors that influence leaf optical features can be internal from the leaf itself or external from the environmental conditions, which impact on vegetation health [113].

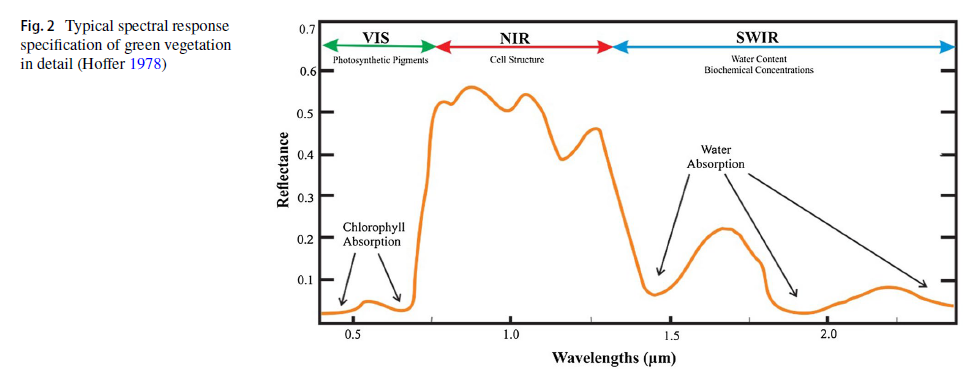


Fig 4. Typical spectral response specification of green vegetation in detail [109]

* + 1. **Vegetation spectral response to soil contamination**

The spectral signatures of vegetation grown in contaminated soils have been investigated in order to understand the role of reflected radiation of the VIS–NIR–SWIR range in detecting plant stress due to soil contamination. Bammel and Birnie [114] studied the possibility of using sagebrush spectral response to explore hydrocarbons. They concluded that the blue shift of the RE is the most efficient representative of hydrocarbon-induced stress in sagebrush. Van der Meer et al. [8] proved that the presence of gas seeps and hydrocarbons also has effects on surface mineralogy and causes increasing spectral responses in the VIS–NIR–SWIR range in anomalous vegetation. The hydrocarbons influence the vegetation root structure and capabilities and hence the spectral features. Smith et al. [115] showed that vegetation grown in soil with high natural gas had increased reflectance in the VIS and decreased reflectance in the NIR–SWIR due to shortage. Noomen et al. [7] proved that ethane (contains up to 10% of natural gas) affects the reflectance in the chlorophyll and water absorption ranges as well as the RE shape and position. Displacement of soil by nitrogen, natural gas, argon and waterlogging was introduced as a reason for these alterations [7]. Souza Filho et al. [116] investigated eucalyptus grown in non-contaminated and contaminated soils with hydrocarbons. They showed that more polluted soils caused higher vegetation reflectance in the VIS and SWIR and lower reflectance in the NIR wavelengths, compared to plants grown outside the contaminated area. Sanches et al. [9] also showed a significant change in reflectance, particularly in the VIS and infrared (leaf and canopy spectra, respectively), in brachiaria plants contaminated with diesel and gasoline. It is clear that the spectral detection of hydrocarbons depends on the level of contaminant and the stage of exposure to a contaminant [9], [117]. The results of studies on the effects of heavy metals on plants’ spectral responses were dissimilar in some cases, as the effects depend on the region and type of plant [118], [119]. Kooistra et al. [120]observed that Pb, Zn, cadmium (Cd), nickel (Ni) and copper (Cu) concentrations in river floodplain soils hold reasonable relationships with the reflectance spectra of grass. In the study of Clevers et al. [121], the maximum FD of grassland spectral features in the RE resulted in a significant, negative correlation with the heavy metal contents in the soil. They stated metals bring lower chlorophyll concentrations; thus, a higher red reflectance is expected. Li et al. [122] used the reflectance spectra of rice plants for successful estimation of soil As contents via their relationship with chlorophyll contents of canopy. Plants growing in heavy metal-polluted areas show a water content drop, chlorophyll hydrolysis rise and cellular structure damage. These variations affect the reflectance spectra of vegetation leaves and canopies, thereby offering a basis for indirectly predicting soil heavy metal concentrations [50].

Based on the previous results, the primary effect of Cd induced stress on rice relates to changes in color and morphology and a corresponding reduction in chlorophyll and LAI [123], [124]. The change of these symptoms in plants can be detected by remote sensing directly [125]. Thus, the shifts in the vegetation spectra of plants under Cd stress occur in the near-infrared part of the spectrum [120]. Therefore, five VIs based on the red edge wavelength were selected for this study: the red-edge chlorophyll index (CIRE), the red edge position (REP), the MERIS terrestrial chlorophyll index (MTCI), and two normalized red edge differences (NDRE1, NDRE2).

The spatio-temporal anomaly is referred to as a spatial-temporal object with thematic attribute values that are significantly different from those of other spatially and temporally referenced objects in its spatial or/and temporal neighborhoods [126], [127]. Therefore, spatio-temporal anomaly detection methods may be feasible for distinguishing between the rice under abrupt stress versus that under stable stress.

Changes in crops VIs were confirmed to be affected by the intrinsic crops growth trend and stable and abrupt stressors [128]. In general, the intrinsic crops growth trend and variation induced by stable stressors have dominant impacts on spatio-temporal shifts of crops spectra within the growing season; abrupt stressors are not constant in space and time, and affect only local or short-term spatiotemporal shifts of crops spectra. In order to quantify the period-toperiod variability of a given index time series (Qi), Qi for each two periods is formulated simply as:

Eq. 1

where , , and are the VIs variation, and the values of the i-th pixel VIs for period m and m+1, respectively. They delineate , − of two different periods, in order to facilitate our interpretation and discussion, with subscripts indicating the range of the DOY used, such as DOY (the time is expressed as the day of the year, DOY; 12/07/2017 is thus called ), (m time is , m + 1 time is ). In [55], the coefficient of spatio-temporal variation (CSTV) is computed. The CSTV is expressed a follows:

Eq. 2

where , , and δ are the i-th pixel value, area-averaged mean value, and standard deviation of Q along a given index, respectively. Thus, CSTV is a useful descriptive parameter for capturing the spatiotemporal variation of VIs. Q and δ are computed as follows:

Eq.3

Eq.4

where j is the total pixel number of Q. The CSTV is actually normalized by a mean and a standard deviation of one (δ) to facilitate the identification of areas where VIs are higher or lower than normal for a given time interval within the growing season. When crops are stressed by stable stressors or are not stressed, the CSTV has a lower value because the VIs change uniformly in this area [50]. However, a higher CSTV indicates that crops are stressed by an abrupt factor during some growth stage(s) [129]. In [55], previous reports on the criteria for spatiotemporal anomalies [127], [130], the CSTV calculation was used with a 95% confidence level defining “stable stress”.

* + 1. **Spaceborne Sensors technique**

Obtaining important information for vegetation stress detection is possible using DRS and IS proximal sensing, UAV and airborne imaging. A pixel spectrum is generated for the potential differentiation of target vegetation attributes using hundreds of connected spectral bands with a narrow bandwidth. Nevertheless, cost-effective and repeated mapping of the vegetation surfaces, efficient and spatially widespread mapping and large-scale monitoring of stressed vegetation can be obtained using spaceborne sensors. The main advantages of the spaceborne approach remote sensing are summarized in the accessibility of high-temporal images, frequent revisit-time, the comprehensive monitoring of large-scale sites, data reduction and efficient classification of results [131], [132]. In recent years, the expansion of spaceborne technologies and applications has attained advantages due to the opening of large satellite information archives to the public (e.g. Landsat) [133], including integrated space missions evolved for the public domain (e.g. the Sentinel missions of the European space agency (ESA)) [134] and the advancement of open source tools for remote sensing data processing [135]. It is expected that such developments will lead to a massive use of satellite data-based techniques for understanding vegetation health, which is based on soil condition. However, the inclusion of satellite remote sensing information with lower spectral and spatial resolution domains into vegetation stress detection are still lacking, essentially due to shortage of appropriate sensors. The capability of satellite platforms for monitoring and determining the vegetation status due to environmental stressors including contamination has been proved theoretically and practically. For instance, Zurita-Milla et al. [136] stated that data fusion of the unmixing-based Landsat and MERIS FR data will be able to successfully assess the stressed vegetation status by evaluating NDVI, the modified transformed chlorophyll index (MTCI) and the modified green vegetation index (MGVI). Misurec et al. [137] detected spatio-temporal changes of forest stands in Ore Mountains, Czech Republic, which suffers from severe environmental pollution. They used the disturbance index (DI) derived from Landsat time-series together with VIs derived from airborne hyperspectral imageries (ASAS obtained in 1998 and APEX obtained in 2013) and concluded that though the alterations between initially moderately-to-heavily harmed and originally damaged stands largely levelled out by 2013, it was yet feasible to identify symptoms of the earlier damage in some cases. Arp [138] studied indirect detection of hydrocarbon seepage using anomalous vegetation. They used Landsat imagery, which showed an anomaly in sagebrush (Artemisia tridentata) fields located in a region of hydrocarbon microseepage. The results proved that the sagebrush anomaly resulted from the rising of gases and waters used to retain reservoir pressures in the field, which produced anoxic, high-pH, high salinity and low-Eh soils [138]. Recently, the potential of hyperspectral spaceborne sensors for detecting hydrocarbon pollution impacts on vegetation has also been studied. Arellano et al. [86] demonstrated the suitability of Hyperion satellite imagery for assessing contamination by oil in the Amazon forest. They highlighted that levels of chlorophyll content, foliar water content and leaf structural changes were decreased in hydrocarbon- polluted tropical forests. To map this effect over broader geographical areas, NDVI was applied to hyperspectral Hyperion satellite imagery and was introduced as a suitable index for petroleum pollution impacts monitoring in forest. The forthcoming sensors from space will generate large data streams for land monitoring, which will shortly become accessible to different user communities [47], [86], [139]. It is expected that in near future, the Italian PRISMA sensor will be available, followed by the Japanese HISUI HSR sensor (with thermal capability), the German EnMap HSR (with free data to the scientific community); thereafter, the Italian-Israeli SHALOM sensor would be in orbit. These sensors, plus the new initiatives such as FLEX sensor (for monitoring fluorescence in vegetation) and Sentinel-2 (with three RE bands) promise that high-quality data will be more frequently available for monitoring vegetation and their stressors from orbit. All the above-mentioned materials showed that the disciplines of optical proximal and remote sensing are experiencing an inimitable increasing in sensors quantity and quality. Moreover, the high potential of the techniques in assessment of vegetation status as indicators of soil contamination has been proved; however, each technique has some advantages and disadvantages, which determine its capability. Therefore, selection of the data source highly depends on the measured attribute, resolution requirement, turnaround and revisit-time, cost and value of the information and data processing requirement [140].

The average reflectance curves between the unstressed and stressed rice are shown in Fig. 5. For ASD reflectance spectra, a large difference in the original spectral reflectance was observed in the visible and near infrared region; namely, the reflectance of unstressed rice was higher than that of stressed rice. For Sentinel-2 reflectance spectra, regardless of whether B02, B03, or B04 was the focus, no clear difference was observed in the mean values of unstressed and stressed rice, indicating that there were no visible symptoms of crops stress in the original reflectance spectra from B02 to B04. From B06, B07, and B8A, the notable common feature was that the spectral curve of unstressed rice was slightly higher than that of stressed rice [55].

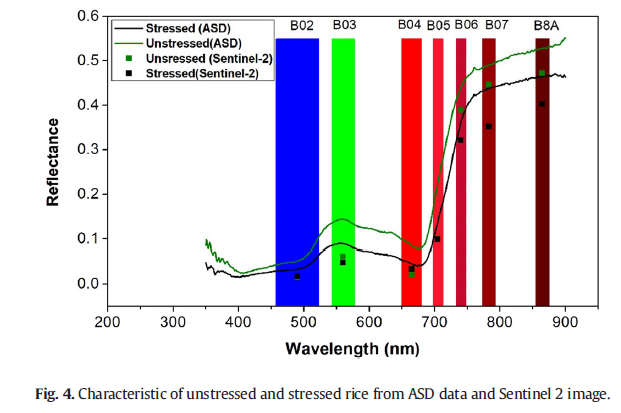


Fig. 5. Characteristic of unstressed and stressed rice from ASD data and Sentinel 2 image.

Liu et al. [55] data was based on the same or similar crops (such as first- and second-season rice crop) in similar growth stages, and images of pure paddy fields were also selected to limit variability. Second, the CSTVs of VIs can vary significantly because of differences sowing date, cultivars of rice, climatic condition, and fertilization strategies, it remains a challenge to discriminate all these factors that potentially influencing crop growth at a large scale. Third, high-resolution spatio-temporal images are required to map crops regions and obtain accurate times and locations of crops under different stressors, especially for abruptly stressed crops [55].

At the same time, using multi- and hyper-spectral sensors I conjunction with SAR data can more effectively monitor crops stressors. Ideally, if clouds and cloud shadows are adequately screened, multi-sensors can exclude large rates of false detection of abrupt stressors because most surface reflectance change is caused by clouds, which have different spectral effects compared to those caused by abrupt stress [55].

* 1. **Factors affecting VIRS detection**

Several factors, including contaminant concentrations and other soil components (e.g. SOM and clay minerals), affect VIRS detection. Because the contaminant concentration determines how much energy is absorbed and emitted, the higher the soil contamination level the easier it is to interpret reflectance spectrum directly [67], [141]. Wu et al. [56] noted that when concentrations of Cr and Cu were higher than 4000 mg/kg, adsorption could be discriminated at wavelengths of around 610 and 830 nm, respectively. However, detection was not possible at concentrations below 1000 mg/kg. Moreover, when contaminant concentrations are limited, spectral peaks may shift from their usual wavelength positions [141]. Because most soil heavy metals only exist in trace amounts, they must be monitored indirectly. The predictability of trace levels of heavy metals varies depends on them in situ behavior. Kemper and Sommer [78] found that Pb could be predicted with a high R2 value (0.940), followed by Hg and As with R2 values of 0.929 and 0.858, respectively. The R2 values for Cd, Cu and Zn were much lower [78]. Hou et al. [142] monitored six heavy metals through hyperspectral VIRS detection, finding that prediction accuracy decreased in the order of Ni > Zn > Pb > Cu > Cr > Cd.

In general, detectability relates to the affinity of contaminants to different soil components [57], [143]. Song et al. [21] found that the detection of heavy metals in soil was correlated to their affinity to Fe2O3, Al2 O3 and SOM [56]. reported that goethite detection (at 500 nm) was positively correlated with various heavy metals. All the predictors mentioned before have certain advantages and disadvantages. Molecular vibration is widely used to predict organic pollutants, but the key wavelengths will shift while the pollutant concentration changes [141]. Soil property-related predictors (i.e. Fe-oxide and clay minerals) are mainly used to predict heavy metals in soil, since heavy metals tend to sorb to them and significant statistic relationship are identified between them [56]. However, Douglas et al. [144] have observed that correlation between the contents of soil property-related predictors (e.g. organic matter and clay) and TPH were not significant, and have obtained similar conclusion. Therefore, the use of soil-property-related predictor in predicting soil organic concentration is limited. Vegetation can potentially indicate the extent of soil pollution, which is a crucial indicator especially when spaceborne spectrometers are employed [97]. However, compared to molecular vibration and soil components, vegetation, incorporates more unstable factors when used to predict soil pollution, such as their different ability on indicating pollutants contents [87]. Further explorations are requisite to interpreting the relationship between vegetation spectrum and soil pollution concentration.

1. **Contaminants detection by Spaceborne spectrometers**

As mentioned before, Spaceborne spectrometry is an efficient, economical and increasingly accessible approach to soil mapping [145]. Earth observation satellites with high resolution sensors and high numbers of spectral bands have been launched, including the European Space Agency’s (ESA’s) Sentinel satellites, NASA’s Landsat program, and China’s HJ-1 [47], [146]. NASA’s Landsat-8 is equipped with an operational land imager and thermal infrared sensor, covering the 11 wavelength bands (4 visible, 1 near-infrared, 2 shortwave infrared, 1 panchromatic, 1 cirrus and 2 thermal infrared) [146]. ESA’s Sentinel-2 satellite has 13 bands, covering 443-2190 nm wavelength [139]. The HJ-1 satellite is equipped with a hyperspectral sensor with 115 spectral bands in the range of 450-950 nm [147]. Peng et al. [148] and Guan et al. [145] used the spectral bands of the Landsat-8 satellite, involving brightness and normalized difference vegetation index (NDVI) with land features (e.g. elevation and slope), to predict the concentrations of As, Cr, Ni, Pb and Zn. Liu e al. [149] used spectral data from the HJ-1 satellite to predict soil concentrations of Cd through multiple nonlinear regression, achieving an R2 of 0.81. Advanced hyperspectral satellites that will provide higher accuracy are due to be launched in the coming years, including the HyspIRI satellite with 214 spectral bands, the CCRSS satellite with 328 spectral bands and the EnMAP satellite with 242 spectral bands [47].

* 1. **Acid mine drainage**

For the first time, Goetz and Rowan[150] carried out geological and soil remote sensing based on satellite data, using a Landsat multispectral scanner (MS) to create iron oxide maps. Since 1995, ASTER was also employed to provide satellite data for geological and soil remote sensing to map surface soil [151]–[153]. According to Mars and Rowan, [154] the ASTER SWIR bands allow for effective mapping of soil due to their relative fine spectral resolution. For monitoring mine tailings site employing simulated 4 broad-band IKONOS image and the full 65-band hyperspectral data (CASI), Levesque et al. [155] observed the similar results. Chevrel et al. [156] studies the potential of the ASTER data in three sites (South Africa, Dominican Republic, and Mexico) and concluded that using the ASTER can bring an invaluable contribution in characterizing, identifying and mapping land use, mining residues, and mining effluents in mining areas. They related this capability to well-targeted spectral bands and the TIR spectral range covering. In [157], the image of the Sentinel-2 satellite was used to evaluate the heavy metals pollution of the area irrigated with wastewater in south of Tehran. In [158], GF-5 hyperspectral satellite images, as well as in-situ sampling data, were carried out to estimate the concentrations of Zn, Ni, and Cu at an open cast coal mine, China.

* 1. **Petroleum hydrocarbons**

Arellano et al. [86] used Hyperion satellite with a spatial resolution of 30 m with each pixel covering the spectral range, 400–2500 nm, for detection of contamination with oil in soils of the Ecuadorian Amazon rainforest. At those sites, crude oil had affected the surroundings and despite the evaporation of gaseous hydrocarbons, liquid hydrocarbons transferred from the open pits to the soil and water. Their investigation demonstrated the suitability of the use of the Hyperion spaceborne image for the detection and characterization of hydrocarbon pollution in tropical forests soils [86]. In [87] Lassalle et al., used hyperspectral remote sensing as a tool for monitoring soil contamination in densely vegetated oil and gas exploration and production regions.

* 1. **Heavy Metals**

Fortunately, remote sensing technology using visible and near-infrared reflectance (VNIR) spectroscopy with relatively continuous spectra and wide spectra coverage supplies a new perspective for convenient and economical monitoring heavy metals contents at large scales [159]. Published researches indicated the capacity of estimating soil heavy metals contents based on remotely sensed images collected by various sensors, including the Earth Observing (EO-1) Advanced Land Imager (ALI) [160], Landsat 8 Operational Land Imager (OLI) [161], HuanJing-1A (HJ-1A) Hyper Spectral Imager (HSI) [149] and Sentinel-2 Multispectral Instrument (MSI) [162].

There are also some soil and geology studies that have been based on the simulated imagery and the synthetic satellite images. For instance, Mielke et al. [163] combined the capabilities of hyper and superspectral spaceborne sensors for soil mapping and monitoring. They explored the potential of spaceborne sensors OLI, Sentinel-2, and EnMAP for spatial extent of mine waste surface mapping. The mines, gold and platinum, have been extracted for about 90 years and contained trace elements (U, Pb, and Cr). They suggested a new index, the iron feature depth (IFD) acquired from OLI data, to map the 900 nm absorption feature for monitoring the spatial extent of mine waste. The mean accuracy for mapping was as follows: EnMAP 100%, Sentinel-2 94.5%, and OLI 92%. Therefore, they proved that Sentinel-2 and EnMAP data might be employed as two sensors for cost saving in mining areas remediation [163]. Rogge et al. [164] also used the simulated the EnMAP scene to assess the sensor’s capability to map Ni, Cu, and platinum group elements. They could assess the value of this upcoming satellite sensor system to support large geological mapping and mineral exploration in Canada. However, Kruse and Perry [165]. Moreover, in [166] there are attempts to determine if a relationship exists between soil’s hyperspectral data and arsenic concentration using NASA’s Hyperion satellite. Liu et al., [55] investigates the possibility of applying multi-temporal Sentinel-2 satellite images to detect heavy metal-induced stress (i.e., Cd stress) in rice crops in four study areas in Zhuzhou City, Hunan Province, China.

1. **Spectral data analysis by machine learning**
   1. **Regression**

Regression algorithms are often used to interpret spectral data [76]. For example, invariable regression is used to predict independent variables from a single dependent variable [50] used this approach to predict soil As levels (R2 = 0.56). However, as multiple dependent variables can usually be extracted from spectral data, multiple linear regression (MLR) is more commonly used. Compared to other advanced multivariate algorithms, MLR is easier to perform and interpret. However, MLR prediction accuracy is reduced when predictor variables involve non-linear relationships. Kemper and Sommer [78] employed MLR to predict the concentration of heavy metals from spectral data (R2 = 0.234e0.957). Ng et al. [69] used this approach to predict soil TPH levels (R2 = 0.71). The most used techniques for interpreting spectral data are principle component regression (PCR) and partial least squares regression (PLSR). PCR is a two-step technique in which predictor variables are transformed into principal components by principal component analysis (PCA), which are then inputted as predictors into MLR [58]. The first step allows multi-linear problems to be solved. As an enhancement to PCR, PLSR has a similar structure but also takes response variables into account in the PCA step. Therefore, PLSR not only handles multi-linear data but also allows for the number of variables to exceeds that of the samples [63], [97]. Douglas et al. [54] and Webster et al. [167] used PLSR to predict soil TPH levels with Vis-NIR and MIR spectral data, reporting R2 values of 0.63 and 0.99, respectively. Other regression approaches include elastic net regression (ENR) and penalized spline regression (PSR). ENR overcomes the problem of overfitting, whereas PSR is able to solve problems of high-dimensional data analysis. Both ENR and PSR have been utilized to predict soil As levels with reported R2 values of 0.97 and 0.89, respectively [76]

While PLSR is the most common regression algorithm for hyperspectral data [168], it has been noted in previous studies that there likely exists a complex nonlinear relationship between hyperspectral data and arsenic content [169], which is why the nonlinear models—BPNN, RF, and KNN—were added.

* 1. **Neural network**

The neural network is composed of artificial neurons which form layers that further link into connections, thus mimicking the human brain [170]. This non-linear method has attracted extensive interest in multiple fields [171], [172]. In soil surveys, back-propagation neural network (BPNN) has obtained attention for its ability to interpret spectral data more effectively than partial least squares regression (PLSR) [59], [75]. Algorithm optimization of BPNN has been explored to improve predictive accuracy. For example, Zhao et al. (2018) used BPNN with a genetic algorithm (GA) to predict soil Hg levels. A combination of particle swarm optimization and BPNN (PSO-BPNN) mitigates slow convergence and avoids trapping in local minima [173] used PSO-BPNN to predict concentrations of Hg, Cd and As with higher accuracy than primary BPNN. Tian et al. [174] optimized BPNN with the combination of GA and the ant colony algorithm to predict heavy metal concentration, with a reported R2 value for Cr detection (0.87) higher than for primary BPNN (0.55). Agrawal [166] used all BPNN models consist of an input layer with 155 input neurons, 1 hidden layer with 7 neurons, and an output layer with 1 neuron. The Tangent Hyperbolic function (tanh) was used as the activation function. Agrawal 2021 used BPNN with a genetic algorithm (GA) to predict soil As levels.

Zhang et al [158], chose tansigmoid and pureliln as the transfer functions of the hidden and output layers according to the previous study when building the BPNN prediction model. Besides, Sigmoid and trainlm were chosen as the activation function and training function, respectively [175]. The other parameters were specified below. Hidden layer, learning rate, momentum factor, and maximum learn epochs were √P (P denotes independent variables), 0.01, 0.3, and 1000, respectively.

Artificial neural network (ANN) models are used for the monitoring of heavy metals [176], and also, some researchers have tried to establish mathematical relationship between land characteristics and satellite data employing smart methods such as ANN [177] and genetic algorithm (GA) [178]–[180]. Among them, the ANN-GA hybrid model has been widely used to study the non-linear relationship between earthly measurements and satellite data. It has been reported that the model developed based on ANNGA is more accurate than the linear models [179], [181]). Generally, in monitoring process, genetic algorithm is employed for optimizations of model parameters [182], [183]. Zhou et al. [184] applied the principal component analysis, ANN, and ANN-GA hybrid approaches to model the distribution of soil heavy metals. They concluded that the most accurate models were obtained by ANN-GA.

In [185] a total of 240 (80%) and 60 (20%) samples were randomly allocated to the training and testing sets, respectively. The accuracy of models was assessed using predictions coefficient of determination (R2) and root mean square error (RMSE). Data analysis and processing of satellite images were performed in ENVI software, and maps were prepared using ArcGIS software version 10.2. The results of training (n = 240) and test (n = 60) models for MSLR and ANN-GA are shown in Table 3.

Table 3. Cross-validation results of ANN-GA and MSLR models for Cd, Pb and Zn (mg kg−1)

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Cd | Pb | Zn |
| R2\_tr (ANN-GA) | 0.85 | 0.91 | 0.90 |
| R2\_tr (MSLR) | 0.41 | 0.58 | 0.57 |
| R2\_ts (ANN-GA) | 0.91 | 0.86 | 0.76 |
| R2\_ts (MSLR) | 0.51 | 0.56 | 0.63 |

(tr train data, ts test data)

To investigate the spatial autocorrelation analysis, the local Moran I [186](Eq. (3)) and G\* statistical method [187] (Eq. (4)) were used to identify the presence of clusters:

Eq. 5

where is the Moran index, N is the number of spatial observation pixel, is the standardized observed value of pixel i, is the standardized observed value of pixel j, and is the the standardized spatial weighting value.

Eq. 6

is the G\* index; N is the number of spatial observation; is the spatial weighting value within a specified distance d of a particular observation pixel i; standardized observed value of pixel j.

* 1. **Random forest**

Random forest (RF) evolved from the decision tree algorithm, a classical and intuitive algorithm that exploits top-down and binary splits to handle regression and classification problems [188]. Because this can lead to high variance and overfilling, bagging (bootstrapping aggregation) has been included. A variety of decision trees are “trained” on extracted subsamples and the average value of splitting points. However, trees generated by bagging may correlate with each other because they are trained with similar samples. RF was developed to de-correlate trees, which selects a subsample of a feature set for each tree, compelling trees to consider all features [189]. It has been increasingly used in environmental applications and achieved superior results in comparison with other predictive techniques [190]. Douglas et al. [54] used RF to predict the concentration of TPH in soil using Vis-NIR data. The reported R2 value and PRD were 0.68 and 1.85, which was higher than that for PLSR (0.54 and 1.51, respectively). Wei et al. [191] used RF to determine soil As concentrations (R2 = 0.95). Chakraborty et al. (2017) reported that the performance of RF in predicting As levels was higher than PSR. Zhang et al [158] used RF and several parameters were tested for determining the optimal value of mtry and ntree of the RF model.

* 1. **Extreme learning machine (ELM)**

The ELM is a new algorithm for single-hidden-layer feed forward networks. For the network, the number of hidden layer nodes needs to be set and the input weight, as well as the offset of hidden elements in the process of execution, does not need to adjust [192]. Hence, the ELM has speedy training speeds and higher generalization performance. No parameters need to be manually adjusted except predefined network architecture [193].

Zhang et al [158], set the activation function of the ELM as a sigmoid function, and the optimal number of hidden nodes was adjusted from 1 to 100 with an increment of 1. Each ELM was run 500 times and the number of hidden layer nodes was obtained by the lowest bias because ELM is sensitive to the random selection of initial weight values.

* 1. **Other algorithms**

Support vector machine (SVM), linear discriminate analysis (LDA) and K-Nearest Neighbors (KNN) algorithms have also been explored in several studies [83], [194]. SVM is an effective and classical classification algorithm, which can also be used for regression. The LDA method is like linear regression but involves data classification. Stazi et al. [143]used SVM to predict the concentrations of As in agricultural soil with 18 variables (R2 = 0.82 and PRD = 2.03). Wei et al. [191] reported As detection with an R2 value of 0.91 using 5 feature variables. K-Nearest Neighbors Regression (KNN) is also predicts its target based on local interpolation of the targets associated with the nearest neighbors in the training set [195].

1. **Data acquisition**
   1. **Soil data collection**

Because VIRS requires a calibration model, both soil contaminant concentrations (e.g., traditional physical sampling) and corresponding VIRS spectra data are simultaneously collected to build a calibration model. Since soil properties can vary significantly, soil samples may be needed to build unique calibration models for each location studied. In some studies, soil samples have been prepared in the laboratory with spiked soil samples [196]. There are no existing studies to indicate the effect of soil sampling depth, however, it should be noted that spaceborne spectrometers will only observe surface soils. Therefore, soil sampling depth is usually limited to less than 20 cm [53]. In [158], each sample consisted of five subsamples collected within 1 m × 1 m plots (i.e., four corners plus center point), and then these five subsamples were mixed as one representative sample. A portable receiver of Global Navigation Satellite System (GNSS) was used to receive the signals from Continuously Operating Reference Stations (CORS) of Qianxun Corporation and to survey the sampling site’s location using Real-time kinematic (RTK) method. Vegetation, weeds, and other non-soil objects on the soil surface were carefully removed during the sampling process. Besides, all samples were put into polyethylene bags. Then the samples were transported to the laboratory for further preprocessing [158].

* 1. **Spectral measurement**

Proximal sensing in the laboratory requires soil samples to be processed. Firstly, debris, organisms and large gravel are removed before sieving (typically 2 mm mesh) [143]. Some studies involved grinding soil to 38-840 mm particle size [149]. The soil is then dried for 1-14 days, either at room temperature or at a constant oven temperature (i.e., 40C) [21],[55]. Some studies applied higher temperatures to speed up drying (e.g. 65C or 105C), but it should be noted that this could remove any volatile content from the soil [143], [144]. Khosravi et al., in [60] dried the samples at 40C and ground to <75 μm (200 mesh) to minimize the impacts of particle size on soil spectral reflectance. Afterwards, they were divided into two parts, one for chemical and the other for spectral analysis. In the next step, the sample is placed on smooth surface (e.g., a glass slide or petri dish) to diffuse reflection and gain a good signal-to-noise ratio [67]. Samples can be smoothed by saturating with distilled water to make a slurry before drying (usually at 40C) [58] or simply smoothed over manually [149]. Measurements are conducted in a darkroom with a light source. In the case of Vis-NIR spectral measurements, the light source could be a tungsten filament lamp or a tungsten halogen lamp with a wavelength of 320e2500 nm [197]. The light source is normally placed 30-70 cm above the soil [198] and the detector a distance of 10-120 cm [191], [196]. Keeping the light source and detector in specific distances from soil ensures that the light can evenly irradiate the surface of the measured object, and maintains the sample in the FOV of the detector [63]. Before measurement, background adsorption is carried out with a white reference material, such as Spectralon, polytetrafluoroethylene (PTFE) or BaSO4 [32]. Additionally, each sample should be measured 3-10 times to reduce error [144]. The concentration of Cr was analyzed by ICP-MS (LabWest Minerals Analysis Pty Ltd., Malaga, WA, Australia). X-Ray Diffraction (XRD) along with study of the thin and polished sections were also carried out for quantitative and qualitative analyses of Fe-oxides/hydroxides and clay minerals. Each sample’s pH was measured using a pH electrode inserted in the slurry containing 50 g of soil in 50 mL of distilled water [193].

Zang et al, [158], in the laboratory, first, air-dried the samples. Then the samples were crushed and grounded using wood sticks and removed impurities including gravel, and other foreign matters. Third, all soil samples were placed into an oven to dry until weight without changing. Then, a 0.7 mm nylon aperture sieve was used to sieve samples and put into clean polyethylene bags for analysis. Then each sample with 4 g weight was placed into a 32 mm mold painted with boric acid to squeeze a tablet under 30-ton pressure for content measuring via X-ray fluorescence. The mean content of each sample was recorded after three measurements to decrease the deviation.

Finally, specific software for X-ray fluorescence (SPECTRO xSORT) named Sample Result Manager was used to pretreat heavy metals concentration data. The quality assurance and quality control (QA/QC) were assessed. The GSS-series and GSD-series geochemical reference standard materials (Institute of Geophysical and Geochemical Prospecting, Lang fang, China) were used to calibrate the X-ray fluorescence (SPECTRO xSORT) instrument and to ensure the relative standard deviation ranged from 3% to 5%. The spectral resolution was 1 nm with a spectral range of 350–2500 nm. A black plastic petri dish with 10cm × 10 cm size was used to fill the sample. A warm-up with 30 min duration was carried out for the ASD to minimize bias. A whiteboard with 99% reflectance was used to calibrate the spectrometer before surveying. A 1000 W halogen lamp with a similar spectral feature of the Sun was utilized as the light source. The observation angle between the light and the vertical direction was set as 15◦. The 30 cm distance was adjusted between the light source and the soil samples. Besides, the distance and angle between the probe and the samples were 15 cm and 90◦, respectively. The field of view (FOV) is 25◦, and the diameter of the field of view is 7 cm. Clearly, the area of the soil sample container completely covered the area of soil spectra. Each sample was scanned 10 times to minimize observation errors originating from stray light, and the final spectrum was determined by the average of 10 scanning spectra.

* 1. **Soil spectral libraries**

To expand the use of VIRS in soil monitoring, international efforts are being made to establish spectral libraries. The first was established by the US National Soil Survey Center in 2006, which contains 3768 samples from the US and 416 samples from countries in Africa (125), Europe (112), Asia (104) and the Americas (75) [199]. Other institutions have published data including 21,500 spectra collected from 4000 soil profiles in Australia, and the spectra of 20,000 samples collected across Europe [200], [201]. In recent years, the Vis-NIR spectra of 23,631 soil samples collected from 35 institutions around the world have been compiled [202].

1. **Statistical analysis methods and modeling strategies**
   1. **Data pre-processing**

Data pre-processing is used to render data valid for model building. Kooistra et al. [32] reported that prediction accuracy and model quality was vastly improved after pre-processing was carried out. Pre-processing usually involves outlier removal, noise minimization and curve smoothing [143]. Data outliers may originate from the sample itself or from experimental operations. Removal of outliners is one of the keys to establishing stable and effective predictive models. Outliers ought to be identified using a systematic method, such as principal component analysis (PCA) [52], with outliers identified by a score matrix. Chakraborty et al. [76]used PCA to preprocess spectral data, identifying 10 outliers. A normal distribution is a prerequisite for some statistical methods, such as Pearson correlation analysis. A normality test can be used to check data normality (e.g., a ShapiroeWilk test and Kolmogorov-Smirnov test showing a p of >0.05). If the data is nonnormal, transformations such as Box-Cox transformation and logarithmic transformation can be applied [48], [53]. Noise in collected spectra will often relate to the roughness of the surveyed land or the observation angle [59]. Bands showing large amounts of noise can be removed (e.g., the initial and tail bands) [142], [203]. Additionally, mathematical transformation methods can be adopted to reduce noise levels (Table 4). The relative effectiveness of data pre-processing has been analyzed in several studies. Liu et al. [204] reported that reflectance data processed by logarithm and continuous removal increased the level of correlation with heavy metals.

Since Aster SWIR sensor stopped functioning in April 2008 due to high SWIR detector temperature [60], VNIR bands of the image was subjected to radiometric calibration using the Internal Average Relative Reflection (IARR) method and then processed to mask the clouds and shadows. For Hyperion, the uncalibrated and overlay bands with low SNR were removed [205]. Atmospheric correction was also implemented using the Fast Lineof- sight Atmospheric Analysis of Hypercubes (FLAASH) algorithm. Sentinel-2A was atmospherically corrected using the SNAP toolbox included the Sen2Cor algorithm [206]. The bilinear interpolation method was then used to resample the image bands to 10 m. Radiance calibration and atmospheric correction using FLAASH algorithm were conducted for Landsat 8-OLI pre-processing. The spectral bands one to seven were down-scaled to 15 m using the Landsat 8-OLI panchromatic band. Geometric rectification of all four images was finally performed using 20 Ground Control Points (GCP) [60]. Chen et al. [75]compared six pre-processing methods, finding that orthogonal signal correction most effectively reduced noise and improved prediction accuracy. However, reported optimal preprocessing methods have varied among studies, owing to the specific features of the spectral data [32], [75]. In practice, the use of multiple data preprocessing methods may be needed to determine the optimal approach.

In [185] Image processing was performed using Landsat ETM + imagery (ID LE07\_L1TP\_166035\_20140527\_ 20161115\_01\_T1, acquisition date May 27, 2014, path 166, row 35) In this research, fast line of sight atmospheric analysis of spectral hypercubes (FLAASH) algorithm was employed to correct the atmospheric effects. Using this algorithm, the values of the spectral radiation were converted to spectral reflection, and the effects associated with the changes in the lighting conditions, season, latitude, and meteorological conditions on the images were removed. In order to remove noise in the images used, minimum noise fraction transform (MNFT) algorithm was employed by ENVI software. This conversion is a linear conversion which is employed for determining the main dimension and volume of the image, separating noise from other information, and entire region sporadically. For this purpose, the geographical position of 40 points was withdrawn using global positioning system device (GPS model Garmin 41838), and the geometrical correction was performed on the image. After atmospheric and geometric corrections, mean values of digital numbers within a radius of 30 m around the sampling points were determined for each band in MATLAB (R2014a). Shapiro-Wilk test was used to assess the normality of residual. To remove the weakest correlated variable, MSLR was applied. This technique minimizes the sum of squared deviations of the observed and predicted dependent variables via the linear transformations of the independent variables. The relationship between heavy metal concentration (as a dependent variable) and averages of pixel values and soil properties (as independent variables) were determined for each band and band ratio by applying MSLR analysis. The band ratios are considered to be absolutely useful methods for highlighting phenomena like heavy metal in multiband images [207], [208]

The direct standardization (DS) algorithm was selected in [158] to correct the GF-5 AHSI imagery for eliminating effects of soil water and other uncertain environmental factors on soil spectra retrieved from GF-5 AHSI imagery. A subset of laboratory-measured (XLab) and field-obtained (XGF- 5) spectra were selected to determine the DS transfer matrix which describes the deviations between field-obtained and laboratory-measured spectra (Eq. (6)).

Eq.6

Where is the transferred field-obtained spectra from the GF-5 AHSI imagery after DS. More details concerning equation (6) can be found in the published literature [209]. Additionally, the performance of the DS algorithm can be affected by the representativeness and the number of transfer samples. Hence, the Kennard–Stone algorithm was used to choose a series of transfer sets with different numbers of samples (t = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, and 110) from the whole dataset for testing the effects of transfer samples on the performance of the DS algorithm [210]. One key point needs to be indicated that the spectral wavelength of laboratory-measured spectra must be selected for matching the corresponding spectral bands of GF-5 AHSI imagery before conducting the DS algorithm. The performance of DS was assessed by calculating the spectral angles mapper (θ) between the DS-transferred field spectra and their corresponding laboratory-measured spectra Eq. (7)[209], [211].

Eq.7

Where and denote the spectral responses of laboratory and field spectra at each wavelength, i and n denote the total number of wavelengths. In this paper, n = 299. Each pixel of GF-5 AHSI imagery was transferred by the optimum DS algorithm.

Furthermore, preprocessing of the hyperspectral satellite data is necessary to reduce noise and data redundancy, reducing the error caused by external influences. Geometric correction involves distortions caused by errors such as the satellite’s positioning on its orbit and angle changes due to Earth’s rotation. NASA performs most base-level geometric processing before releasing Hyperion datasets publicly [212]. NASA’s provided datasets contain hyperspectral values in a 2D format, with swaths angled away from geographic north while measuring regions (due to Earth’s rotation). Given the latitude-longitude coordinates of the four corners, a linear interpolation approach was used to determine the specific swath pixel containing the desired latitude-longitudinal location. As the geographical scale of each individual swath (averaging approximately 1.3 degrees of latitude and 0.4 degrees of longitude) is relatively small, linear interpolation provides very reasonable approximations. Next, to correct for problematic atmospheric conditions, the default numerical radiance values were converted into reflectance values using the well-known formula [213]:

Eq.8

where is reflectance, is the spectral radiance at the sensor’s aperture, d is the Earth– Sun distance in astronomical units (AU), is the mean solar exo-atmospheric irradiance (provided by NASA), and is the solar inclination angle in degrees. As reflectance measures the ratio of the amount of light leaving a target location to the amount of light striking the target, the conversion helps minimize the effects of atmospheric scattering, varying Earth–Sun distances, and solar elevation at time of measurement. Next, potentially troubling Hyperion bands are identified and filtered out based on past literature and basic atmospheric principles [213][214]. The Hyperion Visible-Near-Infrared (400–1100 nm) sensor makes up 70 of the 242 hyperspectral bands, with the other 172 composed from the Short-wave Infrared (1100–2500 nm) sensor. Due to low sensitivity regions and overlap between the two sensors, bands 1 to 7, 58 to 78, and 224 to242 are removed from the dataset. Atmospheric water vapor bands, which absorb almost all relevant solar radiation, and bands likely to contain unstable atmospheric effects are further removed. The final subset consists of 155 bands: 10 to 57 (448–926 nm), 81 to 97 (953–1114 nm), 101 to 119 (1155–1336 nm), 134 to 164 (1488–1790 nm), and 182 to 221 (1972–2365 nm).

Noise reduction with SD transformation is the most effective likely because it can highlight otherwise hard to discern absorption features of spectral curves [161]. Furthermore, taking two derivatives allows for the clearer differences to be observed between spectral reflectance of indicative bands with different arsenic contents, which served to increase model accuracies more than the other kinds of transformations.[166]

Table 4. Spectral transformation methods

|  |  |
| --- | --- |
| **Name** | **Objective** |
| Mean Centering | Eliminate the absolute absorption value of the spectrum, increase the difference between the sample spectra, and improve the robustness and prediction ability of the model |
| Orthogonal signal correction | Filter out signals that are not related to the concentration of the target pollutant in the spectrum |
| Standard Normal Variate (SNV) | Eliminate spectral errors caused by solid particle size surface scattering |
| Multiplicative Scatter Correlation (MSC) | MSC can correct for certain errors caused by light scattering [26]. |
| Savitzky–Golay (SG) smoothing filter | SG smoothing allows for the improvement of the accuracy of noisy synthetic data without losing important baseline trends [37,38]. |
| First Derivative Transformation (FD) | Correct the spectral baseline, eliminate interference from other backgrounds, and improve spectral resolution. |
| Second Derivative Transformation (SD) | SD magnifies smaller differences between wavelengths and reduces some sources of random error caused by external factors through this focus on differences. |
| continuous wavelet transform (CWT) | intensifying significantly spectral responses of GF-5 AHSI images  at a variety of decomposing scales |
| Image processing | fast line of sight atmospheric analysis of spectral hypercubes (FLAASH) algorithm was employed to correct the atmospheric effects |

* 1. **Variable construction**

Variables can be classified as two types: 1) raw or preprocessed spectral feature bands; 2) combined spectral data. The first type provides the most representative information and higher model quality. There are two methods for selecting feature bands: 1) linear regression; or 2) PCA. For linear regression, reflectance correlation coefficients are calculated, with the bands of highest value used as feature variables. The magnitude of correlation coefficients can depend on the pre-processing method applied [204]. In PCA, uncorrelated principal variables are extracted explaining the highest variance. Calibration models such as PCR and PSLR are constructed based on principal components. Other prediction methods also use principal components as prediction variables (e.g. RF) [144]. The number of feature bands used for modeling can range from one to hundreds [149]. Calibration R2 values will tend to increase as the number of feature variable increases, but overfitting may occur at higher numbers. As a rule of thumb, the optimal number of variables is around one third of the number of samples. Combined spectral data can also serve as variables. For this, correlation analysis can be used to select the most effective combination. Liu et al. [149] selected two combinations of spectral data to predict Cd, Hg and As levels in soil. Some commonly used spectral indexes for vegetation, such as the normalized difference vegetation index (NDVI) and the infrared percentage vegetation index (IPVI), have been identified as efficient predictors when using vegetation reflectance data [52]. For example, Shi et al. [50] employed different vegetation indices to predict As levels (R2 = 0.75).

* 1. **Model selection**

Various models for interpreting spectral data were introduced in the previous sections. The model function should be considered firstly in model selection. If the underlying relationship is non-linear, algorithms such as PSR, neural network, RF should be used. PLSR and stepwise regression could help to diminish the risk of collinearity [75], [141]. In [157], linear regression (stepwise) in SPSS software was used to predict the concentration of heavy elements in the soil based on the amount of reflection in the satellite image multivariate. Stepwise regression is a method for analyzing the relationship between independent variables and dependent variables. Model parameters should be considered carefully to avoid overfitting. For instance, in the neural network algorithm, the number of neurons in each hidden layer, the number of hidden lays and the selection of propagation functions can influence model accuracy. Overfitting can occur if the model is too complex. Models with different combinations of parameters should be built, tested and compared. Additionally, models can be optimized with other advanced algorithms, including the genetic, particle swarm optimization, least absolute shrinkage and selection operator algorithms [63], [173]. Such algorithms help improve solution searching and avoid the problem of overfitting.

Linking heavy metal contents to spectral absorption using MSLR and other linear regressions has been utilized by several researchers, leading to the prediction of models with high accuracy [21] [61], [63], [215]–[217] applied stepwise multiple linear regression (MSLR) for predicting As, Cd, Cu, Fe, Hg, Pb, Sb and Zn distribution and obtained the accurate models for most of the elements.

* 1. **Model validation**

Model validation is required to determine prediction error and evaluate model quality. After initial data pre-processing, it is useful to split the data into two separate sets: one set for model training and another for validation [72]. Usually, around 70% of data is used for training and 30% for validation. Model performance can be evaluated systematically using cross validation techniques. In k-fold cross-validation, the data is randomly divided into k equal sized subsamples, with one subsample retained as validation data. The remaining k-1 subsamples are used as training data. The process is repeated k times, with each subsample used once as the validation subset. The average error serves as the performance parameter [55]. The leave-one-out validation procedure is utilized when the number of available samples is small [44]. In this approach, n-1 samples are adopted to train the model and the remaining sample used for validation. The procedure is repeated n times and the root mean square error of cross-validation (RMSECV) serves as the performance parameter. Kooistra et al. [32]used the leave-one out approach to validate a PLS model with 69 samples.

The degree of similarity between each sample’s spectra and the spectra of the same position pixel was calculated using the Spectral Angle Mapper (SAM) [218]. In this method, spectra are considered as space vectors with dimensions equal to the number of spectral bands (Equation (9)).

Eq. 9.

where it is the spectra of the image pixel and r is the spectra of the soil sample measured using a spectrometer. Lower values indicate a higher degree of similarity between the two spectra [218]. The one-way ANOVA method tested the statistical similarity between the spectral properties of images and soil samples. According to the null hypothesis in ANOVA, there was no significant difference between the mean values of the selected image features and the lab spectra. In other words, the null hypothesis is that the samples in the two groups are drawn from populations with the same mean values. This hypothesis was tested by calculating the F-value at the significant level of 0.05. In the case of significant correlation, the spectral features of the image were used as independent variables for the prediction model obtained by lab spectroscopy to determine the concentration and spatial distribution of Cr. The resulting maps were compared to the measured Cr distribution maps.[60].

* 1. **Model quality assessment**

Model quality assessment is a key process in machine learning. Determination coefficients (R2), the root mean square error (RMSE), residual prediction deviation (RPD), the ratio of performance to inter quartile distance (RPIQ), standard error (SE) and bias, can all be used to quantitatively assess model quality (Table 5). The R2 value is the most widely used parameter for assessing model quality, which is the proportion of the variance in a predicted value (dependent variable) that is predictable from the independent variable. The closer R2 is to 1, the better the fit of the model. Reported R2 values in the reviewed literature ranged from 0.11 to 0.99. The RMSE value is the standard deviation of the residuals (prediction errors). The smaller the RMSE, the higher the accuracy of the model. PRD is a goodness-of-fit parameter that is defined as the standard deviation divided by the RMSE, with values greater than 1.8 considered good [71], [144], [219]. PRD values reported in the reviewed studies ranged from 0.51 to 6.23 [76], [78]. In [158], three evaluation indicators including R2, RMSE, and MAE were selected to assess the robustness and precision of the inversion models [192], [220].

Table 5. Evaluation parameters for determining model quality.

|  |  |
| --- | --- |
| Parameters | Equations |
| r (correlation coefficient) |  |
|  |  |
| SE (standard error) |  |
| RMSE (Root mean square effort) |  |
| RPD |  |
| RSD (relative standard deviation) |  |
| Bias |  |

is the observed value of sample i; is the predicted value of sample i;

y is the average of observed value; is the average of predicted value.

* 1. **Image Segmentation**

Rasters resulted from the best models were also segmented using a binary fitness function. Accordingly, if the predicted concentration of Cr for a pixel is greater than a threshold value (i.e., the average measured concentration), the desired pixel got the value of one; otherwise, it obtained the zero value. The binary fitness function used to classify image pixel is as Equation (10) [60]:

Eq. 10

where is the pixel in row i and column j of the image matrix, is the threshold value, and P() and are the predicted and segmented values for the same pixel, respectively [60].

1. **Potential of recently developed and forthcoming spaceborne sensors for soil contamination monitoring**

The identified potential of imaging spectroscopy (IS) is not counterbalanced by the accessibility of satellite imaging data, although their number and availability are constantly growing. Moreover, the opening of large data sources such as Landsat [133], entire space missions expanded for the public domain such as the Copernicus program of the European space agency (ESA) [221][222][134], the WorldView-3 and the WorldView-4 of the American DigitalGlobe [223] has brought advantages for application of spaceborne technologies. Furthermore, in near future a new generation of hyper and superspectral spaceborne sensors with higher signal to noise ratio (SNR) and shorter revisit-time is due to be launched, including HYPXIM [224], HyspIRI [225], EnMAP [226], PRISMA [227], FLEX [228], and SHALOM [229]. Because of such advancements, it is anticipated that these technologies will direct to an exceptional movement in the application of space-based remote sensing techniques for soil contamination monitoring. Khosravi et al. , [193] selected the Hyperion, Sentinel-2A, Aster and Landsat 8-OLI images for their study. Selecting appropriate images was based on spectral coverage, regional parameters and acquisition conditions such as (i) date and time, (ii) cloud coverage, (iii) the coordinates of the area, (iv) water vapor in the atmosphere and (v) climate [60]. Table 6 summarizes the main technical specifications of some of the spaceborne imagers. Soil scientists gradually study dynamic phenomena over space and time including oil and gas seeps, soil contamination, and general soil degradation, which need a longer time series of calibrated data. Recently developed, as well as forthcoming space-based sensors, are very valuable sources of data in this regard and provide an innovative and cost-effective method, which can allow the monitoring of such processes for repetitive coverage of large areas. Owing to the high potential of hyper and superspectral spaceborne images in soil and geological applications, special features of the recently operated the Sentinel-2, as well as forthcoming EnMAP, HyspIRI, EO-1 Hyperion and GaoFen-5 are briefly explained below [47].

Table 6. Technical characteristics of some spaceborne remote sensing imagers.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sensor | Country | Organization | Launching year | GSD (m) | Spectral bands (no.) | Spectral range (nm) | Swath at nadir (km) | Temporal res. (days) |
| Landsat-5 | USA | NASA/NOAA | 1984 | 30, 57, 120 | 7 | 450–2,350 | 185 | 16 |
| Landsat-7 | USA | NASA/USGS | 1999 | 30 | 8 | 450–12,500 | 185 | 16 |
| ASTER | USA | NASA | 1999 | 15, 90 | 15 | 520–11,650 | 60 | 16 |
| IKONOS | USA | GeoEye | 1999 | 0.82, 3.28 | 4 | 445–853 | 11 | 3-5 |
| Hyperion | USA | NASA | 2000 | 30 | 220 | 360–2,600 | 7.65 | 16 |
| CHRIS-PROBA | EU | ESA | 2001 | 17, 34 | 18,37,6 | 400–1,050 | 13 | Varies\* |
| MODIS | USA | NASA | 2002 | 250, 500, 1000 | 19 | 620–965 | 2330 | 16 |
| MERIS | EU | ESA | 2002 | 300, 1200 | 15 | 390–1,040 | 1150 | 35 |
| SPOT-5 | France | CNES | 2002 | 10 | 4 | 480–1,750 | 80 | 2-3 |
| Landsat-8 | USA | NASA/USGS | 2013 | 30 | 11 | 430–12,510 | 185 | 16 |
| WorldView-3 | USA | DigitalGlobe | 2014 | 0.3, 1.24, 3.70 | 16 | 400–2,365 | 13.1 | 1-4 |
| Sentinel-2 | EU | ESA | 2015 | 10, 20, 60 | 13 | 440–2,195 | 290 | 5 |
| PRISMA | Italy | ASI | 2017 | 30 | 247 | 400–2,500 | 30 | NA |
| HISUI | Japan | METI | 2018 | 30 | 185 | 400–2,500 | 30 | NA |
| GaoFen-5 | China | CNSA | 2018 | 30 | 328 | 400–2,500 | 40 | NA |
| EnMAP | Germany | DLR | 2020 | 30 | 242 | 420–2,450 | 30 | 27 |
| HyspIRI | USA | NASA/JPL | 2020 | 60 | 214 | 380–2,510 | 145 | 19 |
| SHALOM | Italy/Israel | ASI/ISA | 2021 | 10 | 241 | 400–2,500 | 10 | 4 |

NA: not available, NASA: national aeronautics and space administration, NOAA: national oceanic and atmospheric administration, USGS: United States geological survey, ESA: European space agency, CNES: centre national d’etudes spatiales, ASI: agenzia spaziale Italiana, METI: ministry of economy, trade and industry, DLR: Deutschen zentrums fur luft- und raumfahrt, CNSA: China national space administration, JPL: jet propulsion laboratory, ISA: Israel space agency. \*Varies: It has a nominal sun synchronous polar orbit (SSO) but no orbit maintenance capability.

* 1. **Sentinel-2**

The Sentinel-2 satellite was successfully launched on 23 June 2015. It was planned to provide continuity to monitor services over global surfaces, leaning on superspectral high-resolution observations. It guarantees continuity of the SPOT and Landsat assignments and offers applications such as land change detection maps and land cover maps [230]. The lately launched WorldView-3 has eight SWIR bands, which four were taken from the ASTER [165]; however, Sentinel-2 is missing such narrow bands in the SWIR range (Fig. 7). Afforded proxies with Sentinel-2 are the same as Landsat obtained ones, even though it had a slightly higher spatial resolution. According to Van der Meer et al., [231] as there is a good correspondence between ASTER and Sentinel-2, several band ratios are proposed for Sentinel-2 to derive some soil attributes. Van der Werff and Van der Meer [223] stated that Sentinel-2 brings us information on soil attributes, together with iron oxide (which has good correlation with PTEs) that previous multispectral missions could not. In addition, it is the first optical Earth observation mission of its kind to include three bands in the “red edge,” which provide key information on vegetation state. Sentinel-2 data has been used to investigate forest fires and crop classifications [232], [233] Though, the simulated imagery of the superspectral Sentinel-2 was used by Mielke et al. [163] to map the spatial extent of mine waste surfaces in South Africa, which contained problematic trace elements of U, Pb, and Cr. The reliability assessment of the sensor to provide operational products would necessitate inquiry of real data [47]. Table 6. Presents the Sentinel satellite bond specifications.

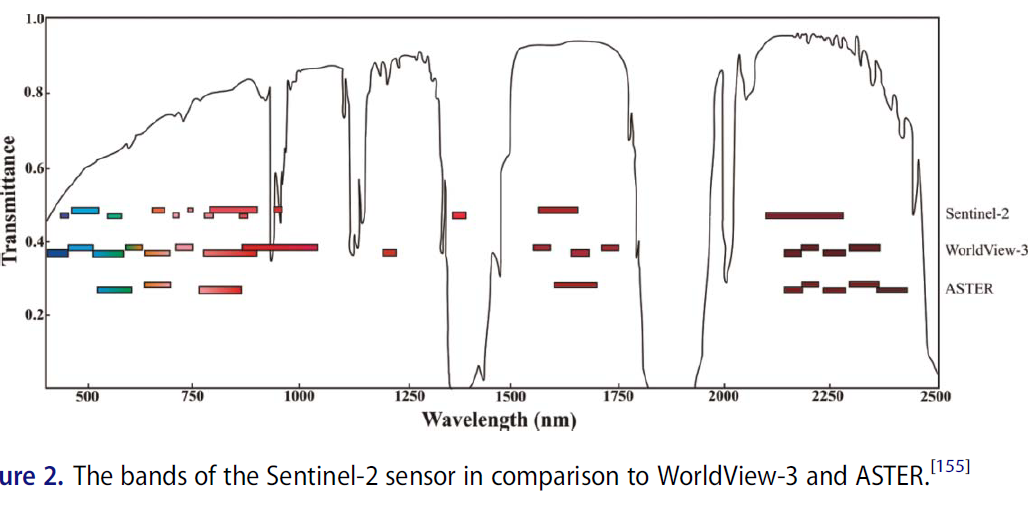


Fig 6. The bands of the Sentinel-2 sensor in comparison to WorldView-3 and ASTER [234].

Table 7. Sentinel satellite bond specifications [157].

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Goals of uses | Wavelength | Band width | Spatial separation | Band number |
| Atmospheric correction (aerosol diffusion) | 443 | 20 | 60 | 1 |
| Plant measurements, carotenoids, browning and soil background. | 490 | 65 | 10 | 2 |
| Green peak, sensitive to total chlorophyll content in plants. | 490 | 65 | 10 | 3 |
| Maximum absorption of chlorophyll | 665 | 30 | 10 | 4 |
| Red edge position | 705 | 15 | 10 | 5 |
| Red edge position, Atmospheric correction, Aerosol load recovery | 740 | 15 | 20 | 6 |
| Leaf area index, near infrared edge | 783 | 20 | 20 | 7 |
| Leaf area index | 842 | 115 | 10 | 8 |
| Near infrared plate sensitive to total chlorophyll, biomass | 865 | 20 | 20 | 8A |
| Atmospheric correction (water vapor absorption) | 945 | 20 | 60 | 9 |
| Atmospheric Correction (Detection of Thin Cirrus Clouds) | 1375 | 30 | 60 | 10 |
| Sensitive to starch lignin and forest and biomass levels, separation of snow, cloud and ice | 1610 | 90 | 20 | 11 |
| Assessment of the conditions of Mediterranean plants, the distinction of clay soils in order to monitor soil erosion, the distinction between living and dead biomass of soil. | 2190 | 180 | 20 | 12 |

* 1. **EnMAP**

The spaceborne EnMAP sensor is devoted to environmental applications of state-of-theart hyperspectral technology [234]. The launch of EnMAP is projected for 2020 and the anticipated mission lifetime is 5 years. It has been introduced as a sensor well suited to soil and geological mapping due to its fine spectral resolution of 6.5 nm in the VIS-NIR range and 10 nm in the SWIR range. It will offer new opportunities for short term minerals and metals detection, and long-term mine waste observation, as well as retrieving accurate soil data for mapping the world’s soils [163], [235]. Using the full wavelength range of EnMAP will provide the chance to track hydrothermal alteration by mineral distribution spectroscopic mapping to imply PTEs through noticeable absorption features of minerals, which are these deposit types characteristic [234]. Fig. 7 represents an EnMAP overpass depicting VIS-NIR and SWIR spectrometers FOV. Mielke et al. [167] employed synthetic EnMAP data to monitor areas around a gold mine; the mean accuracy for mapping of 900 nm Fe absorption features of minerals related to mine waste was 100%. Therefore, they claimed that in the future, in remediation of mining areas, EnMAP might be economical. Rogge et al. [164] also proved that EnMAP provided reconnaissance mapping of Ni, Cu, and platinum group elements over vast subarctic and arctic regions in Canada.

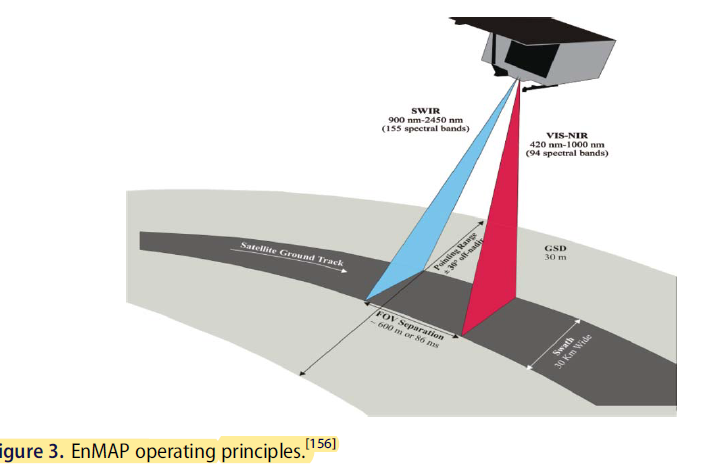


Fig 7. EnMAP operating principles [234].

* 1. **HyspIRI**

The HyspIRI mission will be launched during the next decade and will comprise of two instruments: a VIS-NIR-SWIR imaging spectrometer and an eight-band TIR multispectral imager, plus an intelligent payload module (IPM) for on-board large data sets processing and quick downlink of selected data [46], [236]. The HyspIRI mission expected to provide high-quality global datasets for a vast range of science and application requirements, including monitoring natural and anthropogenic disturbances such as oil spills in soil because of the availability of multiple SWIR bands [237]. For instance, following the deep-water horizon oil spill in Barataria Bay soils to map the soil contamination, Kokaly et al. [238] used hydrocarbon absorptions detectable in SWIR imaging spectrometry data. HyspIRI TIR data would bring the opportunities for new research on hydrocarbon resource explorations with its possibility of temperature anomalies mapping [231], [239]. HyspIRI VIS-NIR-SWIR and TIR data fusion can significantly improve the capability of surface materials (rocks, soils, and vegetation) discrimination,[231], [240] which would be an early attempt for assessment of land surface change, if caused naturally or of anthropogenic source. In all, due to the spatial, spectral, and temporal specifications of recently developed and forthcoming satellite imagers, it seems that there is a potential for developing products that can benefit the soil science and geology community. This is expected to offer faster worldwide support while retaining the maximum level of data consistency. Nonetheless, efforts are still needed for IS products advancement that can solidly support soil spectral libraries (SSLs) development [202], global soil digital mapping, and soil contamination monitoring.

* 1. **EO-1 Hyperion**

Hyperspectral remote sensing is an emerging field that uses imaging spectroscopy to obtain images in continuous spectral ranges from 0.4 to 14 μm [241]. Compared to its predecessor, multispectral imaging, hyperspectral imagery offers a variety of advantages, including high spatial resolution and an appreciably larger amount of spectral information [242] Hyperspectral imagery has been used to remotely study chemical and physical properties of the geology of a given region, with application to fields such as precision agriculture, ecological modeling and monitoring, discovering important ores, and urban planning [149].

NASA’s EO-1 Hyperion satellite, which orbited the Earth daily from 2001 to 2017 and gathered over 54 terabytes of data in the process, pioneers important use of this technology. Hyperion provides high spectral resolution data with 242 bands at 10 nanometer (nm) spectral resolution, 30 m length pixels, and swath width of 7.5 km [60], collecting detailed images (in the form of 3D-cubes) over a large geographical area relatively efficiently.

* 1. **GaoFen-5**

The GF-5 satellite launched on May 9, 2018, the first integrated hyperspectral satellite carrying Advanced Hyperspectral Imager (AHSI) for both land observations and atmospherics all over the world, supplied up-to-date datasets for earth science (Tang, 2018). The spatial resolution (30 m), the spectral resolution of the shortwave infrared band (≤5 nm, up to 0.03 beams), the spectral range (390–2513 nm), and the time resolution (5 days) of GF-5 AHSI were dramatically enhanced comparing with existed hyperspectral satellite sensors [175][158].

1. **Current limitations of spaceborne sensors for soil contamination Monitoring**

Despite the significant progress made in remote sensing of soils during the last decades, and the promising performance of recently developed spaceborne missions, which open opportunities to implement novel and accurate retrieval algorithms in operational processing chains, there are still some limitations in using their data, which still have not been solved. These problems are mainly associated with the data acquisition condition, environmental situation, and the interpretation of different surface variables from satellite data at the temporal and spatial data scales [243]. The main difficulty in soil remote sensing is the presence of vegetation. Soils are usually covered by vegetation and the sensors obtain the signal from the surface, which is covered with vegetation and soil; however, the evaluation of extracted soil spectral information from satellite images is related to the bare soil availability [244]. The vegetation coverage and accessibility of bare soil in satellite images of cultivated lands are limited to a short period of the year. Hence, the possibility to assess soil attributes from space-based remote sensing imagery over vegetated areas is an important prospect, which requires the development of a strategy to remove the vegetation effect to achieve bare soil and reach soils attributes through satellite images in these conditions. To attain bare soil and estimate soil clay content in agriculture areas from Landsat data, Shabou et al. [245] simultaneously analyzed the Landsat, the normalized difference vegetation (NDVI) and MIR data over all test fields to achieve bare soil from satellite images. According to Zhang and Zhou, [243] to obtain quantitative soil moisture under vegetation cover using satellite data, two-source ALEXI models can be utilized to separate the soil and vegetation information and describe the energy exchange among the soil, vegetation, and atmosphere. Garnier et al. [246] showed a relationship between plant’s attributes as responses to soil condition variation. Additionally, Li et al. [247] and Solon et al. [248] investigated the relationship between soil characteristics and vegetation through satellite remote sensing, suggesting that these two elements are closely related. Thus, using vegetation as bioindicators of soil condition can be a solution in vegetated areas [13] Roelofsen et al. [14] stated that deriving the soil conditions through their impact on the natural vegetation is an efficient alternative, as plant communities usually have a narrow tolerance toward soil factors, making them indicative of the site’s conditions. Regarding the airborne remotesensing approach, Noomen et al. [10] used the hyperspectral airborne Probe-1 data in the VIS-NIR-SWIR region to study the long-term hydrocarbon seepage in soil using vegetation response as a proxy of soil condition. The analysis of natural forest vegetation to obtain soil information based on satellite remote sensing was studied by Arellano et al. [86] and Dematte et al. [244] using Hyperion and Landsat, respectively. Their studies showed that this approach could be an efficient way to reach accurate soil results. However, the need to develop techniques in different regions has been pointed out by most authors, since correlation of vegetation with soils is complex. Indeed, obtaining information from vegetated soil is a different method. On the other hand, vegetation expression shows important inferences and correlations with soil attributes. This methodology can assist soil monitoring in vegetated areas with difficult soil access. Another problem in remote sensing of soil arises from the fact that passive sensors’ images are sensitive to weather conditions. In the presence of clouds and shadows in a pixel, the usability of this pixel is prevented [249]. This problem is more serious in the case of using high-resolution images (Landsat, SPOT, ASTER, etc.) and in areas with high frequency of cloud coverage. In the case of Sentinel-2 also, cloud masking in the absence of a TIR band and the need to combine spatial resolutions for 10 m will be the main limitation. In the case of poor atmospheric correction, problems can also occur due to lower sunlight intensity and limitations of simulated datasets [250], The revisit-time of satellites can also be an issue, which still needs to be improved. Generally, in comparison to low temporal resolution satellite images, high-resolution ones have more information in their spatial domain, attained at infrequent time intervals [249]. Thus, the procedure of cloud recovery from a high temporal resolution image depends either on the image features [251] or on a limited number of images [252], while the advantage of high temporal resolution is for low-resolution data. Therefore, many images that cover a short period of time can be processed and the best observations from these images selected, which can be assumed that no cloud has occurred above the area.

There is another limitation in soil remote sensing, which is the heterogeneous nature of remote sensing data. The recorded reflectance by remote sensors, especially on large-scale areas, is resulted from the interactions of electromagnetic radiation with multiple constituents within each pixel that consequently constrains the accuracy of spectral analysis and their interpretation for quantifying soil condition. Over the past decade, although numerous studies addressed the mixing problem and proposed analysis techniques [253], [254], spectra unmixing is still challenging and need more development. In addition, a wide range of soil spectral measurements are being gathered around the globe, which provide different outputs due to various measurement conditions including sampling techniques, sample preparation, instrument specifications, different protocols as well as analytical algorithms, which severely affect the prediction performance of spectroscopic models and outputs [173][255]. These variations of condition justify the importance to establish a simple procedure for standardization that minimizes the systematic and random effects and enables unification of spectral libraries [256] A specific limitation concerning soil contamination is the fact that pollutants do not have direct spectral features and that a correlation with soil characteristics is crucial. Moreover, most of the satellites provide spatial resolution of around 30 m, which can be a limitation in soil contamination assessment. All in all, major issues in the use of remote-sensing data, specifically hyper and superspectral sensors information, for soil applications are (i) the evaluation of soil spectral information in vegetation-covered soils [244]: (ii) the sensor errors correction and atmospheric attenuation removal difficulties [257]; (iii) spectral mixing problem and limited available algorithms for unmixing [254]; (iv) the inadequate SNR of some sensors (e.g., the Hyperion SWIR spectral region) [258]; (v) lack of standardization and an agreed-upon protocol [256]; and (vi) incomplete spectral coverage of some sensors (e.g., the CHRIS-PROBA with lack of SWIR bands) [257]. Therefore, further works are needed to deal with the above-mentioned limitations and issues. The aforementioned factors can play large roles in limiting the applicability of a satellite-based approach in certain regions of the world, particularly those with frequent climatic fluctuations, successive topographical changes, and year-long soil cover. However, outside of these regions, their effects on the accuracy and applicability of the presented methodology are adequately addressed using thorough preprocessing [166].

1. **Summary and future research directions**

As one of the most important components of nature and the primary source of human food supply, the soil needs to be conserved and remove contaminants. In order for humans and other species to survive on the Earth, the soil must be cleared of impurities and prevented from continuing pollution. It is important to identify soil contaminants sources, detect soil contaminants pathways, and rehabilitate them.

This study reviewed the methods of identifying soil pollution by satellite imagery. At first, the types of pollution and identification methods like soil organic matter, Fe-Oxides, and clay minerals in bare soil situations, vegetation spectral signatures, response to soil contamination, and spaceborne sensors technique in the vegetation areas were discussed.

Some contaminants in the soil were detected by satellite images like acid mine drainage, petroleum hydro carbonite, and heavy metals.

Today, with the advancement of computer science, especially in machine learning, there has been a considerable improvement in its use in all sciences. Machine Learning has entered the field of geotechnical engineering in recent decades, and various methods have been used to apply it. In this study, some methods of ML in soil pollutants used in satellite imagery detection were reviewed. Zang et al. [158] have shown that the RF method exhibited better performance in detection of HM than other methods, owing to the information gain strategy for variable selection and the variable importance score. A wide range of the target property was the fundamental factor that affected the model capability. The relatively wide range of Zn, Ni, and Cu variations in the present study may be one of the reasons for better performance with RF models [159], [220]. The better performance of ELM in predicting heavy metal concentrations is owing to the unique ability of neural networks, especially in handling nonlinear internal relationships between heavy metals and spectral reflectance. The estimation accuracy of ELM was better than other liner models, and the calculation speed was also fast.

The ELM with higher generalization ability excluded numerous obstacles such as learning rate, stopping criteria, and local minima. The training speed of ELM can be hundreds of times faster than BPNN in obtaining an excellent model with no iterative steps [193]. Overall, the RF and ELM were suitable for toxic elements estimation using hyperspectral data. The estimation accuracy was significantly improved by using the DS algorithm. For Zn, Ni, and Cu, the R2v were 0.77 (RF), 0.62 (RF), and 0.56 (ELM), respectively [158].

Also, Data collection from soil samples to measure the accuracy of identification of satellite imagery is needed. Data pre-processing, variable construction model selection, validation, and model quality assessment are the essential parts of this issue. The performance of the prediction models in both lab reflectance spectroscopy and satellite imaging spectroscopy was assessed by comparing the predicted values on the validation data set with the observed ones using the metrics; correlation coefficient (R2), Root Mean Square Error (RMSE), and Residual Prediction Deviation (RPD)[60].

Finally, this study reviewed the potential of recently developed and forthcoming spaceborne sensors for soil contamination monitoring and the current limitations of spaceborne sensors for soil contamination monitoring. However, contaminants with different chemical properties induce similar responses on plant reflectance, so it is not yet possible to clearly distinguish the presence of complex mixtures from that of a single contaminant [87]. Moreover, obstacles such as cloud and vegetation coverage must be dealt with. A limited number of studies attempted to explore soil contamination through vegetation spectral data; most of those were based on a single vegetation species (Lassalle et al., 2018; Shi et al., 2016).

1. **References**

[1] “Shi, T., Wang, J., Chen, Y., and Wu, G. (2016). Improving the prediction of arsenic contents in agricultural soils by combining the reflectance spectroscopy of soils and rice plants. International Journal of Applied Earth Observation and Geoinformation 52.”

[2] K. M. Banat, F. M. Howari, and A. A. Al-Hamad, “Heavy metals in urban soils of central Jordan: Should we worry about their environmental risks?,” *Environ. Res.*, vol. 97, no. 3, pp. 258–273, 2005, doi: 10.1016/j.envres.2004.07.002.

[3] P. K. Sahoo, S. M. Equeenuddin, and M. A. Powell, “Trace Elements in Soils around Coal Mines: Current Scenario, Impact and Available Techniques for Management,” *Curr. Pollut. Reports*, vol. 2, no. 1, 2016, doi: 10.1007/s40726-016-0025-5.

[4] Z. Zhang, J. Abuduwaili, and F. Jiang, “Determination of Occurrence Characteristics of Heavy Metals in Soil and Water Environments in Tianshan Mountains, Central Asia,” *Anal. Lett.*, vol. 46, no. 13, pp. 2122–2131, 2013, doi: 10.1080/00032719.2013.784919.

[5] X. L. Xie, X. Z. Pan, and B. Sun, “Visible and Near-Infrared Diffuse Reflectance Spectroscopy for Prediction of Soil Properties near a Copper Smelter,” *Pedosphere*, vol. 22, no. 3, pp. 351–366, 2012, doi: 10.1016/S1002-0160(12)60022-8.

[6] H. O. Rashid *et al.*, “Impact of coal mining on soil, water and agricultural crop production: a cross-sectional study on Barapukuria coal mine industry, Dinajpur, Bangladesh,” *J. Environ. Sci. Res.*, vol. 1, no. 1, p. 0000001, 2014, [Online]. Available: https://www.researchgate.net/publication/261366522\_Impact\_of\_coal\_mining\_on\_soil\_water\_and\_agricultural\_crop\_production\_a\_cross-sectional\_study\_on\_Barapukuria\_coal\_mine\_industry\_Dinajpur\_Bangladesh.

[7] M. F. Noomen, A. K. Skidmore, F. D. van der Meer, and H. H. T. Prins, “Continuum removed band depth analysis for detecting the effects of natural gas, methane and ethane on maize reflectance,” *Remote Sens. Environ.*, vol. 105, no. 3, pp. 262–270, 2006, doi: 10.1016/j.rse.2006.07.009.

[8] F. van der Meer, P. van Dijk, H. van der Werff, and H. Yang, “Remote sensing and petroleum seepage: A review and case study,” *Terra Nov.*, vol. 14, no. 1, pp. 1–17, 2002, doi: 10.1046/j.1365-3121.2002.00390.x.

[9] I. D. Sanches, C. R. Souza Filho, L. A. Magalhães, G. C. M. Quitério, M. N. Alves, and W. J. Oliveira, “Unravelling remote sensing signatures of plants contaminated with gasoline and diesel: An approach using the red edge spectral feature,” *Environ. Pollut.*, vol. 174, pp. 16–27, 2013, doi: 10.1016/j.envpol.2012.10.029.

[10] M. F. Noomen, H. M. A. van der Werff, and F. D. van der Meer, “Spectral and spatial indicators of botanical changes caused by long-term hydrocarbon seepage,” *Ecol. Inform.*, vol. 8, pp. 55–64, 2012, doi: 10.1016/j.ecoinf.2012.01.001.

[11] D. Schumacher, “Hydrocarbon-induced alteration of soils and sediments,” *AAPG Mem.*, no. 66, pp. 71–89, 1996, doi: 10.1306/m66606c6.

[12] M. D. Steven, K. L. Smith, M. D. Beardsley, and J. J. Colls, “Oxygen and methane depletion in soil affected by leakage of natural gas,” *Eur. J. Soil Sci.*, vol. 57, no. 6, pp. 800–807, 2006, doi: 10.1111/j.1365-2389.2005.00770.x.

[13] J. C. Zinnert, S. M. Via, and D. R. Young, “Distinguishing natural from anthropogenic stress in plants: Physiology, fluorescence and hyperspectral reflectance,” *Plant Soil*, vol. 366, no. 1–2, pp. 133–141, 2013, doi: 10.1007/s11104-012-1414-1.

[14] H. D. Roelofsen, P. M. van Bodegom, L. Kooistra, J. J. van Amerongen, and J. P. M. Witte, “An evaluation of remote sensing derived soil pH and average spring groundwater table for ecological assessments,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 43, pp. 149–159, 2015, doi: 10.1016/j.jag.2015.05.005.

[15] A. Akcil and S. Koldas, “Acid Mine Drainage (AMD): causes, treatment and case studies,” *J. Clean. Prod.*, vol. 14, no. 12-13 SPEC. ISS., pp. 1139–1145, 2006, doi: 10.1016/j.jclepro.2004.09.006.

[16] M. E. P. Gomes and P. J. C. Favas, “Mineralogical controls on mine drainage of the abandoned Ervedosa tin mine in north-eastern Portugal,” *Appl. Geochemistry*, vol. 21, no. 8, pp. 1322–1334, 2006, doi: 10.1016/j.apgeochem.2006.06.007.

[17] A. Gholizadeh *et al.*, “Estimation of potentially toxic elements contamination in anthropogenic soils on a brown coal mining dumpsite by reflectance spectroscopy: A case study,” *PLoS One*, vol. 10, no. 2, 2015, doi: 10.1371/journal.pone.0117457.

[18] F. Baruthio, “Toxic effects of chromium and its compounds,” *Biol. Trace Elem. Res.*, vol. 32, no. 1–3, pp. 145–153, 1992, doi: 10.1007/BF02784599.

[19] G. A. Swayze *et al.*, “Using imaging spectroscopy to map acidic mine waste,” *Environ. Sci. Technol.*, vol. 34, no. 1, pp. 47–54, 2000, doi: 10.1021/es990046w.

[20] M. Bussinow, B. Sarapatka, and P. Dlapa, “Chemical degradation of forest soil as a result of polymetallic ore mining activities,” *Polish J. Environ. Stud.*, vol. 21, no. 6, pp. 1551–1561, 2012.

[21] Y. Song, F. Li, Z. Yang, G. A. Ayoko, R. L. Frost, and J. Ji, “Diffuse reflectance spectroscopy for monitoring potentially toxic elements in the agricultural soils of Changjiang River Delta, China,” *Appl. Clay Sci.*, vol. 64, pp. 75–83, 2012, doi: 10.1016/j.clay.2011.09.010.

[22] E. Galunin *et al.*, “Cadmium mobility in sediments and soils from a coal mining area on tibagi river watershed: Environmental risk assessment,” *J. Hazard. Mater.*, vol. 265, pp. 280–287, 2014, doi: 10.1016/j.jhazmat.2013.11.010.

[23] G. Tozsin, “Hazardous elements in soil and coal from the Oltu coal mine district, Turkey,” *Int. J. Coal Geol.*, vol. 131, pp. 1–6, 2014, doi: 10.1016/j.coal.2014.05.011.

[24] D. Pentari, J. Typou, F. Goodarzi, and A. E. Foscolos, “Comparison of elements of environmental concern in regular and reclaimed soils, near abandoned coal mines Ptolemais-Amynteon, northern Greece: Impact on wheat crops,” *Int. J. Coal Geol.*, vol. 65, no. 1–2, pp. 51–58, 2006, doi: 10.1016/j.coal.2005.04.008.

[25] J. Ribeiro, E. Ferreira da Silva, Z. Li, C. Ward, and D. Flores, “Petrographic, mineralogical and geochemical characterization of the Serrinha coal waste pile (Douro Coalfield, Portugal) and the potential environmental impacts on soil, sediments and surface waters,” *Int. J. Coal Geol.*, vol. 83, no. 4, pp. 456–466, 2010, doi: 10.1016/j.coal.2010.06.006.

[26] M. Pietrzykowski, J. Socha, and N. S. van Doorn, “Linking heavy metal bioavailability (Cd, Cu, Zn and Pb) in Scots pine needles to soil properties in reclaimed mine areas,” *Sci. Total Environ.*, vol. 470–471, pp. 501–510, 2014, doi: 10.1016/j.scitotenv.2013.10.008.

[27] Z. Ge, Y., Cui, X., and Bai, “Evaluation on potential ecological risk of heavy metals pollution in reclaimed soil of opencast-taking Pingshuo opencast mine as an example.”

[28] M. You, Y. Huang, J. Lu, and C. Li, “Characterization of Heavy Metals in Soil Near Coal Mines and a Power Plant in Huainan, China,” *Anal. Lett.*, vol. 48, no. 4, pp. 726–737, 2015, doi: 10.1080/00032719.2014.940531.

[29] S. K. Das and G. J. Chakrapani, “Assessment of trace metal toxicity in soils of Raniganj Coalfield, India,” *Environ. Monit. Assess.*, vol. 177, no. 1–4, pp. 63–71, 2011, doi: 10.1007/s10661-010-1618-x.

[30] S. M. Equeenduddin, “Controls of coal and overburden on acid mine drainage and metal mobilization at Makum Coalfield, Assam, India,” 2010.

[31] E. Choe, K. W. Kim, S. Bang, I. H. Yoon, and K. Y. Lee, “Qualitative analysis and mapping of heavy metals in an abandoned Au-Ag mine area using NIR spectroscopy,” *Environ. Geol.*, vol. 58, no. 3, pp. 477–482, 2009, doi: 10.1007/s00254-008-1520-9.

[32] L. Kooistra, R. Wehrens, R. S. E. W. Leuven, and L. M. C. Buydens, “Possibilities of visible-near-infrared spectroscopy for the assessment of soil contamination in river floodplains,” *Anal. Chim. Acta*, vol. 446, no. 1–2, pp. 97–105, 2001, doi: 10.1016/S0003-2670(01)01265-X.

[33] J. Chen, F. Wei, C. Zheng, Y. Wu, and D. C. Adriano, “Background concentrations of elements in soils of China,” *Water. Air. Soil Pollut.*, vol. 57–58, no. 1, pp. 699–712, 1991, doi: 10.1007/BF00282934.

[34] H. T. Shacklette and J. G. Boerngen, “Element Concentrations in Soils and Other Surficial Materials of the Conterminous United States.,” *Geol. Surv. Prof. Pap. (United States)*, 1984.

[35] CCME, “Soil Quality,” pp. 960–960, 2021, doi: 10.1007/978-3-319-95981-8\_300152.

[36] G. Mance, “Pollution threat of heavy metals in aquatic environments,” *Pollut. Threat heavy Met. Aquat. Environ.*, 1987, doi: 10.1016/s0003-2670(00)80771-0.

[37] A. Kabata-Pendias, “Trace elements in soils and plants: Fourth edition,” *Trace Elem. Soils Plants, Fourth Ed.*, pp. 1–520, 2010, doi: 10.1201/b10158.

[38] N. J. Cook *et al.*, “Trace and minor elements in sphalerite: A LA-ICPMS study,” *Geochim. Cosmochim. Acta*, vol. 73, no. 16, pp. 4761–4791, 2009, doi: 10.1016/j.gca.2009.05.045.

[39] R. L. Zimdahl and R. K. Skogerboe, “Behavior of Lead in Soil,” *Environ. Sci. Technol.*, vol. 11, no. 13, pp. 1202–1207, 1977, doi: 10.1021/es60136a004.

[40] T. Asami, “Pollution of soils by cadmium p. 95–111. In Changing metal cycles and human health. Springer.lution of soils by cadmium,” *Chang. Met. cycles Hum. Heal.*, pp. 95–111, 1984.

[41] E. Scudiero, T. H. Skaggs, and D. L. Corwin, “Regional-scale soil salinity assessment using Landsat ETM+ canopy reflectance,” *Remote Sens. Environ.*, vol. 169, pp. 335–343, 2015, doi: 10.1016/j.rse.2015.08.026.

[42] Z. Liu *et al.*, “Hyperspectral discrimination and response characteristics of stressed rice leaves caused by rice leaf folder,” *IFIP Adv. Inf. Commun. Technol.*, vol. 369 AICT, no. PART 2, pp. 528–537, 2012, doi: 10.1007/978-3-642-27278-3\_54.

[43] E. Scudiero, P. Teatini, D. L. Corwin, N. Dal Ferro, G. Simonetti, and F. Morari, “Spatiotemporal response of maize yield to edaphic and meteorological conditions in a saline farmland,” *Agron. J.*, vol. 106, no. 6, pp. 2163–2174, 2014, doi: 10.2134/agronj14.0102.

[44] H. Y. REN, D. F. ZHUANG, A. N. SINGH, J. J. PAN, D. S. QIU, and R. H. SHI, “Estimation of As and Cu Contamination in Agricultural Soils Around a Mining Area by Reflectance Spectroscopy: A Case Study,” *Pedosphere*, vol. 19, no. 6, pp. 719–726, 2009, doi: 10.1016/S1002-0160(09)60167-3.

[45] K. L. Smith, M. D. Steven, and J. J. Colls, “Spectral responses of pot-grown plants to displacement of soil oxygen,” *Int. J. Remote Sens.*, vol. 25, no. 20, pp. 4395–4410, 2004, doi: 10.1080/01431160410001729172.

[46] A. Lausch, S. Erasmi, D. J. King, P. Magdon, and M. Heurich, “Understanding forest health with remote sensing-Part I-A review of spectral traits, processes and remote-sensing characteristics,” *Remote Sens.*, vol. 8, no. 12, 2016, doi: 10.3390/rs8121029.

[47] A. Gholizadeh, M. Saberioon, E. Ben-Dor, and L. Borůvka, “Monitoring of selected soil contaminants using proximal and remote sensing techniques: Background, state-of-the-art and future perspectives,” *Crit. Rev. Environ. Sci. Technol.*, vol. 48, no. 3, pp. 243–278, 2018, doi: 10.1080/10643389.2018.1447717.

[48] S. Chakraborty *et al.*, “Development of a hybrid proximal sensing method for rapid identification of petroleum contaminated soils,” *Sci. Total Environ.*, vol. 514, pp. 399–408, 2015, doi: 10.1016/j.scitotenv.2015.01.087.

[49] A. Gholizadeh and V. Kopačková, “Detecting vegetation stress as a soil contamination proxy\_ a review of optical proximal and remote sensing techniques,” *International Journal of Environmental Science and Technology*. 2019, doi: 10.1007/s13762-019-02310-w.

[50] T. Shi, H. Liu, Y. Chen, J. Wang, and G. Wu, “Estimation of arsenic in agricultural soils using hyperspectral vegetation indices of rice,” *J. Hazard. Mater.*, vol. 308, pp. 243–252, 2016, doi: 10.1016/j.jhazmat.2016.01.022.

[51] T. Shi *et al.*, “Proximal and remote sensing techniques for mapping of soil contamination with heavy metals,” *Appl. Spectrosc. Rev.*, vol. 53, no. 10, pp. 783–805, 2018, doi: 10.1080/05704928.2018.1442346.

[52] T. Shi, Y. Chen, Y. Liu, and G. Wu, “Visible and near-infrared reflectance spectroscopy-An alternative for monitoring soil contamination by heavy metals,” *J. Hazard. Mater.*, vol. 265, pp. 166–176, 2014, doi: 10.1016/j.jhazmat.2013.11.059.

[53] S. Chakraborty *et al.*, “Rapid Identification of Oil-Contaminated Soils Using Visible Near-Infrared Diffuse Reflectance Spectroscopy,” *J. Environ. Qual.*, vol. 39, no. 4, pp. 1378–1387, 2010, doi: 10.2134/jeq2010.0183.

[54] R. K. Douglas, S. Nawar, S. Cipullo, M. C. Alamar, F. Coulon, and A. M. Mouazen, “Evaluation of vis-NIR reflectance spectroscopy sensitivity to weathering for enhanced assessment of oil contaminated soils,” *Sci. Total Environ.*, vol. 626, pp. 1108–1120, 2018, doi: 10.1016/j.scitotenv.2018.01.122.

[55] M. Liu, T. Wang, A. K. Skidmore, and X. Liu, “Heavy metal-induced stress in rice crops detected using multi-temporal Sentinel-2 satellite images,” *Sci. Total Environ.*, vol. 637–638, pp. 18–29, 2018, doi: 10.1016/j.scitotenv.2018.04.415.

[56] Y. Wu *et al.*, “A Mechanism Study of Reflectance Spectroscopy for Investigating Heavy Metals in Soils,” *Soil Sci. Soc. Am. J.*, vol. 71, no. 3, pp. 918–926, 2007, doi: 10.2136/sssaj2006.0285.

[57] Q. X. Xue, Q. M. Yu, F. J. Jun, R. M. Hong, J. Chen, and L. L. Qi, “Reflectance spectroscopy study of Cd contamination in the sediments of the Changjiang River, China,” *Environ. Sci. Technol.*, vol. 41, no. 10, pp. 3449–3454, 2007, doi: 10.1021/es0624422.

[58] Y. Z. Wu, J. Chen, J. F. Ji, Q. J. Tian, and X. M. Wu, “Feasibility of reflectance spectroscopy for the assessment of soil mercury contamination,” *Environ. Sci. Technol.*, vol. 39, no. 3, pp. 873–878, 2005, doi: 10.1021/es0492642.

[59] L. Zhao *et al.*, “Estimation methods for soil mercury content using hyperspectral remote sensing,” *Sustain.*, vol. 10, no. 7, 2018, doi: 10.3390/su10072474.

[60] V. Khosravi, F. Doulati Ardejani, A. Gholizadeh, and M. Saberioon, “Satellite imagery for monitoring and mapping soil chromium pollution in a mine waste dump,” *Remote Sens.*, vol. 13, no. 7, 2021, doi: 10.3390/rs13071277.

[61] E. Choe, F. van der Meer, F. van Ruitenbeek, H. van der Werff, B. de Smeth, and K. W. Kim, “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.*, vol. 112, no. 7, pp. 3222–3233, 2008, doi: 10.1016/j.rse.2008.03.017.

[62] A. Dube, R. Zbytniewski, T. Kowalkowski, E. Cukrowska, and B. Buszewski, “Adsorption and Migration of Heavy Metals in Soil,” *Polish J. Environ. Stud.*, vol. 10, no. 1, pp. 1–10, 2001.

[63] J. Wang, L. Cui, W. Gao, T. Shi, Y. Chen, and Y. Gao, “Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy,” *Geoderma*, vol. 216, pp. 1–9, 2014, doi: 10.1016/j.geoderma.2013.10.024.

[64] S. G. Asmaryan, V. S. Muradyan, L. V. Sahakyan, A. K. Saghatelyan, and T. Warner, “Development of remote sensing methods for assessing and mapping soil pollution with heavy metals,” *Glob. Basis Glob. Spat. Soil Inf. Syst. - Proc. 1st Glob. Conf.*, pp. 429–432, 2014, doi: 10.1201/b16500-77.

[65] S. T. Forrester, L. J. Janik, M. J. McLaughlin, J. M. Soriano-Disla, R. Stewart, and B. Dearman, “Total Petroleum Hydrocarbon Concentration Prediction in Soils Using Diffuse Reflectance Infrared Spectroscopy,” *Soil Sci. Soc. Am. J.*, vol. 77, no. 2, pp. 450–460, 2013, doi: 10.2136/sssaj2012.0201.

[66] A. Hauser, F. Ali, B. Al-Dosari, and H. Al-Sammar, “Solvent-free determination of TPH in soil by near-infrared reflectance spectroscopy,” *Int. J. Sustain. Dev. Plan.*, vol. 8, no. 3, pp. 413–421, 2013, doi: 10.2495/SDP-V8-N3-413-421.

[67] R. N. Okparanma and A. M. Mouazen, “Determination of total petroleum hydrocarbon (TPH) and polycyclic aromatic hydrocarbon (PAH) in soils: A review of spectroscopic and nonspectroscopic techniques,” *Appl. Spectrosc. Rev.*, vol. 48, no. 6, pp. 458–486, 2013, doi: 10.1080/05704928.2012.736048.

[68] E. Hobley, G. R. Willgoose, S. Frisia, and G. Jacobsen, “Vertical distribution of charcoal in a sandy soil: Evidence from DRIFT spectra and field emission scanning electron microscopy,” *Eur. J. Soil Sci.*, vol. 65, no. 5, pp. 751–762, 2014, doi: 10.1111/ejss.12171.

[69] W. Ng, B. P. Malone, and B. Minasny, “Rapid assessment of petroleum-contaminated soils with infrared spectroscopy,” *Geoderma*, vol. 289, pp. 150–160, 2017, doi: 10.1016/j.geoderma.2016.11.030.

[70] J. . J. Workman and L. Weyer, “Practical Guide to Interpretive Near-Infrared Spectroscopy,” *Pract. Guid. to Interpret. Near-Infrared Spectrosc.*, 2007, doi: 10.1201/9781420018318.

[71] R. A. V. Rossel and T. Behrens, “Using data mining to model and interpret soil diffuse reflectance spectra,” *Geoderma*, vol. 158, no. 1–2, pp. 46–54, 2010, doi: 10.1016/j.geoderma.2009.12.025.

[72] R. N. Okparanma, F. Coulon, T. Mayr, and A. M. Mouazen, “Mapping polycyclic aromatic hydrocarbon and total toxicity equivalent soil concentrations by visible and near-infrared spectroscopy,” *Environ. Pollut.*, vol. 192, pp. 162–170, 2014, doi: 10.1016/j.envpol.2014.05.022.

[73] A. Piccolo and F. J. Stevenson, “Infrared spectra of Cu2+ Pb2+ and Ca2+ complexes of soil humic substances,” *Geoderma*, vol. 27, no. 3, pp. 195–208, 1982, doi: 10.1016/0016-7061(82)90030-1.

[74] M. Egli, P. Fitze, and M. Oswald, “Changes in heavy metal contents in an acidic forest soil affected by depletion of soil organic matter within the time span 1969-93,” *Environ. Pollut.*, vol. 105, no. 3, pp. 367–379, 1999, doi: 10.1016/S0269-7491(99)00040-8.

[75] T. Chen, Q. Chang, J. G. P. W. Clevers, and L. Kooistra, “Rapid identification of soil cadmium pollution risk at regional scale based on visible and near-infrared spectroscopy,” *Environ. Pollut.*, vol. 206, pp. 217–226, 2015, doi: 10.1016/j.envpol.2015.07.009.

[76] S. Chakraborty *et al.*, “Rapid assessment of regional soil arsenic pollution risk via diffuse reflectance spectroscopy,” *Geoderma*, vol. 289, pp. 72–81, 2017, doi: 10.1016/j.geoderma.2016.11.024.

[77] L. M. Shuman, “Separating Soil Iron- and Manganese-Oxide Fractions for Microelement Analysis,” *Soil Sci. Soc. Am. J.*, vol. 46, no. 5, pp. 1099–1102, 1982, doi: 10.2136/sssaj1982.03615995004600050044x.

[78] T. Kemper and S. Sommer, “Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy,” *Environ. Sci. Technol.*, vol. 36, no. 12, pp. 2742–2747, 2002, doi: 10.1021/es015747j.

[79] Y. Wu, X. Zhang, Q. Liao, and J. Ji, “Can contaminant elements in soils be assessed by remote sensing technology: A case study with simulated data,” *Soil Sci.*, vol. 176, no. 4, pp. 196–205, 2011, doi: 10.1097/SS.0b013e3182114717.

[80] M. Ibrahim, A. J. Hameed, and A. Jalbout, “Molecular spectroscopic study of River Nile sediment in the greater Cairo region,” *Appl. Spectrosc.*, vol. 62, no. 3, pp. 306–311, 2008, doi: 10.1366/000370208783759795.

[81] J. Kumpiene, A. Lagerkvist, and C. Maurice, “Stabilization of Pb- and Cu-contaminated soil using coal fly ash and peat,” *Environ. Pollut.*, vol. 145, no. 1, pp. 365–373, 2007, doi: 10.1016/j.envpol.2006.01.037.

[82] P. Arellano, K. Tansey, H. Balzter, and D. S. Boyd, “Field spectroscopy and radiative transfer modelling to assess impacts of petroleum pollution on biophysical and biochemical parameters of the Amazon rainforest,” *Environ. Earth Sci.*, vol. 76, no. 5, 2017, doi: 10.1007/s12665-017-6536-6.

[83] G. Lassalle *et al.*, “Detection and discrimination of various oil-contaminated soils using vegetation reflectance,” *Sci. Total Environ.*, vol. 655, pp. 1113–1124, 2019, doi: 10.1016/j.scitotenv.2018.11.314.

[84] G. Lassalle *et al.*, “Application of PROSPECT for estimating total petroleum hydrocarbons in contaminated soils from leaf optical properties,” *J. Hazard. Mater.*, vol. 377, pp. 409–417, 2019, doi: 10.1016/j.jhazmat.2019.05.093.

[85] H. Van Der Werff, M. Van Der Meijde, F. Jansma, F. Van Der Meer, and G. J. Groothuis, “A spatial-spectral approach for visualization of vegetation stress resulting from pipeline leakage,” *Sensors*, vol. 8, no. 6, pp. 3733–3743, 2008, doi: 10.3390/s8063733.

[86] P. Arellano, K. Tansey, H. Balzter, and D. S. Boyd, “Detecting the effects of hydrocarbon pollution in the Amazon forest using hyperspectral satellite images,” *Environ. Pollut.*, vol. 205, pp. 225–239, 2015, doi: 10.1016/j.envpol.2015.05.041.

[87] G. Lassalle, A. Credoz, R. Hédacq, S. Fabre, D. Dubucq, and A. Elger, “Assessing Soil Contamination Due to Oil and Gas Production Using Vegetation Hyperspectral Reflectance,” *Environ. Sci. Technol.*, vol. 52, no. 4, pp. 1756–1764, 2018, doi: 10.1021/acs.est.7b04618.

[88] G. Lassalle *et al.*, “Toward quantifying oil contamination in vegetated areas using very high spatial and spectral resolution imagery,” *Remote Sens.*, vol. 11, no. 19, 2019, doi: 10.3390/rs11192241.

[89] I. D. Sanches, C. R. Souza Filho, L. A. Magalhães, G. C. M. Quitério, M. N. Alves, and W. J. Oliveira, “Assessing the impact of hydrocarbon leakages on vegetation using reflectance spectroscopy,” *ISPRS J. Photogramm. Remote Sens.*, vol. 78, pp. 85–101, 2013, doi: 10.1016/j.isprsjprs.2013.01.007.

[90] E. J. Emengini, G. A. Blackburn, and J. C. Theobald, “Early detection of oil-induced stress in crops using spectral and thermal responses,” *J. Appl. Remote Sens.*, vol. 7, no. 1, p. 073596, 2013, doi: 10.1117/1.jrs.7.073596.

[91] G. A. Blackburn, “Quantifying chlorophylls and carotenoids at leaf and canopy scales: An evaluation of some hyperspectral approaches,” *Remote Sens. Environ.*, vol. 66, no. 3, pp. 273–285, 1998, doi: 10.1016/S0034-4257(98)00059-5.

[92] J. M. Ourcival, R. Joffre, and S. Rambal, “Exploring the relationships between reflectance and anatomical and biochemical properties in Quercus ilex leaves,” *New Phytol.*, vol. 143, no. 2, pp. 351–364, 1999, doi: 10.1046/j.1469-8137.1999.00456.x.

[93] X. Cao, J. Wang, X. Chen, Z. Gao, F. Yang, and J. Shi, “Multiscale remote-sensing retrieval in the evapotranspiration of Haloxylon ammodendron in the Gurbantunggut desert, China,” *Environ. Earth Sci.*, vol. 69, no. 5, pp. 1549–1558, 2013, doi: 10.1007/s12665-012-1989-0.

[94] Y. Huang, Y. Hu, and Y. Liu, “Heavy metal accumulation in iron plaque and growth of rice plants upon exposure to single and combined contamination by copper, cadmium and lead,” *Acta Ecol. Sin.*, vol. 29, no. 6, pp. 320–326, 2009, doi: 10.1016/j.chnaes.2009.09.011.

[95] J. Jiang *et al.*, “Effects of multiple heavy metal contamination and repeated phytoextraction by Sedum plumbizincicola on soil microbial properties,” *Eur. J. Soil Biol.*, vol. 46, no. 1, pp. 18–26, 2010, doi: 10.1016/j.ejsobi.2009.10.001.

[96] P. H. Rosso, J. C. Pushnik, M. Lay, and S. L. Ustin, “Reflectance properties and physiological responses of Salicornia virginica to heavy metal and petroleum contamination,” *Environ. Pollut.*, vol. 137, no. 2, pp. 241–252, 2005, doi: 10.1016/j.envpol.2005.02.025.

[97] T. Shi, H. Liu, J. Wang, Y. Chen, T. Fei, and G. Wu, “Monitoring arsenic contamination in agricultural soils with reflectance spectroscopy of rice plants,” *Environ. Sci. Technol.*, vol. 48, no. 11, pp. 6264–6272, 2014, doi: 10.1021/es405361n.

[98] G. Lassalle, S. Fabre, A. Credoz, D. Dubucq, and A. Elger, “Monitoring oil contamination in vegetated areas with optical remote sensing: A comprehensive review,” *J. Hazard. Mater.*, vol. 393, 2020, doi: 10.1016/j.jhazmat.2020.122427.

[99] B. Adamu, K. Tansey, and B. Ogutu, “Remote sensing for detection and monitoring of vegetation affected by oil spills,” *Int. J. Remote Sens.*, vol. 39, no. 11, pp. 3628–3645, 2018, doi: 10.1080/01431161.2018.1448483.

[100] M. S. Ozigis, J. D. Kaduk, and C. H. Jarvis, “Mapping terrestrial oil spill impact using machine learning random forest and Landsat 8 OLI imagery: a case site within the Niger Delta region of Nigeria,” *Environ. Sci. Pollut. Res.*, vol. 26, no. 4, pp. 3621–3635, 2019, doi: 10.1007/s11356-018-3824-y.

[101] B. Adamu, K. Tansey, and B. Ogutu, “An investigation into the factors influencing the detectability of oil spills using spectral indices in an oil-polluted environment,” *Int. J. Remote Sens.*, vol. 37, no. 10, pp. 2338–2357, 2016, doi: 10.1080/01431161.2016.1176271.

[102] N. N. Onyia, H. Balzter, and J. C. Berrio, “Detecting vegetation response to oil pollution using hyperspectral indices,” *Int. Geosci. Remote Sens. Symp.*, vol. 2018-July, pp. 3963–3966, 2018, doi: 10.1109/IGARSS.2018.8519398.

[103] S. Jacquemoud, F. Baret, and J. F. Hanocq, “Modeling spectral and bidirectional soil reflectance,” *Remote Sens. Environ.*, vol. 41, no. 2–3, pp. 123–132, 1992, doi: 10.1016/0034-4257(92)90072-R.

[104] G. P. Asner, “Biophysical and biochemical sources of variability in canopy reflectance,” *Remote Sens. Environ.*, vol. 64, no. 3, pp. 234–253, 1998, doi: 10.1016/S0034-4257(98)00014-5.

[105] M. A. Cho, P. Debbab, O. Mutanga, N. Dudeni-Tlhoneb, T. Magadla, and S. A. Khuluseb, “Potential utility of the spectral red-edge region of SumbandilaSat imagery for assessing indigenous forest structure and health,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 16, no. 1, pp. 85–93, 2012, doi: 10.1016/j.jag.2011.12.005.

[106] G. Masaitis, G. Mozgeris, and A. Augustaitis, “Spectral reflectance properties of healthy and stressed coniferous trees,” *IForest*, vol. 6, no. JANUARY 2013, pp. 30–36, 2013, doi: 10.3832/ifor0709-006.

[107] M. F. Buitrago Acevedo, T. A. Groen, C. A. Hecker, and A. K. Skidmore, “Identifying leaf traits that signal stress in TIR spectra,” *ISPRS J. Photogramm. Remote Sens.*, vol. 125, pp. 132–145, 2017, doi: 10.1016/j.isprsjprs.2017.01.014.

[108] A. A. Gitelson, Y. Zur, O. B. Chivkunova, and M. N. Merzlyak, “Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy¶,” *Photochem. Photobiol.*, vol. 75, no. 3, p. 272, 2002, doi: 10.1562/0031-8655(2002)075<0272:accipl>2.0.co;2.

[109] Hoffer, A.M., “Biological and physical considerations in applying computer-aided analysis techniques to remote sensor data, in Remote Sensing: The Quantitative Approach,” *Remote Sens. Quant. approach*, vol. 5, pp. 227–289, 1978.

[110] L. Kumar, K. Schmidt, S. Dury, and A. Skidmore, “Imaging Spectrometry and Vegetation Science,” pp. 111–155, 2002, doi: 10.1007/978-0-306-47578-8\_5.

[111] D. N. H. Horler, M. Dockray, and J. Barber, “The red edge of plant leaf reflectance,” *Int. J. Remote Sens.*, vol. 4, no. 2, pp. 273–288, 1983, doi: 10.1080/01431168308948546.

[112] C. Zhang and J. M. Kovacs, “The application of small unmanned aerial systems for precision agriculture: A review,” *Precis. Agric.*, vol. 13, no. 6, pp. 693–712, 2012, doi: 10.1007/s11119-012-9274-5.

[113] Guyot. G, “Signatures spectrales des surfaces naturelles. Caen, Paradigme, France,” 1989.

[114] B. H. Bammel and R. W. Birnie, “Spectral reflectance response of big sagebrush to hydrocarbon-induced stress in the Bighorn Basin, Wyoming,” *Photogramm. Eng. Remote Sensing*, vol. 60, no. 1, pp. 87–96, 1994.

[115] K. L. Smith, M. D. Steven, and J. J. Colls, “Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks,” *Remote Sens. Environ.*, vol. 92, no. 2, pp. 207–217, 2004, doi: 10.1016/j.rse.2004.06.002.

[116] C. R. de Souza Filho, V. Augusto, W. J. Oliveira, and T. Lammoglia, “Detecção de exsudações de hidrocarbonetos por geobotânica e sensoriamento remoto multi-temporal: estudo de caso no Remanso do Fogo (MG),” *Rev. Bras. Geociências*, vol. 38, no. 2, pp. 228–243, 2008, doi: 10.25249/0375-7536.2008382s228243.

[117] E. J. Pell, “Multiple stress-induced foliar senescence and implications for whole-plant longevity,” *Response plants to Mult. Stress.*, 1991.

[118] S. C. Dunagan, M. S. Gilmore, and J. C. Varekamp, “Effects of mercury on visible/near-infrared reflectance spectra of mustard spinach plants (Brassica rapa P.),” *Environ. Pollut.*, vol. 148, no. 1, pp. 301–311, 2007, doi: 10.1016/j.envpol.2006.10.023.

[119] C. Götze, F. Beyer, and C. Gläßer, “Pioneer vegetation as an indicator of the geochemical parameters in abandoned mine sites using hyperspectral airborne data,” *Environ. Earth Sci.*, vol. 75, no. 7, 2016, doi: 10.1007/s12665-016-5367-1.

[120] L. Kooistra *et al.*, “Exploring field vegetation reflectance as an indicator of soil contamination in river floodplains,” *Environ. Pollut.*, vol. 127, no. 2, pp. 281–290, 2004, doi: 10.1016/S0269-7491(03)00266-5.

[121] J. G. P. W. Clevers, L. Kooistra, and E. A. L. Salas, “Study of heavy metal contamination in river floodplains using the red-edge position in spectroscopic data,” *Int. J. Remote Sens.*, vol. 25, no. 19, pp. 3883–3895, 2004, doi: 10.1080/01431160310001654473.

[122] X. Li, X. Liu, M. Liu, C. Wang, and X. Xia, “A hyperspectral index sensitive to subtle changes in the canopy chlorophyll content under arsenic stress,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 36, pp. 41–53, 2015, doi: 10.1016/j.jag.2014.10.017.

[123] J. Barce and C. Poschenrieder, “Plant water relations as affected by heavy metal stress: A review,” *J. Plant Nutr.*, vol. 13, no. 1, pp. 1–37, 1990, doi: 10.1080/01904169009364057.

[124] E. H. Larsson, J. F. Bornman, and H. Asp, “Influence of UV-B radiation and Cd2+ on chlorophyll fluorescence, growth and nutrient content in Brassica napus,” *J. Exp. Bot.*, vol. 49, no. 323, pp. 1031–1039, 1998, doi: 10.1093/jxb/49.323.1031.

[125] T. Rumpf, A. K. Mahlein, U. Steiner, E. C. Oerke, H. W. Dehne, and L. Plümer, “Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance,” *Comput. Electron. Agric.*, vol. 74, no. 1, pp. 91–99, 2010, doi: 10.1016/j.compag.2010.06.009.

[126] V. Barnett and T. Lewis, “Outliers in statistical data John Wiley and Sons,” *New York*, 1994.

[127] T. Cheng and Z. Li, “A multiscale approach for spatio-temporal outlier detection,” *Trans. GIS*, vol. 10, no. 2, pp. 253–263, 2006, doi: 10.1111/j.1467-9671.2006.00256.x.

[128] L. Tian, X. Liu, B. Zhang, M. Liu, and L. Wu, “Extraction of rice heavy metal stress signal features based on long time series leaf area index data using ensemble empirical mode decomposition,” *Int. J. Environ. Res. Public Health*, vol. 14, no. 9, 2017, doi: 10.3390/ijerph14091018.

[129] A. K. Phadikar, S., Sil, J., and Das, “Classification of Rice Leaf Diseases Based on Morphological Changes,” *Int. J. Inf. Electron. Eng.*, vol. 2, no. 3, pp. 460–463, 2012.

[130] S. Shekhar, C. T. Lu, and P. Zhang, “A unified approach to detecting spatial outliers,” *Geoinformatica*, vol. 7, no. 2, pp. 139–166, 2003, doi: 10.1023/A:1023455925009.

[131] P. Thenkabail, J. Lyon, and A. Huete, “Advances in Hyperspectral Remote Sensing of Vegetation and Agricultural Croplands,” *Hyperspectral Remote Sens. Veg.*, pp. 3–36, 2011, doi: 10.1201/b11222-3.

[132] N. Yokoya, J. C. W. Chan, and K. Segl, “Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated EnMAP and Sentinel-2 images,” *Remote Sens.*, vol. 8, no. 3, 2016, doi: 10.3390/rs8030172.

[133] M. A. Wulder and N. C. Coops, “Satellites: Make Earth observations open access.,” *Nature*, vol. 513, no. 7516, pp. 30–31, 2014, doi: 10.1038/513030a.

[134] T. Majasalmi and M. Rautiainen, “The potential of Sentinel-2 data for estimating biophysical variables in a boreal forest: A simulation study,” *Remote Sens. Lett.*, vol. 7, no. 5, pp. 427–436, 2016, doi: 10.1080/2150704X.2016.1149251.

[135] C. Ade and E. Hestir, “Remote sensing and GIs for ecologists: using open source software,” *Photogramm. Eng. Remote Sens.*, vol. 83, no. 6, pp. 391–392, 2017, doi: 10.14358/pers.83.6.391.

[136] R. Zurita-Milla, J. G. P. W. Clevers, and M. E. Schaepman, “Unmixing-based landsat TM and MERIS FR data fusion,” *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 453–457, 2008, doi: 10.1109/LGRS.2008.919685.

[137] J. Mišurec, V. Kopačková, Z. Lhotáková, P. Campbell, and J. Albrechtová, “Detection of spatio-temporal changes of Norway spruce forest stands in ore mountains using landsat time series and airborne hyperspectral imagery,” *Remote Sens.*, vol. 8, no. 2, 2016, doi: 10.3390/rs8020092.

[138] G. K. Arp, “An integrated interpretation for the origin of the Patrick Draw Oil Field sage anomaly,” *Am. Assoc. Pet. Geol. Bull.*, vol. 76, no. 3, pp. 301–306, 1992, doi: 10.1306/bdff87d6-1718-11d7-8645000102c1865d.

[139] M. Berger, J. Moreno, J. A. Johannessen, P. F. Levelt, and R. F. Hanssen, “ESA’s sentinel missions in support of Earth system science,” *Remote Sens. Environ.*, vol. 120, pp. 84–90, 2012, doi: 10.1016/j.rse.2011.07.023.

[140] K. Dalsted, J. F. Paris, D. E. Clay, S. a Clay, C. L. Reese, and J. Chang, “Selecting the Appropriate Satellite Remote Sensing Product for Precision Farming,” *Ssmg*, vol. 8, no. 3, pp. 1–6, 2003.

[141] S. Chakraborty, D. C. Weindorf, B. Li, M. N. Ali, K. Majumdar, and D. P. Ray, “Analysis of petroleum contaminated soils by spectral modeling and pure response profile recovery of n-hexane,” *Environ. Pollut.*, vol. 190, pp. 10–18, 2014, doi: 10.1016/j.envpol.2014.03.005.

[142] L. Hou, X. Li, and F. Li, “Hyperspectral-based Inversion of Heavy Metal Content in the Soil of Coal Mining Areas,” *J. Environ. Qual.*, vol. 48, no. 1, pp. 57–63, 2019, doi: 10.2134/jeq2018.04.0130.

[143] S. R. Stazi *et al.*, “Hyperspectral Visible–Near Infrared Determination of Arsenic Concentration in Soil,” *Commun. Soil Sci. Plant Anal.*, vol. 45, no. 22, pp. 2911–2920, 2014, doi: 10.1080/00103624.2014.954716.

[144] R. K. Douglas, S. Nawar, M. C. Alamar, A. M. Mouazen, and F. Coulon, “Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques,” *Sci. Total Environ.*, vol. 616–617, pp. 147–155, 2018, doi: 10.1016/j.scitotenv.2017.10.323.

[145] Q. Guan *et al.*, “Prediction of heavy metals in soils of an arid area based on multi-spectral data,” *J. Environ. Manage.*, vol. 243, pp. 137–143, 2019, doi: 10.1016/j.jenvman.2019.04.109.

[146] D. Roy, M. Wulder, and T. R. Loveland, “Landsat-8 : Science and Product Vision for Terrestrial Global Change Research Remote Sensing of Environment Landsat-8 : Science and product vision for terrestrial global change research,” *Remote Sens. Environ.*, vol. 145, no. December 2018, 2014, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003442571400042X%0Ahttps://www.sciencedirect.com/science/article/pii/S003442571400042X%0Ahttp://dx.doi.org/10.1016/j.rse.2014.02.001.

[147] F. Liu, X. Liu, L. Zhao, C. Ding, J. Jiang, and L. Wu, “The Dynamic Assessment Model for Monitoring Cadmium Stress Levels in Rice Based on the Assimilation of Remote Sensing and the WOFOST Model,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 3, pp. 1330–1338, 2015, doi: 10.1109/JSTARS.2014.2371058.

[148] Y. Peng *et al.*, “Digital mapping of toxic metals in qatari soils using remote sensing and ancillary data,” *Remote Sens.*, vol. 8, no. 12, 2016, doi: 10.3390/rs8121003.

[149] P. Liu *et al.*, “Integrating a hybrid back propagation neural network and particle swarm optimization for estimating soil heavy metal contents using hyperspectral data,” *Sustain.*, vol. 11, no. 2, 2019, doi: 10.3390/su11020419.

[150] A. F. H. Goetz and L. C. Rowan, “Geologic remote sensing,” *Science (80-. ).*, vol. 211, no. 4484, pp. 780–791, 1981, doi: 10.1126/science.211.4484.781.

[151] M. Abrams and S. J. Hook, “Simulated Aster data for geologic studies - Geoscience and Remote Sensing, IEEE Transactions on,” vol. 33, no. 3, pp. 692–699, 1998.

[152] M. Abrams, “The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER): Data products for the high spatial resolution imager on NASA’s Terra platform,” *Int. J. Remote Sens.*, vol. 21, no. 5, pp. 847–859, 2000, doi: 10.1080/014311600210326.

[153] Y. Yamaguchi, A. B. Kahle, H. Tsu, T. Kawakami, and M. Pniel, “Overview of advanced spaceborne thermal emission and reflection radiometer (ASTER),” *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1062–1071, 1998, doi: 10.1109/36.700991.

[154] J. C. Mars and L. C. Rowan, “Spectral assessment of new ASTER SWIR surface reflectance data products for spectroscopic mapping of rocks and minerals,” *Remote Sens. Environ.*, vol. 114, no. 9, pp. 2011–2025, 2010, doi: 10.1016/j.rse.2010.04.008.

[155] J. Lévesque, K. Staenz, and T. Szeredi, “The impact of spectral band characteristics on unmixing of hyperspectral data for monitoring mine tailings site rehabilitation,” *Can. J. Remote Sens.*, vol. 26, no. 3, pp. 231–240, 2000, doi: 10.1080/07038992.2000.10874772.

[156] S. Chevrel, A. Bourguignon, F. Cottard, and Y. Itard, “Exploitation of ASTER imagery in mining-related environmental management,” *Pecora 16 - Glob. Priorities L. Remote Sens.*, pp. 1–10, 2005.

[157] F. Mirzaei, Y. Abbasi, and T. Sohrabi, “Modeling the distribution of heavy metals in lands irrigated by wastewater using satellite images of Sentinel-2,” *Egypt. J. Remote Sens. Sp. Sci.*, vol. 24, no. 3, pp. 537–546, 2021, doi: 10.1016/j.ejrs.2021.03.002.

[158] B. Zhang, B. Guo, B. Zou, W. Wei, Y. Lei, and T. Li, “Retrieving soil heavy metals concentrations based on GaoFen-5 hyperspectral satellite image at an opencast coal mine, Inner Mongolia, China,” *Environ. Pollut.*, vol. 300, no. February, p. 118981, 2022, doi: 10.1016/j.envpol.2022.118981.

[159] K. Tan *et al.*, “Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning,” *J. Hazard. Mater.*, vol. 401, 2021, doi: 10.1016/j.jhazmat.2020.123288.

[160] L. Y. Yang, X. H. Gao, W. Zhang, F. F. Shi, L. H. He, and W. Jia, “Estimating heavy metal concentrations in topsoil from vegetation reflectance spectra of Hyperion images: A case study of Yushu County, Qinghai, China,” *Chinese J. Appl. Ecol.*, vol. 27, no. 6, pp. 1775–1784, 2016, doi: 10.13287/j.1001-9332.201606.030.

[161] W. Liu *et al.*, “Hyperspectral inversion of mercury in reed leaves under different levels of soil mercury contamination,” *Environ. Sci. Pollut. Res.*, vol. 27, no. 18, pp. 22935–22945, 2020, doi: 10.1007/s11356-020-08807-z.

[162] B. Dkhala, N. Mezned, C. Gomez, and S. Abdeljaouad, “Hyperspectral field spectroscopy and SENTINEL-2 Multispectral data for minerals with high pollution potential content estimation and mapping,” *Sci. Total Environ.*, vol. 740, 2020, doi: 10.1016/j.scitotenv.2020.140160.

[163] C. Mielke, N. K. Boesche, C. Rogass, H. Kaufmann, C. Gauert, and M. de Wit, “Spaceborne mine waste mineralogy monitoring in South Africa, applications for modern push-broom missions: Hyperion/OLI and EnMAP/Sentinel-2,” *Remote Sens.*, vol. 6, no. 8, pp. 6790–6816, 2014, doi: 10.3390/rs6086790.

[164] D. Rogge, B. Rivard, K. Segl, B. Grant, and J. Feng, “Mapping of NiCu-PGE ore hosting ultramafic rocks using airborne and simulated EnMAP hyperspectral imagery, Nunavik, Canada,” *Remote Sens. Environ.*, vol. 152, pp. 302–317, 2014, doi: 10.1016/j.rse.2014.06.024.

[165] F. A. Kruse and S. L. Perry, “Mineral mapping using simulated worldview-3 short-wave-infrared imagery,” *Remote Sens.*, vol. 5, no. 6, pp. 2688–2703, 2013, doi: 10.3390/rs5062688.

[166] A. Agrawal and M. R. Petersen, “Detecting arsenic contamination using satellite imagery and machine learning,” *Toxics*, vol. 9, no. 12, 2021, doi: 10.3390/toxics9120333.

[167] G. T. Webster *et al.*, “Rapid prediction of total petroleum hydrocarbons in soil using a hand-held mid-infrared field instrument,” *Talanta*, vol. 160, pp. 410–416, 2016, doi: 10.1016/j.talanta.2016.07.044.

[168] A. Gholizadeh, D. Žižala, M. Saberioon, and L. Borůvka, “Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging,” *Remote Sens. Environ.*, vol. 218, pp. 89–103, 2018, doi: 10.1016/j.rse.2018.09.015.

[169] Y. Q. Song, X. Zhao, H. Y. Su, B. Li, Y. M. Hu, and X. Sen Cui, “Predicting spatial variations in soil nutrients with hyperspectral remote sensing at regional scale,” *Sensors (Switzerland)*, vol. 18, no. 9, 2018, doi: 10.3390/s18093086.

[170] C. Laberge, D. Cluis, and G. Mercier, “Metal bioleaching prediction in continuous processing of municipal sewage with Thiobacillus ferrooxidans using neural networks,” *Water Res.*, vol. 34, no. 4, pp. 1145–1156, 2000, doi: 10.1016/S0043-1354(99)00246-8.

[171] O. Abedinia, N. Amjady, and N. Ghadimi, “Solar energy forecasting based on hybrid neural network and improved metaheuristic algorithm,” *Comput. Intell.*, vol. 34, no. 1, pp. 241–260, 2018, doi: 10.1111/coin.12145.

[172] J. Park *et al.*, “Real time vehicle speed prediction using a Neural Network Traffic Model,” *Proc. Int. Jt. Conf. Neural Networks*, pp. 2991–2996, 2011, doi: 10.1109/IJCNN.2011.6033614.

[173] Z. Liu, Y. Lu, Y. Peng, L. Zhao, G. Wang, and Y. Hu, “Estimation of soil heavy metal content using hyperspectral data,” *Remote Sens.*, vol. 11, no. 12, 2019, doi: 10.3390/rs11121464.

[174] S. Tian *et al.*, “Hyperspectral prediction model of metal content in soil based on the genetic ant colony algorithm,” *Sustain.*, vol. 11, no. 11, 2019, doi: 10.3390/su11113197.

[175] X. Meng *et al.*, “Regional soil organic carbon prediction model based on a discrete wavelet analysis of hyperspectral satellite data,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 89, 2020, doi: 10.1016/j.jag.2020.102111.

[176] L. Boszke and A. Astel, “Application of neural-based modeling in an assessment of pollution with mercury in the middle part of the Warta River,” *Environ. Monit. Assess.*, vol. 152, no. 1–4, pp. 133–147, 2009, doi: 10.1007/s10661-008-0302-x.

[177] O. Şenkal, “Modeling of solar radiation using remote sensing and artificial neural network in Turkey,” *Energy*, vol. 35, no. 12, pp. 4795–4801, 2010, doi: 10.1016/j.energy.2010.09.009.

[178] Z. (2015). Wu, L., Wang, Y., Long, J., & Liu, “An unsupervised change detection approach for remote sensing image using principal component analysis and genetic algorithm. In Image and graphics (pp. 589–602). New York City: Springer International Publishing.”

[179] Q. Miao, R. Liu, Y. Quan, and J. Song, “Remote sensing image fusion based on shearlet and genetic algorithm,” *Int. J. Bio-Inspired Comput.*, vol. 9, no. 4, pp. 240–250, 2017, doi: 10.1504/IJBIC.2017.084330.

[180] R. Lasaponara, G. Leucci, N. Masini, R. Persico, and G. Scardozzi, “Towards an operative use of remote sensing for exploring the past using satellite data: The case study of Hierapolis (Turkey),” *Remote Sens. Environ.*, vol. 174, pp. 148–164, 2016, doi: 10.1016/j.rse.2015.12.016.

[181] Z. Xiao *et al.*, “Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 209–223, 2014, doi: 10.1109/TGRS.2013.2237780.

[182] Y. Icaga, “Genetic algorithm usage in water quality monitoring networks optimization in Gediz (Turkey) River Basin,” *Environ. Monit. Assess.*, vol. 108, no. 1–3, pp. 261–277, 2005, doi: 10.1007/s10661-005-4328-z.

[183] C. L. Chang, S. L. Lo, and S. L. Yu, “The parameter optimization in the inverse distance method by genetic algorithm for estimating precipitation,” *Environ. Monit. Assess.*, vol. 117, no. 1–3, pp. 145–155, 2006, doi: 10.1007/s10661-006-8498-0.

[184] P. Zhou, Y. Zhao, Z. Zhao, and T. Chai, “Source mapping and determining of soil contamination by heavy metals using statistical analysis, artificial neural network, and adaptive genetic algorithm,” *J. Environ. Chem. Eng.*, vol. 3, no. 4, pp. 2569–2579, 2015, doi: 10.1016/j.jece.2015.08.003.

[185] A. Nadari, M. A. Delavar, B. Kaboudin, and M. S. Askari, “Assessment of spatial distribution of soil heavy metals using ANN-GA, MSLR and satellite imagery,” *Environ. Monit. Assess.*, vol. 189, no. 5, 2017, doi: 10.1007/s10661-017-5821-x.

[186] L. Anselin, “Local Indicators of Spatial Association—LISA,” *Geogr. Anal.*, vol. 27, no. 2, pp. 93–115, 1995, doi: 10.1111/j.1538-4632.1995.tb00338.x.

[187] A. Getis and J. K. Ord, “The analysis of spatial association by use of distance statistics,” *Adv. Spat. Sci.*, vol. 61, pp. 127–145, 2010, doi: 10.1007/978-3-642-01976-0\_10.

[188] K. Ellis, J. Kerr, S. Godbole, G. Lanckriet, D. Wing, and S. Marshall, “A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers,” *Physiol. Meas.*, vol. 35, no. 11, pp. 2191–2203, 2014, doi: 10.1088/0967-3334/35/11/2191.

[189] V. Svetnik, A. Liaw, C. Tong, J. Christopher Culberson, R. P. Sheridan, and B. P. Feuston, “Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling,” *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 1947–1958, 2003, doi: 10.1021/ci034160g.

[190] X. Zhu *et al.*, “Machine learning exploration of the critical factors for CO2 adsorption capacity on porous carbon materials at different pressures,” *J. Clean. Prod.*, vol. 273, 2020, doi: 10.1016/j.jclepro.2020.122915.

[191] L. Wei, Z. Yuan, Y. Zhong, L. Yang, X. Hu, and Y. Zhang, “An improved gradient boosting regression tree estimation model for soil heavy metal (Arsenic) pollution monitoring using hyperspectral remote sensing,” *Appl. Sci.*, vol. 9, no. 9, 2019, doi: 10.3390/app9091943.

[192] L. Guo *et al.*, “Mapping field-scale soil organic carbon with unmanned aircraft system-acquired time series multispectral images,” *Soil Tillage Res.*, vol. 196, 2020, doi: 10.1016/j.still.2019.104477.

[193] V. Khosravi, F. Doulati Ardejani, S. Yousefi, and A. Aryafar, “Monitoring soil lead and zinc contents via combination of spectroscopy with extreme learning machine and other data mining methods,” *Geoderma*, vol. 318, pp. 29–41, 2018, doi: 10.1016/j.geoderma.2017.12.025.

[194] J. shan, J. Zhao, L. Liu, Y. Zhang, X. Wang, and F. Wu, “A novel way to rapidly monitor microplastics in soil by hyperspectral imaging technology and chemometrics,” *Environ. Pollut.*, vol. 238, pp. 121–129, 2018, doi: 10.1016/j.envpol.2018.03.026.

[195] Fabian Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011, [Online]. Available: http://scikit-learn.sourceforge.net.

[196] 2019. Pelta, R., Ben-Dor, E., “Assessing the detection limit of petroleum hydrocarbon in soils using hyperspectral remote-sensing. Rem. Sens. Environ. 224, 145e153.”

[197] B. B. M. Sridhar, R. K. Vincent, S. J. Roberts, and K. Czajkowski, “Remote sensing of soybean stress as an indicator of chemical concentration of biosolid amended surface soils,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 13, no. 4, pp. 676–681, 2011, doi: 10.1016/j.jag.2011.04.005.

[198] Q. Shen, K. Xia, S. Zhang, C. Kong, Q. Hu, and S. Yang, “Hyperspectral indirect inversion of heavy-metal copper in reclaimed soil of iron ore area,” *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 222, 2019, doi: 10.1016/j.saa.2019.117191.

[199] D. J. Brown, K. D. Shepherd, M. G. Walsh, M. Dewayne Mays, and T. G. Reinsch, “Global soil characterization with VNIR diffuse reflectance spectroscopy,” *Geoderma*, vol. 132, no. 3–4, pp. 273–290, 2006, doi: 10.1016/j.geoderma.2005.04.025.

[200] A. Stevens, M. Nocita, G. Tóth, L. Montanarella, and B. van Wesemael, “Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy,” *PLoS One*, vol. 8, no. 6, 2013, doi: 10.1371/journal.pone.0066409.

[201] R. A. V. Rossel and R. Webster, “Predicting soil properties from the Australian soil visible-near infrared spectroscopic database,” *Eur. J. Soil Sci.*, vol. 63, no. 6, pp. 848–860, 2012, doi: 10.1111/j.1365-2389.2012.01495.x.

[202] Z. Rossel, R.V., Behrens, T., Ben-Dor, E., Brown, D., Dematt^e, J., Shepherd, K.D., Shi, “A global spectral library to characterize the world’s soil-Web of Science Core Collection,” [Online]. Available: https://www-webofscience-com.acceso.unicauca.edu.co/wos/woscc/full-record/WOS:000374624800011.

[203] R. Pelta and E. Ben-Dor, “Assessing the detection limit of petroleum hydrocarbon in soils using hyperspectral remote-sensing,” *Remote Sens. Environ.*, vol. 224, pp. 145–153, 2019, doi: 10.1016/j.rse.2019.01.026.

[204] K. Liu, D. Zhao, J. yong Fang, X. Zhang, Q. yun Zhang, and X. ke Li, “Estimation of Heavy-Metal Contamination in Soil Using Remote Sensing Spectroscopy and a Statistical Approach,” *J. Indian Soc. Remote Sens.*, vol. 45, no. 5, pp. 805–813, 2017, doi: 10.1007/s12524-016-0648-4.

[205] T. Zhou, Z. Li, and J. Pan, “Multi-feature classification of multi-sensor satellite imagery based on dual-polarimetric sentinel-1A, landsat-8 OLI, and hyperion images for urban land-cover classification,” *Sensors (Switzerland)*, vol. 18, no. 2, 2018, doi: 10.3390/s18020373.

[206] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, “Sen2Cor for Sentinel-2,” p. 3, 2017, doi: 10.1117/12.2278218.

[207] M. F. Kaiser, H. Aboulela, H. El Serehy, and H. E. Edin, “Spectral enhancement of SPOT imagery data to assess marine pollution near Port Said, Egypt,” *Int. J. Remote Sens.*, vol. 31, no. 7, pp. 1753–1764, 2010, doi: 10.1080/01431160902926624.

[208] T. Slonecker, G. B. Fisher, D. P. Aiello, and B. Haack, “Visible and infrared remote imaging of hazardous waste: A review,” *Remote Sens.*, vol. 2, no. 11, pp. 2474–2508, 2010, doi: 10.3390/rs2112474.

[209] W. Ji, R. A. Viscarra Rossel, and Z. Shi, “Improved estimates of organic carbon using proximally sensed vis-NIR spectra corrected by piecewise direct standardization,” *Eur. J. Soil Sci.*, vol. 66, no. 4, pp. 670–678, 2015, doi: 10.1111/ejss.12271.

[210] B. Zou, X. Jiang, H. Feng, Y. Tu, and C. Tao, “Multisource spectral-integrated estimation of cadmium concentrations in soil using a direct standardization and Spiking algorithm,” *Sci. Total Environ.*, vol. 701, 2020, doi: 10.1016/j.scitotenv.2019.134890.

[211] Y. Luo and C. Tu, *Twenty years of research and development on soil pollution and remediation in China*. 2018.

[212] P. Jarecke, K. Yokoyama, and P. Barry, “On-orbit radiometric calibration the Hyperion instrument,” *Int. Geosci. Remote Sens. Symp.*, vol. 6, pp. 2825–2827, 2001, doi: 10.1109/igarss.2001.978176.

[213] B. Datt, T. R. McVicar, T. G. Van Niel, D. L. B. Jupp, and J. S. Pearlman, “Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6 PART I, pp. 1246–1259, 2003, doi: 10.1109/TGRS.2003.813206.

[214] H. Li, D. Zhang, Y. Zhang, and Y. Xu, “Research of image preprocessing methods for EO-1 Hyperion hyperspectral data in tidal flat area,” *Geoinformatics 2008 Jt. Conf. GIS Built Environ. Classif. Remote Sens. Images*, vol. 7147, p. 71471G, 2008, doi: 10.1117/12.813253.

[215] J. Moros *et al.*, “Use of reflectance infrared spectroscopy for monitoring the metal content of the estuarine sediments of the Nerbioi-Ibaizabal River (Metropolitan Bilbao, Bay of Biscay, Basque Country),” *Environ. Sci. Technol.*, vol. 43, no. 24, pp. 9314–9320, 2009, doi: 10.1021/es9005898.

[216] S. S. Gill and N. Tuteja, “Cadmium stress tolerance in crop plants: Probing the role of sulfur,” *Plant Signal. Behav.*, vol. 6, no. 2, pp. 215–222, 2011, doi: 10.4161/psb.6.2.14880.

[217] Y. Chen *et al.*, “Mapping of Cu and Pb contaminations in soil using combined geochemistry, topography, and remote sensing: A case study in the le’an river floodplain, China,” *Int. J. Environ. Res. Public Health*, vol. 9, no. 5, pp. 1874–1886, 2012, doi: 10.3390/ijerph9051874.

[218] F. A. Kruse *et al.*, “The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data,” *Remote Sens. Environ.*, vol. 44, no. 2–3, pp. 145–163, 1993, doi: 10.1016/0034-4257(93)90013-N.

[219] A. Otten, A. Alphenaar, C. Pijls, F. Spuij, and H. Wit, *In Situ Soil Remediation*, vol. 6, no. 12. Dordrecht: Springer Netherlands, 1997.

[220] Y. Hong *et al.*, “Diagnosis of cadmium contamination in urban and suburban soils using visible-to-near-infrared spectroscopy,” *Environ. Pollut.*, vol. 291, 2021, doi: 10.1016/j.envpol.2021.118128.

[221] C. Donlon *et al.*, “The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission,” *Remote Sens. Environ.*, vol. 120, pp. 37–57, 2012, doi: 10.1016/j.rse.2011.07.024.

[222] J. Verrelst *et al.*, “Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review,” *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 273–290, 2015, doi: 10.1016/j.isprsjprs.2015.05.005.

[223] H. van der Werff and F. van der Meer, “Sentinel-2 for mapping iron absorption feature parameters,” *Remote Sens.*, vol. 7, no. 10, pp. 12635–12653, 2015, doi: 10.3390/rs71012635.

[224] S. Michel, P. Gamet, and M. J. Lefevre-Fonollosa, “HYPXIM A hyperspectral satellite defined for science, security and defence users,” *Work. Hyperspectral Image Signal Process. Evol. Remote Sens.*, 2011, doi: 10.1109/WHISPERS.2011.6080864.

[225] R. Houborg, M. Anderson, F. Gao, M. Schull, and C. Cammalleri, “Monitoring water and carbon fluxes at fine spatial scales using HyspIRI-like measurements,” *Int. Geosci. Remote Sens. Symp.*, pp. 7302–7305, 2012, doi: 10.1109/IGARSS.2012.6351975.

[226] K. Richter, T. Hank, C. Atzberger, M. Locherer, and W. Mauser, “Regularization strategies for agricultural monitoring: The EnMAP vegetation analyzer (AVA),” *Int. Geosci. Remote Sens. Symp.*, pp. 6613–6616, 2012, doi: 10.1109/IGARSS.2012.6352083.

[227] S. Pignatti *et al.*, “Development of algorithms and products for supporting the Italian hyperspectral PRISMA mission: The SAP4PRISMA project,” *Int. Geosci. Remote Sens. Symp.*, pp. 127–130, 2012, doi: 10.1109/IGARSS.2012.6351620.

[228] S. Kraft, U. Del Bello, M. Bouvet, and M. Drusch, “FLEX: ESA’s Earth Explorer 8 candidate mission, Proceedings of 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany,” pp. 7125–7128, 2012.

[229] E. Ben Dor, A. Kafri, and G. Varacalli, “Shalom : spaceborne hyperspectral applicative land and ocean mission: a joint project of asi-isa an update for 2014,” *IGARSS -A Spec. Sess. hyperspectral Sens. orbit*, no. July, 2014.

[230] M. Drusch *et al.*, “Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services,” *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012, doi: 10.1016/j.rse.2011.11.026.

[231] F. D. Van der Meer, H. M. A. van der Werff, and F. J. A. van Ruitenbeek, “Potential of ESA’s Sentinel-2 for geological applications,” *Remote Sens. Environ.*, vol. 148, pp. 124–133, 2014, doi: 10.1016/j.rse.2014.03.022.

[232] A. Fernández-Manso, O. Fernández-Manso, and C. Quintano, “SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 50, pp. 170–175, 2016, doi: 10.1016/j.jag.2016.03.005.

[233] M. Belgiu and O. Csillik, “Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis,” *Remote Sens. Environ.*, vol. 204, pp. 509–523, 2018, doi: 10.1016/j.rse.2017.10.005.

[234] L. Guanter *et al.*, “The EnMAP spaceborne imaging spectroscopy mission for earth observation,” *Remote Sens.*, vol. 7, no. 7, pp. 8830–8857, 2015, doi: 10.3390/rs70708830.

[235] C. Rogass, K. Segl, C. Mielke, Y. Fuchs, and H. Kaufmann, “EnGeoMap—A geological mapping tool applied to the EnMAP mission,” *EARSeL eProceedings*, vol. 12, no. 1, pp. 12–17, 2013, [Online]. Available: https://gfzpublic.gfz-potsdam.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item\_247780.

[236] C. M. Lee *et al.*, “An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities,” *Remote Sens. Environ.*, vol. 167, pp. 6–19, 2015, doi: 10.1016/j.rse.2015.06.012.

[237] I. Leifer *et al.*, “State of the art satellite and airborne marine oil spill remote sensing: Application to the BP Deepwater Horizon oil spill,” *Remote Sens. Environ.*, vol. 124, pp. 185–209, 2012, doi: 10.1016/j.rse.2012.03.024.

[238] R. F. Kokaly *et al.*, “Spectroscopic remote sensing of the distribution and persistence of oil from the Deepwater Horizon spill in Barataria Bay marshes,” *Remote Sens. Environ.*, vol. 129, pp. 210–230, 2013, doi: 10.1016/j.rse.2012.10.028.

[239] A. Peltier, A. Finizola, G. A. Douillet, E. Brothelande, and E. Garaebiti, “Structure of an active volcano associated with a resurgent block inferred from thermal mapping: The Yasur-Yenkahe volcanic complex (Vanuatu),” *J. Volcanol. Geotherm. Res.*, vol. 243–244, pp. 59–68, 2012, doi: 10.1016/j.jvolgeores.2012.06.022.

[240] B. Ribeiro da Luz and J. K. Crowley, “Spectral reflectance and emissivity features of broad leaf plants: Prospects for remote sensing in the thermal infrared (8.0-14.0 μm),” *Remote Sens. Environ.*, vol. 109, no. 4, pp. 393–405, 2007, doi: 10.1016/j.rse.2007.01.008.

[241] S. Sandor-Leahy and J. Shepanski, “Hyperspectral Remote Sensing: Data Collection and Exploitation,” *Encycl. Anal. Chem.*, 2006, doi: 10.1002/9780470027318.a2309.

[242] P. K. Garg, “Effect of contamination and adjacency factors on snow using spectroradiometer and hyperspectral images,” *Hyperspectral Remote Sens. Theory Appl.*, pp. 167–196, 2020, doi: 10.1016/B978-0-08-102894-0.00016-4.

[243] D. Zhang and G. Zhou, “Estimation of soil moisture from optical and thermal remote sensing: A review,” *Sensors (Switzerland)*, vol. 16, no. 8, 2016, doi: 10.3390/s16081308.

[244] J. A. M. Demattê, V. M. Sayão, R. Rizzo, and C. T. Fongaro, “Soil class and attribute dynamics and their relationship with natural vegetation based on satellite remote sensing,” *Geoderma*, vol. 302, pp. 39–51, 2017, doi: 10.1016/j.geoderma.2017.04.019.

[245] M. Shabou *et al.*, “Soil clay content mapping using a time series of Landsat TM data in semi-arid lands,” *Remote Sens.*, vol. 7, no. 5, pp. 6059–6078, 2015, doi: 10.3390/rs70506059.

[246] E. Garnier *et al.*, “Assessing the effects of land-use change on plant traits, communities and ecosystem functioning in grasslands: A standardized methodology and lessons from an application to 11 European sites,” *Ann. Bot.*, vol. 99, no. 5, pp. 967–985, 2007, doi: 10.1093/aob/mcl215.

[247] W. Q. Li, X. Z. Liu, M. A. Khan, and B. Gul, “Relationship between soil characteristics and halophytic vegetation in coastal region of north china,” *Pakistan J. Bot.*, vol. 40, no. 3, pp. 1081–1090, 2008.

[248] J. Solon, E. Roo-Zielińska, and M. Degórski, “Landscape scale of topography-soil-vegetation relationship: Influence of land use and land form,” *Polish J. Ecol.*, vol. 60, no. 1, pp. 3–17, 2012.

[249] B. A. Latif, R. Lecerf, G. Mercier, and L. Hubert-Moy, “Preprocessing of low-resolution time series contaminated by clouds and shadows,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2083–2096, 2008, doi: 10.1109/TGRS.2008.916473.

[250] J. Verrelst, J. P. Rivera, L. Alonso, L. Guanter, and J. Moreno, “Evaluating machine learning regression algorithms for operational retrieval of biophysical parameters: Opportunities for Sentinel,” *Eur. Sp. Agency, (Special Publ. ESA SP*, vol. 707 SP, 2012.

[251] F. Chen, Z. Zhao, L. Peng, and D. Yan, “Clouds and cloud shadows removal from high-resolution remote sensing images,” *Int. Geosci. Remote Sens. Symp.*, vol. 6, pp. 4256–4259, 2005, doi: 10.1109/IGARSS.2005.1525858.

[252] F. Melgani, “Contextual reconstruction of cloud-contaminated multitemporal multispectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 442–455, 2006, doi: 10.1109/TGRS.2005.861929.

[253] B. Somers, G. P. Asner, L. Tits, and P. Coppin, “Endmember variability in Spectral Mixture Analysis: A review,” *Remote Sens. Environ.*, vol. 115, no. 7, pp. 1603–1616, 2011, doi: 10.1016/j.rse.2011.03.003.

[254] R. Heylen, M. Parente, and P. Gader, “A review of nonlinear hyperspectral unmixing methods,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, 2014, doi: 10.1109/JSTARS.2014.2320576.

[255] A. Gholizadeh, N. Carmon, A. Klement, E. Ben-Dor, and L. Borůvka, “Agricultural soil spectral response and properties assessment: Effects of measurement protocol and data mining technique,” *Remote Sens.*, vol. 9, no. 10, 2017, doi: 10.3390/rs9101078.

[256] E. Ben Dor, C. Ong, and I. C. Lau, “Reflectance measurements of soils in the laboratory: Standards and protocols,” *Geoderma*, vol. 245–246, pp. 112–124, 2015, doi: 10.1016/j.geoderma.2015.01.002.

[257] R. Casa, F. Castaldi, S. Pascucci, A. Palombo, and S. Pignatti, “A comparison of sensor resolution and calibration strategies for soil texture estimation from hyperspectral remote sensing,” *Geoderma*, vol. 197–198, pp. 17–26, 2013, doi: 10.1016/j.geoderma.2012.12.016.

[258] L. B. Liao, P. J. Jarecke, D. A. Gleichauf, and T. R. Hedman, “Performance characterization of the Hyperion Imaging Spectrometer instrument,” *Earth Obs. Syst. V*, vol. 4135, p. 264, 2000, doi: 10.1117/12.494253.