

Article

Multi-Compartment Pollution Assessment of Soil Metals and Air Pollutants in Lebanon's Bekaa Valley: Integrating Geochemistry, Remote Sensing, and Multi-Scale Spatial Analysis

Marie Therese Abi Saab ¹, Elie Saliba ², Salim Fahed ¹, Rhend Sleiman ¹, Dany Romanos ¹, Yara Khairallah ¹, Claudine Sebaaly ³, Rossella Albrizio ^{4,*} and Mohamed Houssemeddine Sellami ⁴

¹ Climate and Water Unit, Lebanese Agricultural Research Institute (LARI), Jdeidet El-Metn P.O. Box 90-1965, Lebanon; mtabisaab@lari.gov.lb (M.T.A.S.); salimfahed@gmail.com (S.F.); rhend.t.sleiman@gmail.com (R.S.); danny_romanos@hotmail.com (D.R.); yara.khairallah@hotmail.com (Y.K.)

² The School of Engineering, Department of Agriculture, The Holy Spirit University of Kaslik (USEK), P.O. Box 446, Jounieh 1200, Lebanon; elie997saliba@hotmail.com

³ Faculty of Agronomy, Department of Plant Protection, Lebanese University, Dekwenh P.O. Box 90775, Lebanon; claudinesebaaly@gmail.com

⁴ Institute for Mediterranean Agricultural and Forest Systems (ISAFOM), National Research Council of Italy (CNR), P.le Enrico Fermi 1, 80055 Portici, Italy; mohamed.sellami@isafom.cnr.it

* Correspondence: rossella.albrizio@cnr.it

Abstract

This study presents the first integrated multi-compartment assessment linking soil heavy-metal contamination and satellite-derived tropospheric air pollution in Lebanon's Bekaa Valley. Six metals (Cd, Cr, Cu, Ni, Pb, Zn) were measured in 295 agricultural topsoils using DTPA extraction—which targets the labile, bioavailable pool most relevant to food-chain transfer in calcareous soils—and were compared with Sentinel-5P TROPOMI vertical columns of SO₂, NO₂ and HCHO extracted at point, 1000 m and 5000 m scales. Chromium was the dominant element, and composite pollution indices indicated overall good but transitional soil quality, with a majority of sites at warning level. Principal component analysis, spatial autocorrelation and land-use analysis converged on three plausible source groupings—predominantly geogenic (Cr, Ni), localized anthropogenic (Cd, Pb) and agricultural (Cu)—although, in the absence of isotopic or other direct source tracing, these assignments remain indicative rather than confirmed. Soil metals and atmospheric columns were largely uncorrelated ($|r| < 0.3$ for 17 of 18 pairs); this decoupling is consistent with the contrasting temporal integration and spatial support of the two compartments rather than with a single shared pathway. Multi-scale analysis supported a parsimonious dual-scale (point + 5000 m) sampling framework. The study offers a transferable assessment framework for other Mediterranean regions and underscores the need for locally calibrated, total-metal soil background values. Because these indices and threshold comparisons are computed from DTPA-extractable concentrations, they serve here as one-directional, bioavailability-based screening tools rather than regulatory determinations: an exceedance conservatively flags a site for confirmatory total-metal analysis, whereas a non-exceedance does not establish compliance.

Keywords: heavy metal contamination; Nemerow Pollution Index; source apportionment; TROPOMI; local spatial autocorrelation; spatial co-occurrence; Mediterranean agricultural regions



Academic Editors: Marcos Francos, Luís André Mendes, Ana Belén Muñoz and Fernando Peregrina

Received: 29 June 2026

Revised: 21 July 2026

Accepted: 1 August 2026

Published: 5 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

Heavy metal contamination of agricultural soils has become a major environmental challenge worldwide due to rapid urbanization, industrial development, and agricultural intensification. Heavy metals are capable of accumulating in soils for long periods, thereby posing risks to crop productivity, food safety, ecosystem functioning, and human health [1–3]. Their potential transfer through food chains and subsequent bioaccumulation further amplifies these risks. Recent global assessments indicate that soil metal pollution is widespread and increasingly threatens agricultural systems, particularly in semi-arid and Mediterranean regions where intensive land use, water scarcity, and geological conditions may favor metal accumulation [4]. Among the most frequently reported contaminants in these environments are chromium (Cr), nickel (Ni), and copper (Cu), which often originate from both natural and anthropogenic sources.

A major challenge in soil contamination research is the identification and discrimination of pollution sources. Conventional soil surveys provide direct and reliable measurements of contamination but are inherently constrained by sampling density, cost, and logistical limitations. As a result, contamination assessments often lack the spatial continuity required to fully characterize regional-scale pollution patterns. In parallel, advances in Earth observation technologies have transformed environmental monitoring by providing spatially continuous information on atmospheric pollutants. In particular, the Copernicus Sentinel-5P mission, equipped with the TROPOspheric Monitoring Instrument (TROPOMI) (The instrument was developed through a consortium involving Airbus Defence and Space Netherlands and the Netherlands Organisation for Applied Scientific Research (TNO), with major instrument development activities based in Leiden, The Netherlands), delivers near-daily observations of tropospheric pollutants such as sulfur dioxide (SO₂), nitrogen dioxide (NO₂), and formaldehyde (HCHO), which are widely recognized indicators of industrial activity, combustion processes, traffic emissions, and other anthropogenic sources [5,6]. The rationale for examining relationships between atmospheric pollutants and soil heavy metals is based on their potential origin from common anthropogenic emission sources. Industrial activities, transportation, fossil fuel combustion, waste burning, and intensive agricultural practices release both gaseous pollutants (e.g., SO₂, NO₂, and volatile organic compounds leading to HCHO formation) and metal-bearing particulates. While gaseous pollutants characterize current atmospheric pollution loads, heavy metals accumulate progressively in soils through dry and wet deposition and other environmental pathways. Therefore, the objective of this study is not to infer a direct causal relationship between atmospheric gases and soil metal concentrations, but rather to investigate whether atmospheric pollution patterns can reveal spatial signatures of common contamination sources.

Although soil contamination and atmospheric pollution are often driven by common emission sources, these environmental compartments are generally investigated separately. Consequently, the extent to which satellite-derived atmospheric pollutants can explain or predict spatial patterns of heavy metal accumulation in soils remains poorly understood. Existing studies have primarily focused either on mapping atmospheric pollution using remote sensing or on assessing heavy metal contamination through ground-based sampling [7–10]. Few studies have integrated both datasets within a unified spatial framework to evaluate whether atmospheric pollutant distributions can provide insights into soil contamination processes and source attribution. Addressing this gap is important because establishing such relationships could support cost-effective contamination screening and environmental monitoring in regions where extensive soil sampling programs are not feasible. To date, only a handful of studies have attempted to jointly analyse ground-based soil contamination and Sentinel-5P/TROPOMI atmospheric retrievals within a common spatial framework; to date these fields have advanced largely along separate tracks—remote

sensing of soil heavy metals on one side [11], and Sentinel-5P/TROPOMI mapping and validation of tropospheric SO₂, NO₂ and HCHO on the other [12,13]—so the extent to which the two compartments co-vary in space remains largely untested. Accordingly, the present study is framed around two explicit working hypotheses. H1: If soil heavy metals and tropospheric pollutants share common, spatially persistent anthropogenic emission sources, their spatial patterns should display detectable positive covariation. The corresponding null hypothesis (H0) is that the two compartments are spatially decoupled because of their contrasting temporal integration (decadal soil accumulation versus a single-season atmospheric snapshot) and spatial support (sub-metre point samples versus multi-kilometre satellite pixels). H2: The spatial structure of soil metals reflects a mixture of geogenic and localized anthropogenic controls that can be disentangled through multivariate and multi-scale spatial analysis.

The Bekaa Valley in Lebanon provides an ideal setting for investigating these interactions. As the country's principal agricultural region, accounting for approximately 42% of Lebanon's cultivated land [14], the valley is exposed to multiple potential contamination sources, including industrial emissions, dense transportation corridors, intensive agricultural practices, and uncontrolled waste disposal. These pressures occur within a topographically enclosed basin where atmospheric circulation is partially restricted, potentially enhancing pollutant accumulation and deposition. At the same time, the region lacks nationally established geochemical background values for heavy metals, forcing environmental assessments to rely largely on international reference standards that may not adequately reflect local geological and pedological conditions [15,16]. Previous investigations in Lebanon have examined specific aspects of environmental contamination, including soil-groundwater vulnerability, heavy metal accumulation in agricultural soils, fertilizer-derived trace element inputs, and sediment contamination [15,17–21]. However, no study has jointly analyzed soil heavy metal concentrations and satellite-derived atmospheric pollutants using multi-scale spatial statistics to investigate contamination patterns and potential common sources in the Bekaa Valley.

The present study addresses this knowledge gap through an integrated assessment of soil and atmospheric pollution. Specifically, it aims to: (i) characterize the contamination status and spatial distribution of six priority heavy metals (Cd, Cr, Cu, Ni, Pb, and Zn) using the Nemerow Pollution Index and enrichment factor analysis; (ii) quantify tropospheric SO₂, NO₂, and HCHO concentrations derived from Sentinel-5P observations at point, 1000 m, and 5000 m spatial scales; (iii) evaluate statistical relationships between soil metals and atmospheric pollutants to assess the hypothesis of shared anthropogenic emission sources; (iv) investigate spatial clustering and heterogeneity using Global (Moran's I) and local (LISA, Getis-Ord Gi*) indicators of spatial autocorrelation together with coefficient of variation analyses; and (v) infer potential dominant contamination sources through principal component analysis (PCA).

Beyond its regional relevance, this study addresses a broader methodological challenge in environmental monitoring: determining whether satellite-derived atmospheric pollution indicators can provide meaningful information about soil contamination patterns. By integrating field-based geochemical measurements, Sentinel-5P observations, multi-scale spatial analysis, and multivariate source apportionment techniques, the study offers a transferable framework for contamination assessment in data-scarce agricultural regions. The proposed approach is particularly relevant for Mediterranean and semi-arid environments worldwide, where monitoring resources are limited, contamination pressures are increasing, and cost-effective methods for identifying pollution hotspots and emission sources are urgently needed.

2. Materials and Methods

2.1. Study Area

The Bekaa Valley is an intermontane graben situated in Eastern Lebanon (33.6–33.9° N, 35.8–36.0° E; 850–1000 m a.s.l.), bounded by Mount Lebanon to the west and the Anti-Lebanon range to the east. The region experiences a semi-arid Mediterranean climate (Köppen BSk–Csa), with mean annual precipitation ranging from 230 mm in the north to 610 mm in the central valley. Dominant soils are Fluvisols, Cambisols, Regosols, Vertisols, and Luvisols, developed on calcareous parent material [15]. The valley encompasses Lebanon's most productive agricultural land (42% of the national cultivated area), with principal crops including cereals, potatoes, vegetables, and fruit trees/vineyards [14,22]. The study area spans 52 villages encompassing industrial, agricultural, and residential land uses, with the Litani River serving as the primary irrigation source despite documented contamination from untreated municipal and industrial wastewater [23].

2.2. Soil Sampling and Laboratory Analysis

A total of 295 georeferenced sampling locations were distributed across the Bekaa Valley using a stratified random design. Sampling was conducted in September 2019, coinciding with the end of the irrigation season and the peak dry period to ensure stable metal concentrations (Figure 1). At each site, composite soil samples were collected from the 0–25 cm layer using a stainless-steel auger, homogenized by quartile mixing (~1 kg), and stored in sealed polyethylene bags. This depth was selected because it corresponds to the cultivated topsoil layer, where agricultural practices such as irrigation, fertilization, manure application, and pesticide use predominantly influence heavy-metal accumulation. In addition, the surface soil is the primary recipient of atmospheric deposition from industrial emissions, traffic-related pollutants, and windborne dust, making it the most relevant layer for assessing the combined effects of agricultural and airborne sources of contamination. The 0–25 cm layer also encompasses the main root zone of most annual crops cultivated in the Bekaa Valley and is therefore particularly suitable for evaluating metal bioavailability and potential transfer to plants. In the laboratory (Lebanese Agricultural Research Institute), samples were air-dried, disaggregated, and sieved through a 2 mm stainless-steel mesh. Heavy metals (Cd, Cr, Cu, Ni, Pb, and Zn) were extracted using the DTPA (diethylenetriaminepentaacetic acid) buffered solution method following ISO 14870:2001 [24,25] and quantified by flame Atomic Absorption Spectrophotometry (AAS). DTPA is recommended for assessing the reactive metal pool in calcareous soils [26]. This method was adopted for our conditions, considering that our local soils are calcareous and that we also need to know the metals' availability to get transported into the food chain. Quality assurance included analysis of procedural blanks (every 20 samples), certified reference materials, and duplicate measurements (relative standard deviation < 10% for all metals). Detection limits (mg/kg) were Cd = 0.001, Cr = 0.5, Cu = 0.1, Ni = 0.5, Pb = 0.1, and Zn = 0.1. It should be emphasised that DTPA extraction targets the labile, plant-available metal pool rather than total (pseudo-total) metal content; consequently, the concentrations reported here are not directly equivalent to the total-metal basis on which most regulatory guidelines are defined. All subsequent guideline comparisons are therefore treated as conservative, bioavailability-based screening indicators rather than formal total-metal exceedances, and confirmatory total digestion (e.g., aqua regia with ICP-MS) is recommended for definitive assessment (see Sections 2.3 and 4). Because DTPA recovers only a fraction of the total-metal pool, this screening is inherently one-directional: a DTPA concentration exceeding a total-metal guideline is a conservative indication of probable total-metal exceedance, whereas a DTPA concentration below that guideline cannot be taken to demonstrate compliance, because the larger total concentration may still exceed the threshold.

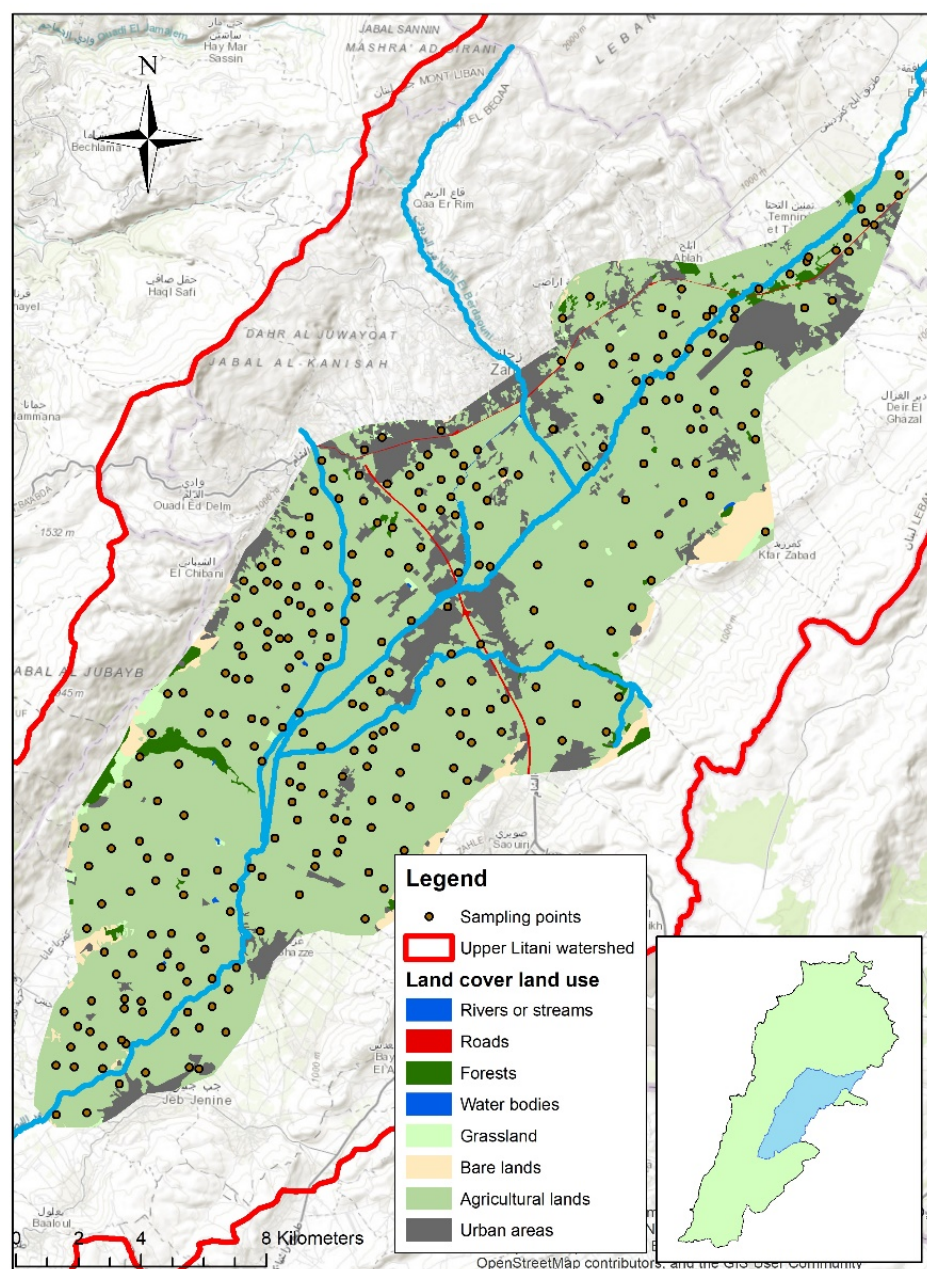


Figure 1. Study area and locations of the sampled soil.

2.3. Pollution Assessment Indices

The Nemerow Pollution Index (NPI) was calculated to evaluate overall multi-element soil contamination. For each metal “*i*”, the single-element pollution index (PI) was computed as $i = C_i/S_i$, where C_i is the measured concentration and S_i the guideline-based screening reference value. Since Lebanon lacks nationally established soil quality guidelines, reference standards were primarily derived from the Canadian Council of Ministers of the Environment (CCME) Soil Quality Guidelines for the Protection of Environmental and Human Health [16], supplemented by internationally adopted reference values commonly used in soil pollution literature [27,28]. The reference standards adopted for NPI calculation were: Cd = 3, Cr = 100, Cu = 100, Ni = 50, Pb = 100, and Zn = 300 mg/kg, representing widely used international benchmarks. For the exceedance analysis, the more conservative CCME agricultural soil quality guidelines were additionally applied: Cr = 64, Cu = 63, and Ni = 50 mg/kg [16], enabling sensitivity assessment of contamination classification to regulatory threshold selection. The composite NPI was then derived

as $NPI = \sqrt{(P_{\max}^2 + P_{\text{mean}}^2)}/2$, balancing extreme and average contamination [27,28]. Sites were classified as: Clean ($NPI \leq 0.7$), Warning ($0.7 < NPI \leq 1.0$), Slightly Polluted ($1.0 < NPI \leq 2.0$), Moderately Polluted ($2.0 < NPI \leq 3.0$), and Heavily Polluted ($NPI > 3.0$). Enrichment Factors (EF) were computed as $EF = C_{\text{measured}}/C_{\text{background}}$, using geogenic background concentrations: Cd = 0.3, Cr = 60, Cu = 25, Ni = 40, Pb = 20, and Zn = 70 mg/kg, and classified following Sutherland (2000) [29]: Minimal ($EF < 2$), Moderate (2–5), Significant (5–20), and Very High (>20). The CCME guidelines were adopted because they are internationally recognised, explicitly derived for the protection of both environmental and human health, and provide dedicated agricultural-land values; in the absence of nationally established Lebanese thresholds, they offer a transparent and reproducible benchmark. We acknowledge, however, that they were not calibrated to Lebanese lithology and are defined on a total-metal basis, whereas our measurements represent the DTPA-extractable fraction; exceedance percentages are consequently reported as screening indicators. Priority sites (“hotspots”) were defined statistically, in relative terms, as the upper decile (top 10%) and upper ventile (top 5%) of the composite NPI distribution, thereby identifying the most contaminated sites within the surveyed area rather than sites exceeding an absolute regulatory limit. Throughout this study, the NPI, the single-element pollution index, and the CCME threshold comparisons are therefore used strictly as internally consistent, bioavailability-based screening devices for ranking and prioritising sites; they are not, and should not be read as, regulatory (total-metal) assessments of soil quality or determinations of compliance. Accordingly, the pollution-class labels (e.g., “Clean”, “Warning”) describe a site’s relative standing within this DTPA-based screening scheme and do not certify that its total-metal concentrations lie below any guideline value.

2.4. Satellite-Derived Air Pollutants and Land Cover

Tropospheric vertical column densities (VCD) of SO₂, NO₂, and HCHO were extracted from Sentinel-5P/TROPOMI Level 2 products using Google Earth Engine [30,31]. TROPOMI estimates VCD using the Differential Optical Absorption Spectroscopy (DOAS) retrieval at 0.01 arc-degree spatial resolution. Sentinel-5P/TROPOMI products provide tropospheric vertical column densities of SO₂, NO₂, and HCHO rather than concentrations measured at a specific altitude above ground level. The retrieved values represent the integrated abundance of these pollutants throughout the tropospheric atmospheric column over each pixel. To characterize atmospheric pollution conditions during the period corresponding to soil sampling, mean pollutant values were calculated for the June–October 2019 period. Air pollutant concentrations were extracted at three spatial scales around each sampling point: (i) point level (exact coordinates), (ii) 1000 m circular buffer (neighborhood scale), and (iii) 5000 m circular buffer (landscape scale). This multi-scale extraction strategy enables discrimination between localized emission signatures and regional background concentrations. Land use/land cover (LULC) was classified from Sentinel-2/ESA CCI Land Cover products into seven categories: Agricultural land, Artificial land, Forest land, Grassland, Natural land without vegetation, Water body, and River. Proportional coverage (%) was computed within 1000 m and 5000 m buffers. The 1000 m and 5000 m radii were chosen to represent the neighbourhood (local-emission) and landscape (regional-background) scales, respectively, and to bracket the effective spatial support of the TROPOMI retrieval: the 5000 m radius approximates the $\sim 3.5 \times 7$ km pixel footprint, ensuring that landscape-scale values integrate at least one quasi-independent satellite observation, whereas the 1000 m radius resolves sub-pixel land-use heterogeneity immediately around each sampling point. The 1000 m radius is intended to capture the near-source zone in which local dry deposition and short-range dispersion gradients dominate, whereas the 5000 m radius represents the landscape scale over which regional

advective mixing and background concentrations prevail. These radii are consistent with the buffer sizes (typically a few hundred metres to several kilometres) widely adopted in land-use-regression and buffer-based environmental exposure studies [32], and with the effective spatial support of the TROPOMI retrieval discussed above. It is emphasised that this multi-scale spatial pairing is designed to test for the spatial co-occurrence of common emission signatures, and not to infer any causal or temporal link between present-day atmospheric columns and soil metals that have accumulated over years to decades.

2.5. Statistical Analysis

Each variable was first summarised using descriptive statistics, namely the mean, standard deviation, coefficient of variation (CV), skewness, and kurtosis. To measure how variables varied together, Pearson correlation coefficients were then computed for three sets of relationships: (i) among the six soil metals, (ii) between the metals and the three air pollutants (SO₂, NO₂, and HCHO), and (iii) between the land-use/land-cover (LULC) classes and pollutant concentrations at each spatial scale.

To identify shared contamination sources among the metals, Principal Component Analysis (PCA) with Varimax rotation was applied to the z-standardised metal concentrations. The number of components to retain was decided using three complementary criteria rather than a single rule: the Kaiser criterion (eigenvalues ≥ 1.0), visual inspection of the scree plot (Supplementary Figure S1), and the interpretability of the resulting components. The dataset is well suited to PCA; with 295 sites and only six metals, there are about 49 observations per variable, far above the 5–10 per variable usually recommended for stable, reproducible results.

Spatial patterns were examined at two levels. Global spatial autocorrelation was assessed with Moran's I statistic, using k-nearest-neighbour ($k = 8$) row-standardised spatial weights. To locate the specific clusters responsible for any global signal, two local indicators of spatial association were then computed from the same $k = 8$ weights: the Local Moran's I (LISA) statistic [33], which classifies each site as a High–High, Low–Low, High–Low, or Low–High cluster, and the Getis–Ord Gi* statistic [34], which identifies statistically significant hot- and cold-spots. LISA clusters were considered significant at a two-sided $p < 0.05$ (based on the analytical randomisation standardised statistic), and Gi* hot- and cold-spots were identified at $|Z| > 1.96$ (95%) and $|Z| > 2.58$ (99%).

Group differences were tested with two methods. One-way ANOVA compared air-pollutant concentrations across the point, 1000 m, and 5000 m buffer zones, with eta-squared (η^2) reported as the effect size. Two-sample Welch *t*-tests compared LULC coverage between At-Risk ($\text{NPI} > 0.7$) and Clean ($\text{NPI} \leq 0.7$) sites, with Cohen's *d* expressing the practical size of each difference. Finally, redundancy between spatial scales was checked by correlating the same variable measured at different buffer sizes.

Statistical significance was set at $\alpha = 0.05$ throughout. With 295 sampling sites, the study is well powered: a two-tailed Pearson correlation test at $\alpha = 0.05$ has an 80% probability of detecting a correlation as weak as $|r| \approx 0.16$, so even subtle associations between soil metals and air pollutants could be detected. All analyses were carried out in R version 4.5.1 [35]. Data handling and visualisation used the tidyverse [36]; spatial operations and buffer calculations used sf [37]; the spatial-autocorrelation tests (global Moran's I, LISA, and Getis–Ord Gi*) used spdep [38]; and principal component analysis was performed with FactoMineR [39] and visualised with factoextra [40].

3. Results

3.1. Heavy Metal Concentrations and Contamination Status

Table 1 presents descriptive statistics for the six heavy metals across 295 sampling sites. Chromium exhibited the highest mean concentration (86.06 ± 31.84 mg/kg; median: 95.19), followed by Zn (57.56 ± 18.77), Ni (30.32 ± 14.46), and Cu (7.62 ± 13.70 mg/kg). Cadmium and Pb had the lowest means at 0.012 ± 0.031 and 0.714 ± 1.699 mg/kg, respectively. Coefficients of variation revealed two distinct variability regimes: Cd (266.5%), Pb (238.0%), and Cu (179.8%) exhibited extreme heterogeneity ($CV > 100\%$), diagnostic of localized point-source contamination, whereas Cr (37.0%), Zn (32.6%), and Ni (47.7%) displayed moderate variability consistent with mixed diffuse and geogenic contributions. Strong positive skewness for Cd (2.31), Cu (4.74), and Pb (2.05) confirmed right-skewed distributions with occasional high-concentration outliers representing contamination hotspots.

Table 1. Descriptive statistics of heavy metal concentrations (mg/kg) in Bekaa Valley agricultural soils ($n = 295$). CV = Coefficient of Variation.

Metal	<i>n</i>	Mean	SD	Median	Min	Max	CV (%)	Skew
Cd	295	0.012	0.031	0.000	0.000	0.111	266.5	2.31
Cr	295	86.060	31.835	95.187	11.503	189.960	37.0	−0.70
Cu	295	7.619	13.695	3.047	0.000	101.950	179.8	4.74
Ni	295	30.317	14.458	29.721	5.611	90.682	47.7	0.66
Pb	295	0.714	1.699	0.012	0.000	6.111	238.0	2.05
Zn	295	57.564	18.770	57.693	9.015	119.082	32.6	0.21

Screening-level threshold comparison identified Cr as the primary contaminant of concern: approximately 79% of sites (233 of 295) exceeded the CCME soil quality guideline of 64 mg/kg [16], while 113 sites (38.3%) exceeded the more permissive internationally adopted reference value of 100 mg/kg used for NPI computation. This large discrepancy underscores the sensitivity of contamination classification to regulatory threshold selection. Nickel exceeded its CCME threshold of 50 mg/kg at 24 sites (8.1%), and Cu surpassed the CCME guideline of 63 mg/kg at approximately 2% of sites (though only a single site exceeded the 100 mg/kg reference used for NPI). No exceedances were recorded for Cd, Pb, or Zn under any benchmark. Enrichment factor analysis (Figure 2) confirmed predominantly minimal enrichment ($EF < 2$) across all metals (100% for Cd, Pb, and Zn; 99.3% for Ni; 98.0% for Cu; 91.2% for Cr), with Cr alone showing moderate enrichment (EF 2–5) at 8.8% of sites (maximum $EF = 3.17$), suggesting localized anthropogenic augmentation above geogenic baselines. The spatial distribution of individual metals (Figure 3) reveals distinct geographic patterns for each element. Because these comparisons are based on DTPA-extractable rather than total concentrations, the exceedance percentages should be read as conservative, bioavailability-oriented screening indicators. The apparent contrast between the generally good composite NPI status and the high single-element Cr exceedance is reconciled by noting that the NPI integrates all six metals against the 100 mg/kg international reference, whereas the 79% figure reflects a single element evaluated against the more stringent CCME agricultural threshold (64 mg/kg); much of the Cr signal is, moreover, attributable to geogenic enrichment from calcareous/ophiolitic parent materials (Section 4.1). More generally, because DTPA recovers only part of the total-metal pool, the exceedance counts reported here represent a conservative lower bound on the true total-metal exceedance, and—crucially—the sites that do not exceed a threshold cannot

be regarded as compliant; confirmatory total digestion would be required to establish either outcome.

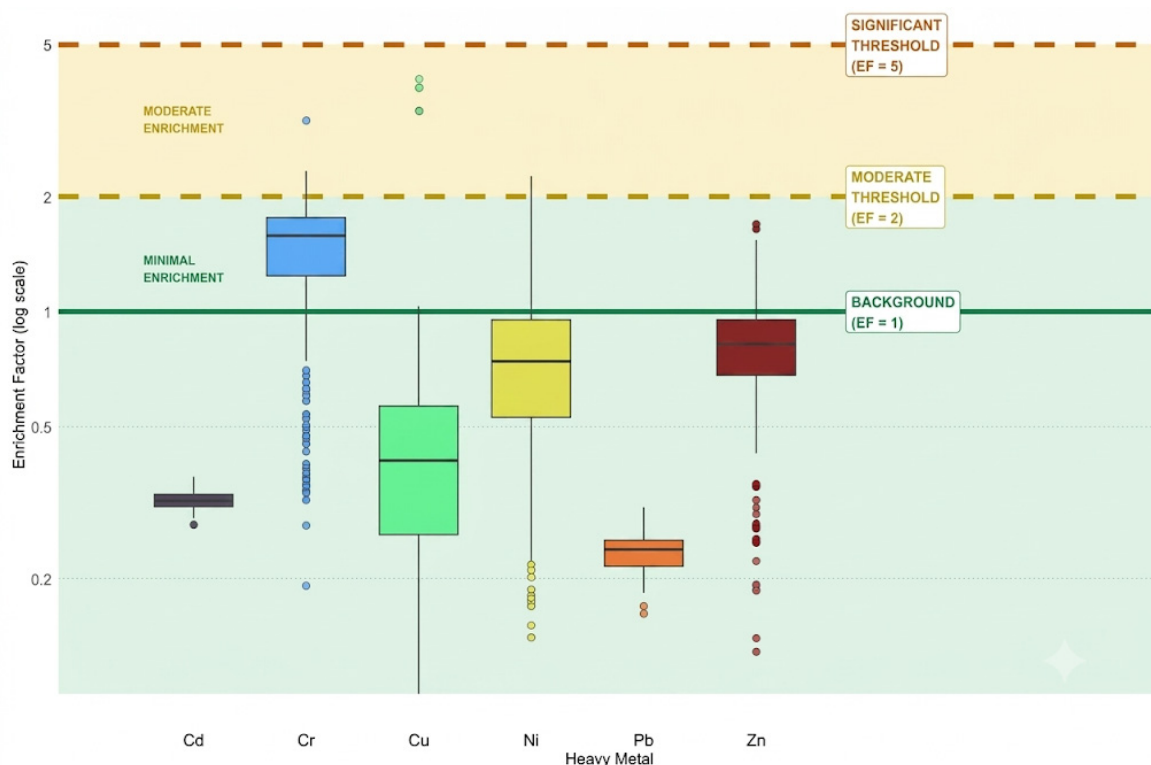


Figure 2. Enrichment Factor (EF) distribution by heavy metal with classification zones: green = Minimal ($EF \leq 2$), yellow = Moderate ($2 < EF \leq 5$), orange = Significant ($5 < EF \leq 20$), red = Very High ($EF > 20$). $EF = \text{Measured concentration} / \text{Natural background concentration}$. Solid green line = natural background ($EF = 1$). Classification after Sutherland (2000) [29].

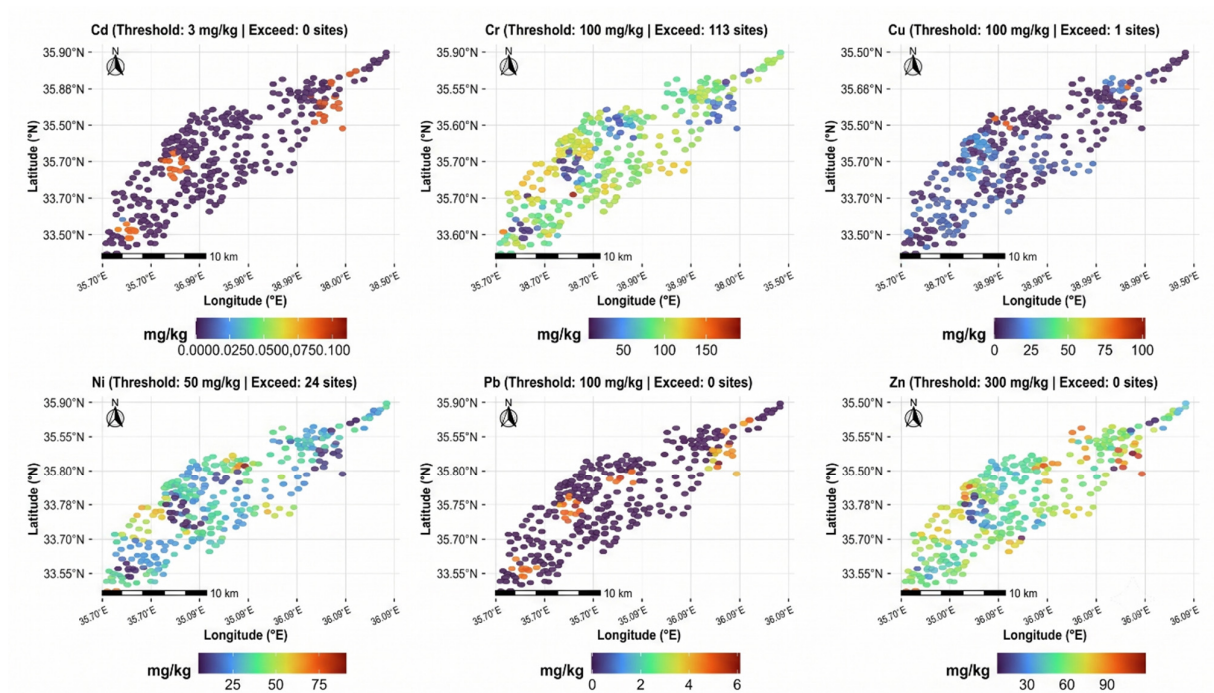


Figure 3. Spatial distribution of individual heavy metals across the study area. Color intensity represents concentration magnitude (mg/kg). Color scale: low (blue) to high (red) concentrations. Subplot titles indicate regulatory thresholds and the number of exceedance sites.

3.2. Nemerow Pollution Index and Hotspot Identification

The composite NPI classified the vast majority of sites below the pollution threshold: 133 (45.1%) as Clean ($\text{NPI} \leq 0.7$), 156 (52.9%) as Warning ($0.7 < \text{NPI} \leq 1.0$), and only 6 (2.0%) as Slightly Polluted ($1.0 < \text{NPI} \leq 2.0$). No sites reached moderate or heavy pollution levels. The mean NPI was 0.667 ± 0.232 (range: 0.133–1.383). Spatial mapping (Figure 4) reveals a north–south gradient in contamination severity. The 30 highest-priority sites (top 10%; $\text{NPI} \geq 0.90$) and 15 critical hotspots (top 5%; $\text{NPI} \geq 0.97$) were concentrated in the central and southern portions of the study area (Figure 5), spatially coinciding with zones of mixed industrial–agricultural land use and proximity to the Litani River. Here, hotspots are defined in relative, distribution-based terms as the upper decile (top 10%, $\text{NPI} \geq 0.90$) and upper ventile (top 5%, $\text{NPI} \geq 0.97$) of the composite NPI distribution; they identify the most contaminated sites within the surveyed area for prioritised screening and confirmatory analysis, not sites exceeding an absolute regulatory limit. As emphasised in Section 2.3, these NPI classes are relative, bioavailability-based screening categories derived from DTPA-extractable concentrations; in particular, a “Clean” or “Warning” classification reflects a low DTPA-based index value and does not certify compliance with total-metal soil-quality guidelines.

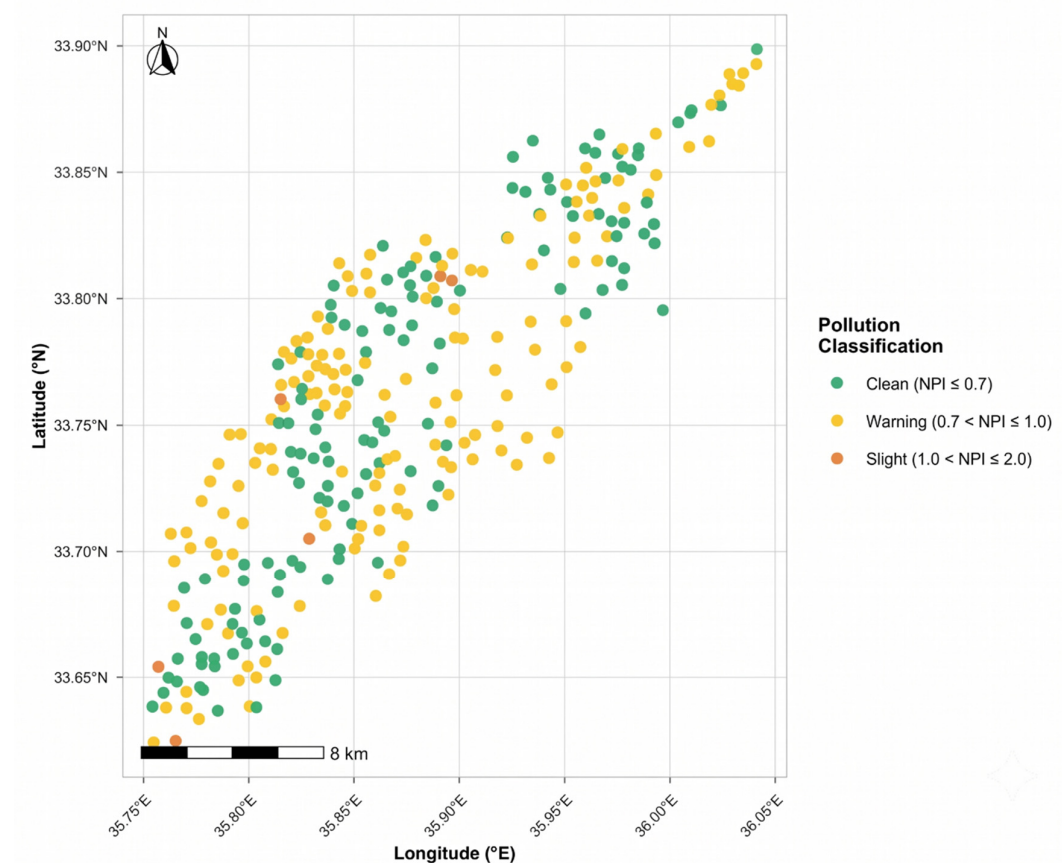


Figure 4. Pollution classification map based on NPI thresholds. Green = Clean ($\text{NPI} \leq 0.7$, $n = 133$), Yellow = Warning ($0.7 < \text{NPI} \leq 1.0$, $n = 156$), Orange = Slight ($1.0 < \text{NPI} \leq 2.0$, $n = 6$).

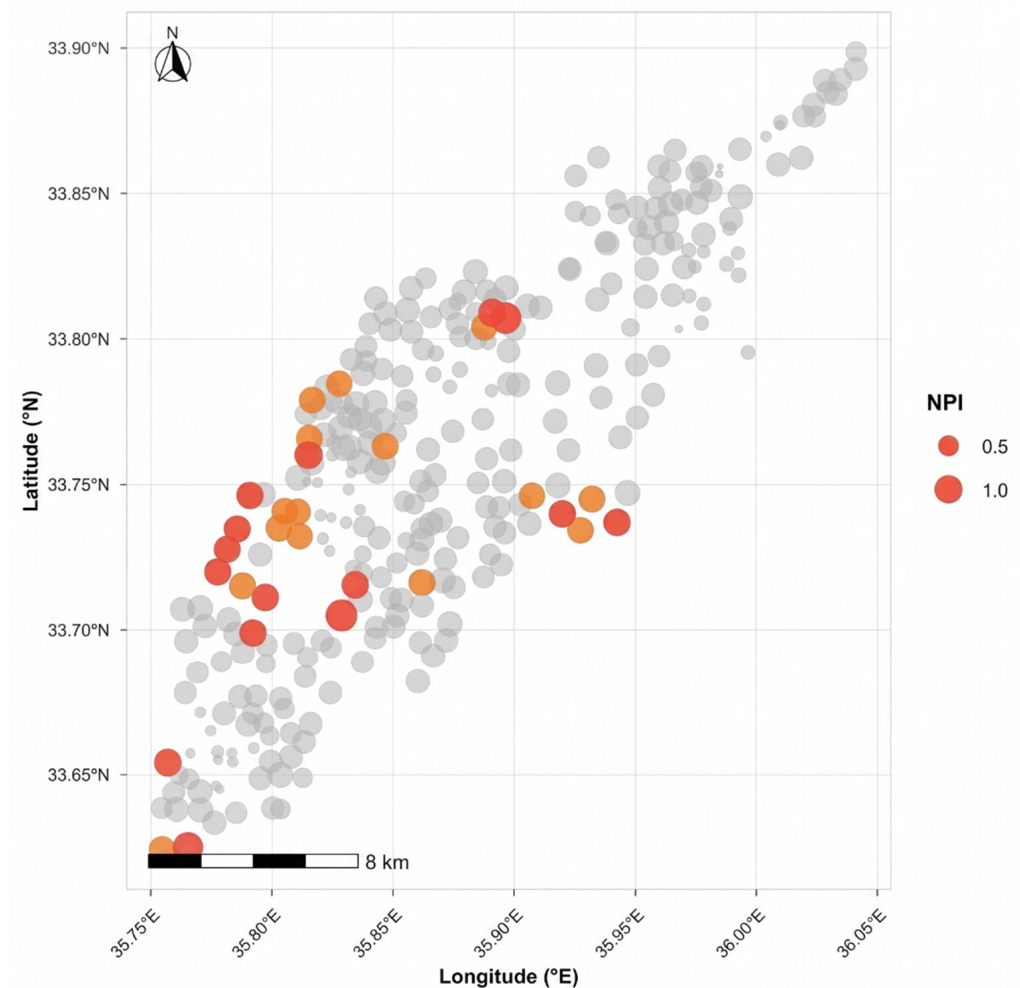


Figure 5. Pollution hotspot identification map. Orange = top 10% ($NPI \geq 0.90$, $n = 30$); Red = top 5% ($NPI \geq 0.97$, $n = 15$). Maximum $NPI = 1.383$.

3.3. Spatial Autocorrelation Analysis

Global Moran's I analysis (Table 2) demonstrated statistically significant positive spatial autocorrelation (clustering) for all ten variables tested ($p < 0.001$). Among heavy metals, Cd exhibited the strongest clustering ($I = 0.472$, $Z = 17.49$), followed by Pb (0.441), Ni (0.405), Cr (0.366), Zn (0.330), and Cu (0.169). Air pollutants showed markedly stronger spatial structure: NO_2 ($I = 0.960$), SO_2 (0.869), and HCHO (0.828), reflecting the smoother spatial gradients inherent to tropospheric satellite retrievals. The composite NPI exhibited substantial clustering ($I = 0.360$, $Z = 13.30$). It is important to stress that positive spatial autocorrelation indicates only that similar values cluster in space; such clustering is consistent with localized (point-source) inputs but can equally arise from spatially structured geogenic gradients, land-use patterns, or hydrological and depositional processes, and therefore cannot by itself confirm a point-source origin. When Moran's I is combined with the coefficient of variation, the elements displaying both strong clustering and very high CV (Cd, Pb) are the most consistent with localized anthropogenic inputs, whereas the moderately variable, spatially continuous elements (Cr, Ni) are more consistent with geogenic gradients; Cu, with the weakest clustering ($I = 0.169$), is consistent with a diffuse source. The specific sites driving this clustering—statistically significant hotspots, cold-spots and spatial outliers—are located using local indicators of spatial association in Section 3.7.

Table 2. Global spatial autocorrelation analysis (Moran's I). Spatial weights: k-nearest neighbors (k = 8), row-standardized. All Z-scores > 1.96; all *p* < 0.001.

Variable	Type	Moran's I	Z-Score	<i>p</i> -Value	Pattern	Indicative Source Pattern
Cd	Heavy Metal	0.472	17.49	<0.001	Clustered	Localized/anthropogenic
Cr	Heavy Metal	0.366	13.51	<0.001	Clustered	Geogenic gradient
Cu	Heavy Metal	0.169	6.60	<0.001	Clustered	Diffuse/agricultural
Ni	Heavy Metal	0.405	14.98	<0.001	Clustered	Geogenic gradient
Pb	Heavy Metal	0.441	16.34	<0.001	Clustered	Localized/anthropogenic
Zn	Heavy Metal	0.330	12.22	<0.001	Clustered	Geogenic gradient
NPI	Composite Index	0.360	13.30	<0.001	Clustered	Mixed
HCHO	Air Pollutant	0.828	30.51	<0.001	Clustered	Regional
NO ₂	Air Pollutant	0.960	35.19	<0.001	Clustered	Regional
SO ₂	Air Pollutant	0.869	31.94	<0.001	Clustered	Regional

3.4. Inter-Metal Correlations and Principal Component Analysis

Pearson correlation analysis revealed two clearly distinct metal association groups (Table 3). A very strong positive correlation linked Pb and Cd ($r = 0.878$, $p < 0.001$), while both exhibited strong negative correlations with Cr ($r = -0.757$ and -0.701) and moderate negative correlations with Ni ($r = -0.577$ and -0.538). Chromium and Ni were moderately positively correlated ($r = 0.575$), and both were associated positively with Zn ($r = 0.391$ and 0.528 , respectively). Copper was statistically independent of all other metals ($|r| < 0.08$ in all pairs), suggesting an isolated, distinct source.

Table 3. Pearson correlation matrix of heavy metal concentrations.

	Pb	Cu	Cd	Cr	Ni
Cu	−0.066	—			
Cd	0.878	−0.073	—		
Cr	−0.757	0.056	−0.701	—	
Ni	−0.577	−0.017	−0.538	0.575	—
Zn	−0.274	−0.078	−0.227	0.391	0.528

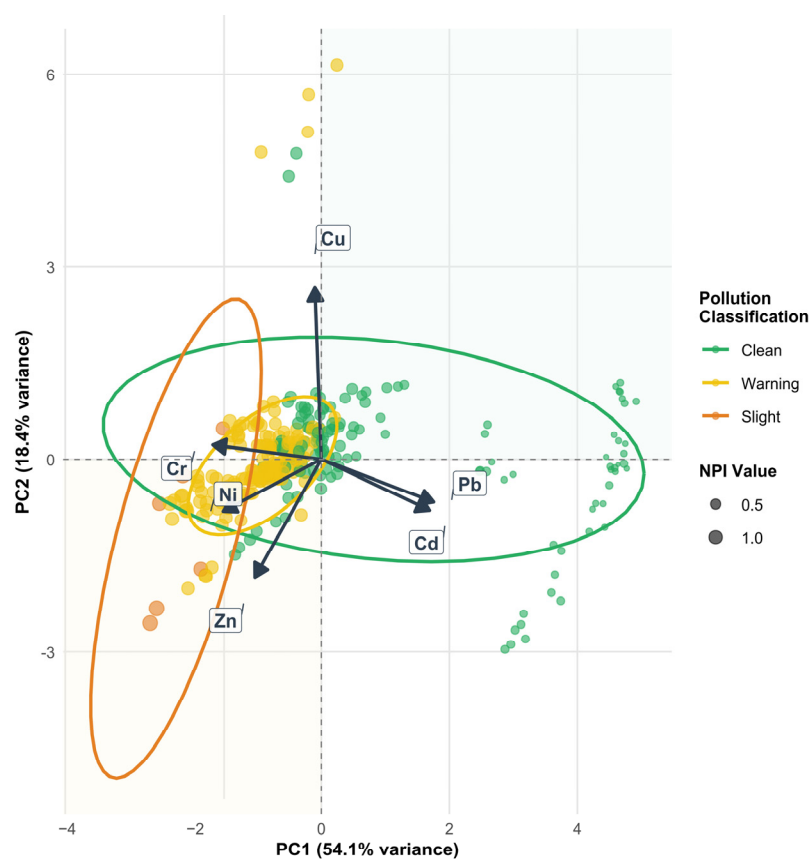
Values in bold are different from 0 with a significance level $\alpha = 0.05$.

PCA extracted two components with eigenvalues ≥ 1.0 , cumulatively explaining 72.5% of the total variance (Table 4), a solution supported by both the Kaiser criterion and the scree plot (Supplementary Figure S1), where the curve bends sharply (the “elbow”) after the second component, indicating that two components are enough to summarize the data. PC1 (54.1%) showed positive loadings on Pb (0.502) and Cd (0.483) and negative loadings on Cr (−0.484) and Ni (−0.438), representing an agricultural input axis. A third component (PC3, eigenvalue = 0.855, 14.3% variance) fell below the Kaiser threshold and is therefore not retained as an independent source dimension; it is reported in Table 4 for completeness only and is not used for source interpretation. The PCA biplot (Figure 6) visually confirms the source separation, with 95% confidence ellipses showing progressive displacement of Warning and Slight-pollution sites along the PC1 axis.

Table 4. PCA eigenvalues, variance explained, and component loadings. Loadings $\geq |0.4|$ are considered dominant.

	PC1	PC2	PC3
Eigenvalue	3.246	1.102	0.855
Variance (%)	54.10	18.37	14.25
Cumulative (%)	54.10	72.46	86.71
Cd	0.483	−0.226	−0.296
Cr	− 0.484	0.063	0.065
Cu	−0.029	0.762	−0.647
Ni	− 0.438	−0.230	−0.237
Pb	0.502	−0.190	−0.255
Zn	−0.296	− 0.525	−0.607

Boldface factor loadings indicate the most relevant characters for each principal component.

**Figure 6.** PCA biplot combining sample scores (points) and variable loadings (arrows, similar directions suggest common sources). Sites colored by NPI pollution classification with 95% confidence ellipses. PC1 separates anthropogenic (Cd, Pb) from geogenic (Cr, Ni) metals; PC2 isolates the agricultural Cu source.

3.5. Tropospheric Air Pollutants and Metal–Air Correlations

Mean tropospheric column densities were remarkably stable across buffer scales: $\text{HCHO} \approx 1.56 \times 10^{-4}$, $\text{NO}_2 \approx 1.39 \times 10^{-4}$, and $\text{SO}_2 \approx 1.22 \times 10^{-4}$ mol/m². SO_2 exhibited the greatest spatial heterogeneity (CV = 23.5–34.2%), possibly reflecting localized industrial emissions, followed by NO_2 (CV = 9.0–9.9%) and HCHO (CV = 4.2–6.3%). Importantly, CV decreased consistently with increasing buffer size for all three pollutants, confirming that spatial averaging at larger scales smooths localized emission signatures. One-way ANOVA

detected no significant concentration differences across buffer zones: HCHO ($F = 2.41$, $p = 0.09$), NO_2 ($F = 0.24$, $p = 0.79$), and SO_2 ($F = 1.75$, $p = 0.18$); all $\eta^2 < 0.01$. Inter-buffer redundancy was very high for Point vs. 1000 m (NO_2 : $r = 0.996$; SO_2 : $r = 0.964$; HCHO: $r = 0.971$).

Pearson correlation between heavy metals and air pollutants at the 1000 m buffer scale (Figure 7) revealed predominantly weak to negligible associations: $|r| < 0.3$ for 17 of 18 tested pairs. Only one pair, Cu– NO_2 ($r = 0.118$, $p = 0.043$), reached nominal significance; however, given the negligible shared variance ($r^2 \approx 1.4\%$) and the number of pairwise tests performed (which inflates the probability of a chance result), we do not interpret this coefficient as ecologically meaningful evidence of a shared traffic-related source. Cd– NO_2 ($r = -0.104$, $p = 0.075$) and Cr–HCHO ($r = -0.112$, $p = 0.055$) approached but did not attain statistical significance. The spatial distribution of air pollutants (Figure 8) shows that SO_2 hotspots are spatially concentrated, which may point to discrete industrial emission sources in those areas.

3.6. Multi-Scale Land Use–Pollution Associations

A total of 19 statistically significant correlations ($p < 0.05$) were identified between LULC variables and heavy metal concentrations across buffer zones (Figure 9). The strongest associations were: Water body at 1000 m with Ni ($r = -0.609$, $p = 0.006$) and Cd ($r = 0.524$, $p = 0.021$); Agricultural land at 5000 m with Ni ($r = -0.349$, $p < 0.001$) and Zn ($r = -0.252$, $p < 0.001$); and Natural barren land at 5000 m with Zn ($r = 0.334$, $p < 0.001$). Notably, Forest cover at 5000 m exhibited a strong negative correlation with atmospheric HCHO ($r = -0.723$, $p < 0.001$), which may reflect a protective forest canopy effect on local air quality, although confounding spatial factors cannot be ruled out. The 5000 m buffer yielded the highest percentage of significant LULC–pollution correlations (34.7%) compared to 1000 m (23.8%), while the 1000 m buffer produced stronger mean absolute correlation coefficients ($|r| = 0.158$ vs. 0.110). NPI differed significantly between areas of high vs. low agricultural coverage (Welch $t = -3.44$, $p = 0.00067$), though the result likely reflects a spatial confound whereby intensive agriculture occupies the flat valley center where geogenic Cr and Ni concentrations from calcareous parent materials are lower.

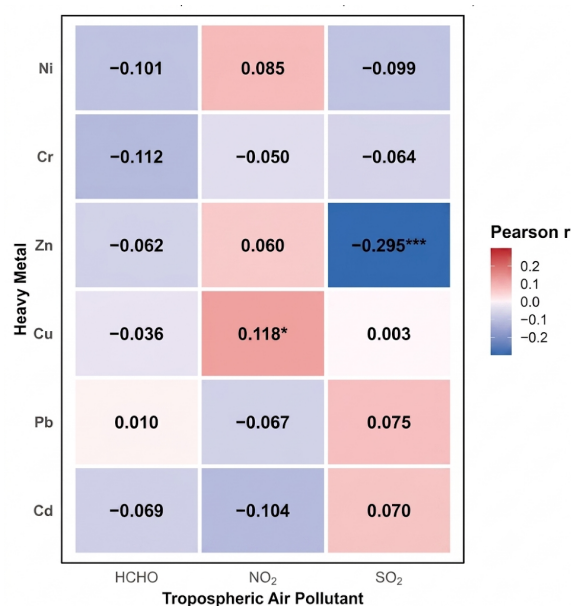


Figure 7. Correlation heatmap between heavy metals and air pollutants (1000 m buffer). Predominantly weak correlations ($|r| < 0.3$) indicate independent emission sources. Significance * = $p < 0.05$ and *** = $p < 0.001$.

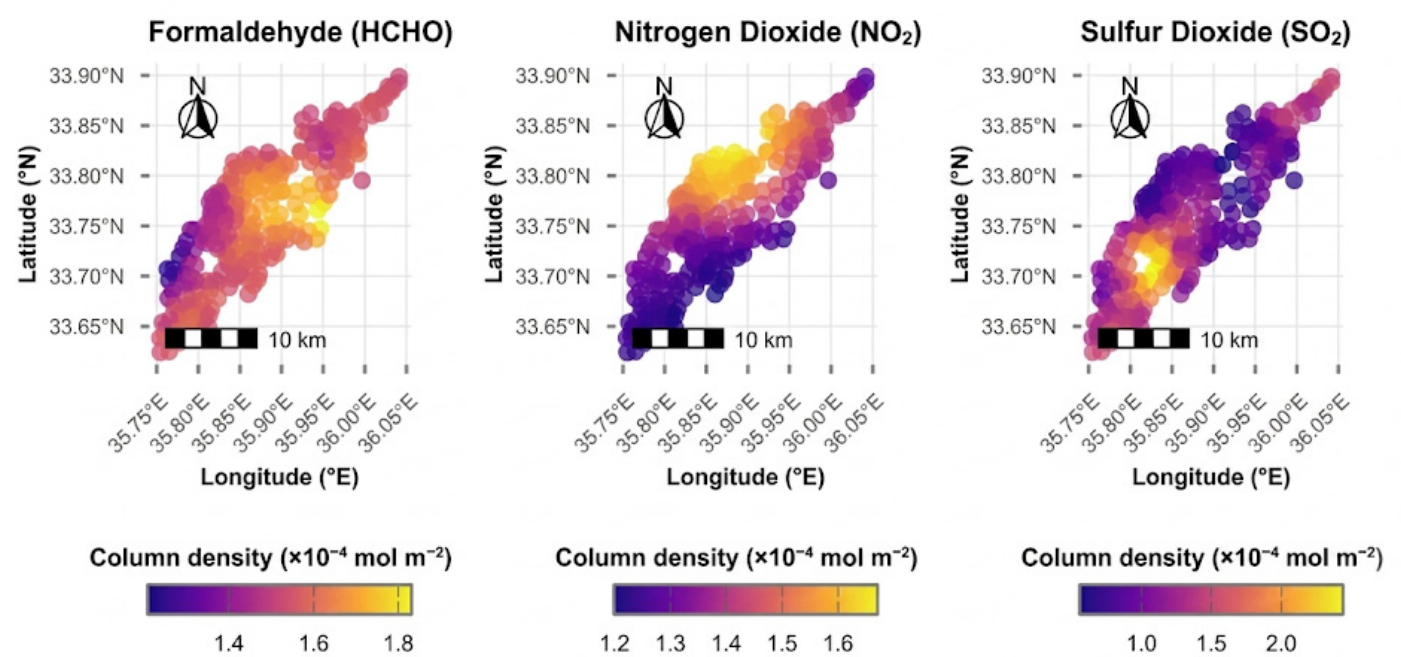


Figure 8. Spatial distribution of tropospheric air pollutants at the 1000 m buffer scale (Sentinel-5P TROPOMI, June–October 2019). SO₂ exhibits the highest spatial variability (CV = 32.4%) with localized industrial emission hotspots. Color scale: Low (dark) to high (bright) concentrations.

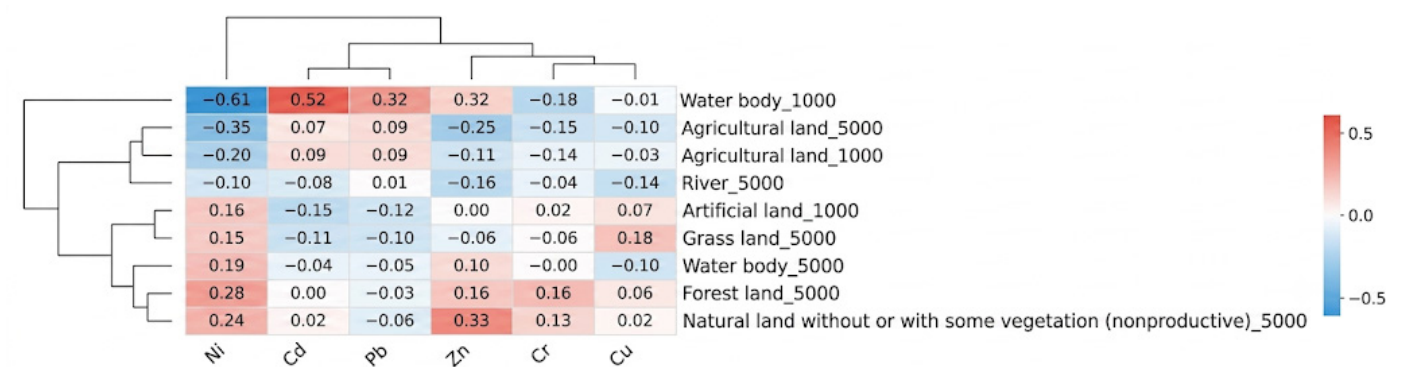


Figure 9. Correlation heatmap of land cover type $\times$ heavy metal across buffer zones. Red = positive; Blue = negative correlation. Hierarchical clustering reveals the Cr–Ni geogenic cluster vs. the Cd–Pb anthropogenic cluster. Nineteen significant correlations were identified ($p < 0.05$).

3.7. Local Spatial Autocorrelation: Cluster and Hotspot Identification

The Global Moran’s I established that soil contamination is significantly clustered across the study area (Section 3.3), but this single, area-wide statistic cannot indicate where those clusters lie. To map the individual hotspots and cold-spots—a necessary first step for ranking the highest-priority sampling sites—we therefore applied two local indicators of spatial association: the Local Moran’s I (LISA) and the Getis–Ord Gi* statistic. Both were computed for every metal and for the composite NPI using the same spatial weights as the global analysis ($k = 8$ nearest neighbours, row-standardized) and the analytical randomisation significance test (Table 5, Figure 10).

Table 5. Local spatial autocorrelation of soil heavy metals and the composite NPI.

Variable	HH	LL	HL	LH	Gi* Hot ($p \leq 0.05$)	Gi* Cold ($p \leq 0.05$)	Gi* Hot ($p \leq 0.01$)
Cd	33	0	0	0	46	0	42
Pb	36	0	1	0	44	0	33
Ni	32	31	2	2	38	36	36
Cr	22	43	2	2	35	47	14
Zn	24	22	1	3	34	39	18
Cu	8	0	1	6	22	0	15
NPI	19	38	2	2	35	41	14

Local Moran's I (LISA) cluster counts (HH = High-High, LL = Low-Low, HL = High-Low, LH = Low-High) and Getis-Ord Gi* hot-/cold-spot counts at the 95% confidence level (analytical randomisation test, $p \leq 0.05$); the final column reports Gi* hotspots significant at the 99% level. Spatial weights as in Table 2 ($k = 8$ nearest neighbours).

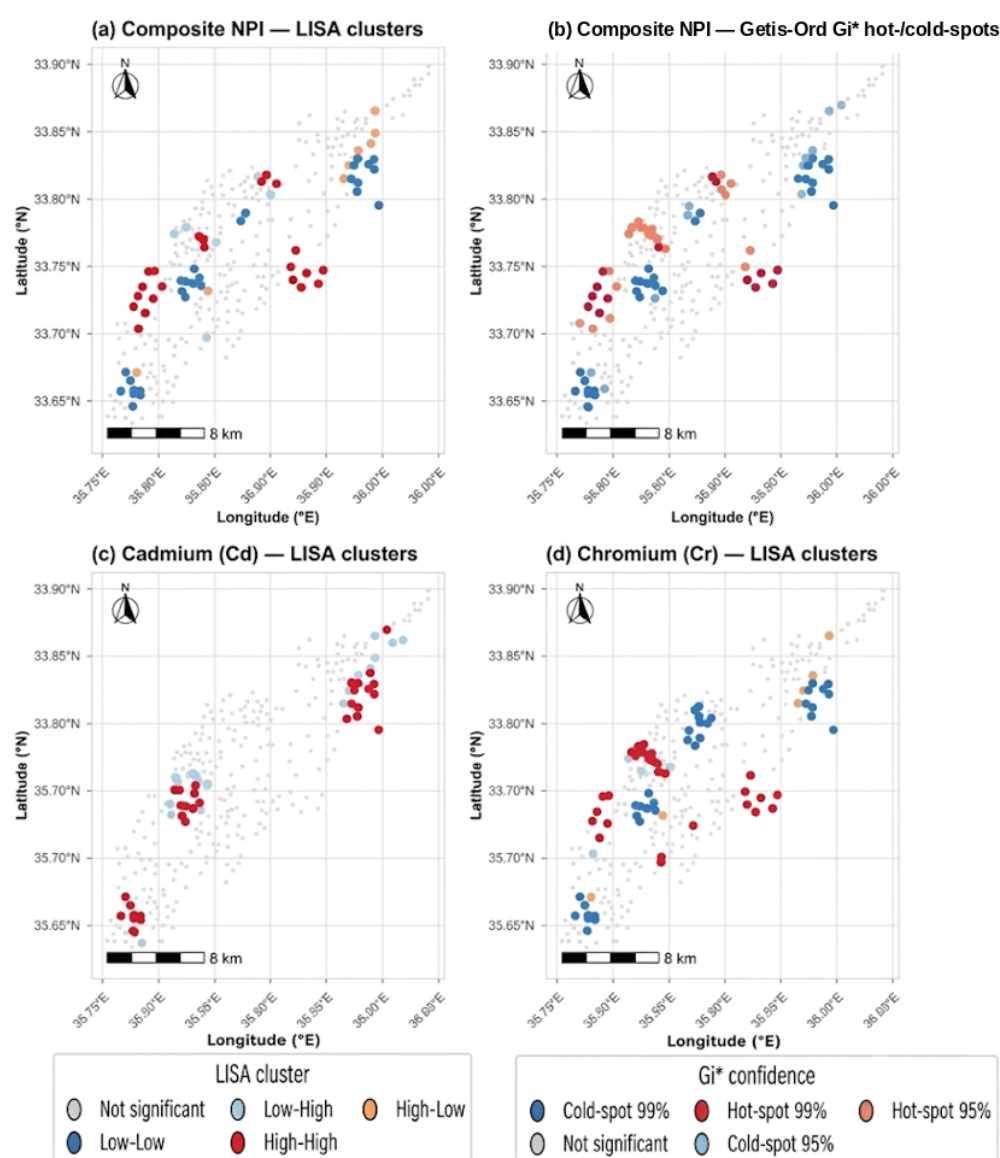


Figure 10. Local spatial autocorrelation of soil contamination in the Bekaa Valley. (a) Local Moran's I (LISA) cluster map and (b) Getis-Ord Gi* hot-/cold-spot map of the composite Nemerow Pollution Index (NPI); (c) LISA cluster map of cadmium (Cd, anthropogenic marker) and (d) chromium (Cr, geogenic marker). LISA clusters and Gi* hot-/cold-spots are significant at $p \leq 0.05$ (analytical randomisation test); Gi* confidence is additionally shown at the 99% level. Spatial weights: $k = 8$ nearest neighbours, row-standardized, as in Table 2. Each panel includes a north arrow and a graphical scale bar.

For the composite NPI, the LISA analysis identified 19 High–High core-cluster sites and 38 Low–Low sites, plus only four spatial outliers (two High–Low and two Low–High). The Getis–Ord G_i^* statistic flagged 35 significant hotspot sites ($p \leq 0.05$, of which 14 at $p \leq 0.01$), concentrated in the central–southern valley, and 41 cold-spot sites in the north (Figure 10a,b). Importantly, of the 30 sites in the upper NPI decile (Section 3.2), only 14 coincided with a statistically significant G_i^* hotspot (and 15 with a LISA High–High cluster); the remaining upper-decile sites were largely locally non-significant (13 sites) or High–Low spatial outliers (2 sites)—isolated high values surrounded by lower ones. This shows that defining hotspots purely by their rank in the value distribution (the top decile) mixes genuine spatial clusters with isolated or cluster-edge sites. The local statistics therefore provide a more defensible, statistically grounded basis for selecting priority zones for confirmatory sampling.

The local statistics also independently support the source groupings inferred in Section 3.4. The likely anthropogenic pair, Cd and Pb, formed tight High–High clusters (33 and 36 sites, respectively) with no Low–Low clusters and the largest numbers of high-confidence G_i^* hotspots (42 and 33 sites at $p \leq 0.01$)—a pattern typical of localized, point-source inputs. By contrast, the likely geogenic pair, Cr and Ni, showed a two-sided (bipolar) structure, combining High–High clusters (22 and 32 sites) with extensive Low–Low clusters (43 and 31 sites); this points to a broad, spatially continuous lithological gradient rather than to discrete sources. Copper had the weakest local structure (only 8 High–High sites and no Low–Low clusters), consistent with its low global Moran’s I and a diffuse, agriculturally driven distribution. Finally, the clear spatial separation between the Cd hotspots and the Cr clusters (Figure 10c,d) reinforces the conclusion that these elements are controlled by different processes.

4. Discussion

4.1. Contamination Status in a Mediterranean and Global Context

The Bekaa Valley currently maintains an overall good soil quality status (mean NPI = 0.667), yet the dominance of warning-level sites (52.9%) highlights a system under increasing environmental pressure. This transitional condition is consistent with recent global assessments showing that agricultural soils in semi-arid and Mediterranean regions are increasingly shifting from safe to precautionary thresholds due to cumulative anthropogenic inputs and climate-related stressors [4,41]. Chromium emerges as the principal contaminant, with approximately 79% of sites exceeding the CCME guideline of 64 mg/kg and 38.3% exceeding the 100 mg/kg international reference value, aligning with global observations that identify Cr and Ni as dominant elements in soils derived from ultramafic and calcareous parent materials [4]. The moderate variability (CV = 37%) and spatial continuity of Cr suggest a predominantly geogenic origin, further supported by its inverse association with anthropogenic elements (Pb, Cd). However, localized enrichment (EF up to 3.17) indicates that anthropogenic contributions—potentially from industrial emissions, wastewater irrigation, or agrochemicals—cannot be excluded. The low mean concentrations of Cd and Pb are encouraging from a food safety perspective. Nevertheless, their extremely high coefficients of variation (>230%) and strong spatial clustering indicate localized contamination hotspots, a pattern widely reported in agricultural landscapes impacted by point-source pollution such as waste dumping, traffic emissions, and small-scale industries [2,42]. The Cr exceedance figures cited above are, again, DTPA-based screening flags rather than regulatory determinations: they conservatively indicate sites that warrant confirmatory total-metal analysis, but the remaining sites cannot be assumed to comply with total-metal guidelines on the basis of DTPA data alone.

4.2. Integrated Source Apportionment and Process Interpretation

The interpretation of contamination sources was guided by both the statistical analyses and the environmental context of the Bekaa Valley. The region is subject to a combination of natural and anthropogenic pressures, including weathering of carbonate and marl parent materials, intensive agricultural practices involving fertilizers and Cu-based fungicides, major transportation corridors linking Beirut to Syria, industrial activities concentrated around urban centers, quarrying operations, and widespread waste dumping and open burning reported in several parts of the valley. These potential sources provide an important framework for interpreting the observed metal distribution patterns.

The combination of PCA, spatial autocorrelation, coefficient of variation, and land-use correlation analyses suggests the presence of three major contamination components reflecting both natural and human influences. The first component, represented by Cr, Ni, and Zn, is characterized by negative loadings on PC1, moderate coefficients of variation (32–47%), and positive correlations with forest and natural land cover at the 5000 m scale. These patterns are consistent with a substantial geogenic contribution, likely associated with the weathering of local parent materials. Previous studies have highlighted the dominant role of lithology in controlling baseline concentrations of Cr and Ni in Mediterranean soils, particularly in calcareous and alkaline environments [41]. The negative correlations with agricultural land (Ni: $r = -0.349$; Zn: $r = -0.252$ at 5000 m) may further reflect dilution effects associated with cultivation practices, soil mixing, or sediment redistribution.

A second component, dominated by Cd and Pb, loads positively on PC1 and exhibits high coefficients of variation ($CV > 230\%$), pronounced spatial clustering (Moran's $I = 0.44$ – 0.47), and a strong inter-metal correlation ($r = 0.878$). These characteristics indicate localized inputs and suggest a common anthropogenic origin. Within the Bekaa Valley, potential contributors include industrial activities, emissions from major transportation corridors, waste burning, uncontrolled dumping sites, and historical deposition from legacy leaded gasoline use. Similar Cd–Pb associations have been reported in agricultural and peri-urban environments affected by mixed industrial and traffic-related contamination [43,44]. However, because direct source-tracing techniques such as isotopic fingerprinting were not employed, these source assignments should be considered indicative rather than definitive.

Copper displayed a distinct behavior, loading primarily on PC2 (0.762) and exhibiting weak correlations with all other metals ($|r| < 0.08$). This independence suggests that Cu is controlled by processes different from those governing the remaining trace elements. Given the widespread cultivation of orchards and vineyards in the Bekaa Valley and the long-term use of Cu-based fungicides, agricultural practices likely represent an important source of Cu enrichment. This interpretation is supported by previous studies identifying pesticide application as a major driver of copper accumulation in agricultural soils [45]. The relatively low Moran's I value (0.169) further indicates a diffuse spatial pattern consistent with widespread agricultural inputs rather than localized point sources.

Overall, the convergence of multivariate statistics, spatial analyses, and land-use relationships supports a mixed contamination scenario in which natural geological controls dominate the distribution of Cr and Ni, agricultural practices contribute substantially to Cu accumulation, and localized anthropogenic activities are likely responsible for elevated Cd and Pb concentrations. Nevertheless, confirmation of these source categories would require complementary approaches such as atmospheric deposition monitoring, historical emission inventories, or isotopic source fingerprinting. This mixed scenario is independently reinforced by the local spatial-autocorrelation analysis (Section 3.7): the anthropogenic Cd–Pb pair forms tight High–High clusters with no Low–Low clusters, whereas the geogenic Cr–Ni pair shows a bipolar High–High/Low–Low structure indicative of a continuous lithological gradient, and Cu shows only weak, diffuse local structure.

4.3. Decoupling Between Soil and Atmospheric Pollution Compartments

One of the most significant findings of this study—and, given the study design, an anticipated one—is the pervasive weak coupling between soil metals and tropospheric air pollutants. Despite the use of high-resolution Sentinel-5P data, correlations remained negligible ($|r| < 0.3$ for 17/18 pairs), so that the data support the null hypothesis (H0) of spatially decoupled compartments rather than H1. This decoupling can be attributed to three primary factors that were expected a priori: (i) temporal mismatch—soil metals reflect decades of cumulative deposition, whereas satellite-derived concentrations represent a single-season atmospheric snapshot (June–October 2019); (ii) spatial scale mismatch—soil samples are point-based (sub-meter), while TROPOMI pixels average over 3.5×7 km footprints; and (iii) process-level differences—soil contamination is governed by deposition, adsorption, irrigation quality, and agrochemical inputs, whereas atmospheric pollutant concentrations respond to emission rates, photochemistry, and meteorological transport. Rather than a null or negative outcome, this result is itself informative: it quantifies the limits of substituting atmospheric remote sensing for direct soil monitoring and delineates the conditions (long-term, deposition-based observation) under which the two compartments could be meaningfully linked. Similar conclusions have been reported in recent satellite-based studies emphasising the difficulty of linking soil contamination with atmospheric observations without long-term integrated datasets [6,10]. The single weak Cu–NO₂ association ($r = 0.118$, $r^2 \approx 1.4\%$) is not sufficient to support any causal inference.

4.4. Multi-Scale Analysis and Methodological Implications

The multi-scale buffer analysis yields important methodological insights for environmental monitoring design. The 5000 m buffer captures broader landscape influences and yields the highest proportion of significant LULC–pollution relationships (34.7%), while the 1000 m buffer captures stronger localized interactions (mean $|r| = 0.158$). The extremely high redundancy in air pollutant data across scales ($r > 0.96$ for Point vs. 1000 m) confirms that Sentinel-5P products are effectively scale-invariant at local study-area dimensions, consistent with recent validation studies [7]. In contrast, LULC variables show moderate to low inter-scale correlation ($r = 0.41$ – 0.46), indicating that landscape context is genuinely scale-dependent. The recommended dual-scale framework (point + 5000 m) provides an optimal balance between local precision and regional representativeness and can serve as a reference protocol for future multi-compartment environmental studies in data-scarce regions.

4.5. Implications for Environmental Management in Lebanon

This study highlights a critical gap in Lebanon’s environmental governance: the absence of locally calibrated soil background values. Reliance on international standards may lead to systematic misclassification of naturally enriched soils as contaminated, particularly for geogenically abundant elements such as Cr. The 38.3% exceedance rate relative to the commonly adopted 100 mg/kg international reference value rises to approximately 79% under the CCME agricultural soil quality guideline (64 mg/kg) [16]—yet both may partially reflect natural calcareous/ophiolitic enrichment rather than anthropogenic pollution. This benchmark sensitivity analysis demonstrates that the choice of regulatory standard significantly affects contamination assessment results and reinforces the urgent need for establishing region-specific geochemical baselines, disaggregated by parent material and soil type [46]. Furthermore, the coexistence of soil contamination hotspots and atmospheric pollution clusters within a topographically confined valley suggests compounded environmental risk. Bekaa’s geomorphology limits pollutant dispersion, enhancing both soil accumulation and air stagnation—conditions likely to worsen under projected cli-

mate change scenarios of increased aridity and reduced precipitation [14]. As with every exceedance statistic in this study, the 38.3% and ~79% figures are bioavailability-based screening flags computed on a DTPA basis; they are not equivalent to total-metal regulatory exceedances, and sites below a threshold are not thereby shown to comply.

4.6. Study Limitations and Future Directions

Several limitations should be acknowledged. First, the DTPA extraction method measures the plant-available fraction of metals rather than total metal content and may therefore underestimate absolute contamination levels. Future studies should complement DTPA extraction with total digestion methods (e.g., aqua regia) to provide a more comprehensive assessment of metal accumulation. The magnitude and even the direction of this bias are element-specific: in the studied calcareous soils, strongly retained or lithogenic elements such as Cr and Pb typically have low DTPA-extractable-to-total ratios, so their DTPA-based exceedance percentages act as conservative lower bounds that may substantially understate total-metal exceedance, whereas more labile elements yield extractable fractions that represent a larger share of the total. The reported exceedance rates should therefore be read as bioavailability-based screening indicators rather than as total-metal exceedances, and confirmatory aqua-regia/ICP-MS digestion is recommended to establish definitive total concentrations. It follows that the NPI classification and the exceedance statistics presented in this study must not be interpreted as regulatory-compliance assessments. Consistent with the extraction chemistry, the inference they support is strictly one-directional: a DTPA-based exceedance is a conservative flag for probable total-metal exceedance, whereas a site that does not exceed a screening threshold on a DTPA basis has not been shown to comply with total-metal guidelines. Definitive classification in either direction requires total-digestion (aqua-regia/ICP-MS) confirmation.

Second, the soil survey was conducted during a single sampling campaign (September 2019), providing only a temporal snapshot of contamination conditions. Seasonal and inter-annual variability in soil metal concentrations, particularly for elements influenced by irrigation practices, land management, and episodic pollution events, remains uncharacterized.

Third, as already discussed in Section 4.3, relationships between soil heavy metals and satellite-derived atmospheric pollutants are constrained by fundamental temporal and spatial-support mismatches: the Sentinel-5P products represent tropospheric vertical column densities integrated over multi-kilometre pixels and a single 2019 season, whereas soil metal integrates decades of cumulative processes. To avoid repetition, we note here only that the observed correlations should not be read as causal links between contemporary atmospheric pollution and soil metal accumulation, but at most as indicators of potential spatial co-occurrence of persistent common sources; establishing causal relationships would require long-term atmospheric deposition monitoring, historical emission inventories, and time-series analysis.

Fourth, the absence of nationally established Lebanese soil background values necessitated reliance on global reference concentrations, which may not fully reflect local geological and pedological conditions. The development of regional geochemical baselines would substantially improve future contamination assessments and source attribution analyses.

Finally, this study focused on environmental contamination patterns and did not assess metal bioavailability in crops, transfer through the food chain, or associated human health risks. Addressing these questions would require plant tissue analyses, bioaccumulation studies, and quantitative exposure assessments.

These limitations identify several priorities for future research, including longitudinal monitoring programs, atmospheric deposition studies, development of national geochemi-

cal baseline datasets, isotopic source fingerprinting (e.g., Pb isotopes), and comprehensive ecological and human health risk assessments.

5. Conclusions

This study provides the first integrated multi-compartment assessment of soil and atmospheric pollution in Lebanon's Bekaa Valley, combining ground-based geochemistry of six heavy metals, satellite-derived air pollutant measurements, and multi-scale spatial statistics. The principal findings and their implications are:

1. The Bekaa Valley's agricultural soils maintain overall good environmental quality (mean NPI = 0.667; 98% of sites below the pollution threshold), but the high proportion of warning-level sites (52.9%) indicates an environmental system approaching critical limits that requires proactive management rather than reactive remediation. This status is expressed on a DTPA-based, bioavailability-oriented screening basis and should not be read as demonstrating regulatory (total-metal) compliance; in particular, the 98% of sites below the screening threshold have not been shown to comply with total-metal guidelines, and total-digestion confirmation remains necessary.
2. Three plausible contamination source groupings were inferred from convergent PCA, correlation, coefficient of variation and spatial analyses: a predominantly geogenic Cr–Ni–Zn assemblage controlled by lithological parent material; a localized Cd–Pb grouping ($r = 0.878$) with consistent anthropogenic emissions (industrial/traffic) inputs; and an isolated agricultural Cu source (PC2 loading = 0.762) linked to fungicide application. Because these assignments are based on covariation rather than direct source fingerprinting, they should be regarded as indicative working hypotheses requiring confirmation (e.g., by Pb-isotope analysis) rather than as established attributions.
3. Soil metals and tropospheric air pollutants operate through largely independent pathways ($|r| < 0.3$ for 17/18 pairs), suggesting that atmospheric remote sensing alone cannot substitute for direct soil monitoring and that compartment-specific management strategies are essential.
4. Multi-scale analysis supports a dual-buffer (point + 5000 m) analytical framework: air pollutant buffers exhibit very high inter-scale redundancy ($r > 0.96$), while land cover buffers capture genuinely scale-dependent landscape information ($r \approx 0.4$ – 0.5).
5. For policy, and keeping the above constraints in mind, the results point to four main needs: (a) develop national soil background values calibrated to Lebanon's diverse geological substrates and defined on a total-metal basis; (b) undertake prioritised screening and confirmatory (total-digestion) analysis at the statistically significant hotspot sites delineated by local spatial-autocorrelation analysis (Getis–Ord G_i^* /LISA, Section 3.7) before any remediation decision; (c) work towards integrated soil–air environmental monitoring frameworks; and (d) promote sustainable agricultural practices to limit long-term metal accumulation from agrochemicals.

These conclusions are necessarily exploratory: they rest on a single-season, DTPA-based soil survey and on source assignments derived from statistical covariation rather than direct source tracing, and should be confirmed before being used as a basis for regulatory action. Future research should prioritize temporal monitoring to assess contamination trajectories, isotopic fingerprinting (Pb isotopes) for definitive source attribution, metal bioavailability and crop uptake assessments to quantify food safety risks, and extension of this multi-compartment methodology to other data-scarce Mediterranean agricultural regions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app16157791/s1>, Figure S1: Scree Plot: Eigenvalues and Variance Explained. Bars = Variance (%) | Line = Eigenvalue. Components with eigenvalue ≥ 1 (above the red line) are retained for interpretation.

Author Contributions: Conceptualization, M.T.A.S. and R.A.; methodology, M.T.A.S., M.H.S. and R.A.; software, M.H.S.; formal analysis, M.H.S., E.S. and M.T.A.S.; investigation, E.S., S.F., D.R., Y.K. and R.S.; data curation, E.S., D.R. and Y.K.; writing—original draft, E.S. and M.T.A.S.; writing—review and editing, E.S., M.T.A.S., C.S., R.S., R.A. and M.H.S.; supervision, R.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request. Sentinel-5P TROPOMI data are freely available via the Copernicus Open Access Hub. The Google Earth Engine script used for air pollutant extraction can be provided upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Billah, M.M.; Kokushi, E.; Uno, S. Distribution, Geochemical Speciation, and Bioavailable Potencies of Cadmium, Copper, Lead, and Zinc in Sediments from Urban Coastal Environment in Osaka Bay, Japan. *Water Air Soil Pollut.* **2019**, *230*, 157. [\[CrossRef\]](#)
2. Bing, H.; Wu, Y.; Zhou, J.; Sun, H.; Wang, X.; Zhu, H. Spatial variation of heavy metal contamination in the riparian sediments after two-year flow regulation in the Three Gorges Reservoir, China. *Sci. Total Environ.* **2019**, *649*, 1004–1016. [\[CrossRef\]](#)
3. Ren, Z.; Xiao, R.; Zhang, Z.; Lv, X.; Fei, X. Risk assessment and source identification of heavy metals in agricultural soil: A case study in the coastal city of Zhejiang Province, China. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 2109–2118. [\[CrossRef\]](#)
4. Hou, D.; Jia, X.; Wang, L.; McGrath, S.P.; Zhu, Y.-G.; Hu, Q.; Zhao, F.-J.; Bank, M.S.; O'Connor, D.; Nriagu, J. Global soil pollution by toxic metals threatens agriculture and human health. *Science* **2025**, *388*, 316–321. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Krotkov, N.A.; McLinden, C.A.; Li, C.; Lamsal, L.N.; Celarier, E.A.; Marchenko, S.V.; Swartz, W.H.; Bucsela, E.J.; Joiner, J.; Duncan, B.N.; et al. Aura OMI observations of regional SO₂ and NO₂ pollution changes from 2005 to 2015. *Atmos. Chem. Phys.* **2016**, *16*, 4605–4629. [\[CrossRef\]](#)
6. Maurya, N.K.; Pandey, P.C.; Sarkar, S.; Kumar, R.; Srivastava, P.K. Spatio-Temporal Monitoring of Atmospheric Pollutants Using Earth Observation Sentinel 5P TROPOMI Data: Impact of Stubble Burning a Case Study. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 301. [\[CrossRef\]](#)
7. Kaplan, G.; Yigit Avdan, Z. Space-borne air pollution observation from sentinel-5p tropomi: Relationship between pollutants, geographical and demographic data. *Int. J. Eng. Geosci.* **2020**, *5*, 130–137. [\[CrossRef\]](#)
8. Ghoveh Nodoushan, M.A.; Pormazar, S.M. Air quality and its impact on climate change in major urban areas across Cairo and Moscow cities. *J. Nat. Spat. Sci.* **2023**, *5*, 50.
9. Velayarce, D.; Bustos, Q.; García, M.P.; Timaná, C.; Carbajal, R.; Salvatierra, N.; Horna, D.; Murray, V. Air Quality Analysis in Lima, Peru Using the NO₂ Levels during the COVID-19 Pandemic Lockdown. *Atmosphere* **2022**, *13*, 373. [\[CrossRef\]](#)
10. Faisal, M.; Jaelani, L.M. Spatio-temporal analysis of nitrogen dioxide (NO₂) from Sentinel-5P imageries using Google Earth Engine changes during the COVID-19 social restriction policy in Jakarta. *Nat. Hazards Res.* **2023**, *3*, 344–352. [\[CrossRef\]](#)
11. Lovynska, V.; Bayat, B.; Bol, R.; Moradi, S.; Rahmati, M.; Raj, R.; Sytnyk, S.; Wiche, O.; Wu, B.; Montzka, C. Monitoring Heavy Metals and Metalloids in Soils and Vegetation by Remote Sensing: A Review. *Remote Sens.* **2024**, *16*, 3221. [\[CrossRef\]](#)
12. Wiczorek, B. Air Pollution Patterns Mapping of SO₂, NO₂, and CO Derived from TROPOMI over Central-East Europe. *Remote Sens.* **2023**, *15*, 1565. [\[CrossRef\]](#)
13. Wang, C.; Wang, T.; Wang, P.; Wang, W. Assessment of the Performance of TROPOMI NO₂ and SO₂ Data Products in the North China Plain: Comparison, Correction and Application. *Remote Sens.* **2022**, *14*, 214. [\[CrossRef\]](#)
14. Verner, D.; Ashwill, M.; Christensen, J.; McDonnell, R.; Redwood, J.; Jomaa, I.; Saade, M.; Massad, R.; Chehade, A.; Bitar, A.W.; et al. *Droughts and Agriculture in Lebanon*; World Bank Group: Washington, DC, USA, 2018. [\[CrossRef\]](#)

15. Darwish, T.; Jomaa, I.; Sukarieh, W.; Chihny, R. Vulnerability of soil resources to heavy metals contamination in Central Bekaa-Lebanon. *Sci. Meet. Abst.* **2000**, 214. Available online: <https://inis.iaea.org/records/ar61n-y8x85> (accessed on 31 July 2026).
16. CCME (Canadian Council of Ministers of the Environment). Canadian Soil Quality Guidelines (CSQG) for the Protection of Environmental and Human Health: Summary Tables. In *Canadian Environmental Quality Guidelines, 1999*; Canadian Council of Ministers of the Environment: Winnipeg, MB, Canada, 2007; pp. 1–9.
17. Darwish, T.; Jomaa, I.; Awad, M.; Boumetri, R. Preliminary contamination hazard assessment of land resources in Central Bekaa plain of Lebanon. *Leb. Sci. J.* **2008**, 9, 3–15.
18. Darwish, T.M.; Hamze, M.; Nsouli, B. Quaternary agricultural soils of Lebanon–East Mediterranean. *Leb. Sci. J.* **2021**, 22, 185–208.
19. Kassir, L.N.; Darwish, T.; Shaban, A.; Olivier, G.; Ouaini, N. Mobility and bioavailability of selected trace elements in Mediterranean red soil amended with phosphate fertilizers: Experimental study. *Geoderma* **2012**, 189–190, 357–368. [\[CrossRef\]](#)
20. Merhaby, D.; Ouddane, B.; Net, S.; Halwani, J. Assessment of trace metals contamination in surficial sediments along Lebanese Coastal Zone. *Mar. Pollut. Bull.* **2018**, 133, 881–890. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Moustafa, A.; Hamze, M.; Baroudi, M. Assessment of heavy metal and metalloids pollution in soils and groundwater in Akkar Northen Lebanon. *J. Appl. Sci. Res.* **2019**, 15, 28–38. [\[CrossRef\]](#)
22. Caiserman, A.; Dumas, D.; Bennafla, K.; Faour, G.; Amiraslani, F. Application of Remotely Sensed Imagery and Socioeconomic Surveys to Map Crop Choices in the Bekaa Valley (Lebanon). *Agriculture* **2019**, 9, 57. [\[CrossRef\]](#)
23. Abi Saab, M.T.; Jammoul, D.; Makhoul, H.; Fahed, S.; Lebous, N.; Hajjar, C.; Abi Saad, R.; Younes, M.; Hajj, M.; Todorovic, M. Assessing the performance of constructed wetland for water quality management of a Southern Mediterranean River. *Water Environ. J.* **2018**, 32, 508–518. [\[CrossRef\]](#)
24. Lindsay, W.L.; Norvell, W.A. Development of a DTPA Soil Test for Zinc, Iron, Manganese, and Copper. *Soil Sci. Soc. Am. J.* **1978**, 42, 421–428. [\[CrossRef\]](#)
25. ISO 14870:2001; Soil Quality—Extraction of Trace Elements by Buffered DTPA Solution. International Organization for Standardization: Geneva, Switzerland, 2001. Available online: <https://www.iso.org/standard/25232.html> (accessed on 21 July 2022).
26. Gao, H.; Koopmans, G.F.; Song, J.; Groenenberg, J.E.; Liu, X.; Comans, R.N.J.; Weng, L. Evaluation of heavy metal availability in soils near former zinc smelters by chemical extractions and geochemical modelling. *Geoderma* **2022**, 423, 115970. [\[CrossRef\]](#)
27. Nemerow, N.L. *Stream, Lake, Estuary, and Ocean Pollution*, 2nd ed.; Van Nostrand Reinhold: New York, NY, USA, 1991.
28. Cheng, J.; Shi, Z.; Zhu, Y. Assessment and mapping of environmental quality in agricultural soils of Zhejiang Province, China. *J. Environ. Sci.* **2007**, 19, 50–54. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Sutherland, R.A. Bed sediment-associated trace metals in an urban stream, Oahu, Hawaii. *Environ. Geol.* **2000**, 39, 611–627. [\[CrossRef\]](#)
30. Google Earth Engine. A Planetary-Scale Platform for Earth Science Data & Analysis. Available online: <https://earthengine.google.com> (accessed on 21 July 2022).
31. ESA Copernicus Sentinel-5P. Available online: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-5P (accessed on 21 July 2022).
32. Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, 42, 7561–7578. [\[CrossRef\]](#)
33. Anselin, L. Local Indicators of Spatial Association—LISA. *Geogr. Anal.* **1995**, 27, 93–115. [\[CrossRef\]](#)
34. Getis, A.; Ord, J.K. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr. Anal.* **1992**, 24, 189–206. [\[CrossRef\]](#)
35. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2025. Available online: <https://www.R-project.org/> (accessed on 1 July 2025).
36. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, 4, 1686. [\[CrossRef\]](#)
37. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* **2018**, 10, 439–446. [\[CrossRef\]](#)
38. Bivand, R.S.; Wong, D.W.S. Comparing implementations of global and local indicators of spatial association. *TEST* **2018**, 27, 716–748. [\[CrossRef\]](#)
39. Lê, S.; Josse, J.; Husson, F. FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* **2008**, 25, 1–18. [\[CrossRef\]](#)
40. Kassambara, A.; Mundt, F. *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses*, R package version 1.0.7; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: <https://CRAN.R-project.org/package=factoextra> (accessed on 1 July 2025).
41. Alnaimy, M.A.; Elrys, A.S.; Zelenakova, M.; Pietrucha-Urbanik, K.; Merwad, A.-R.M. The Vital Roles of Parent Material in Driving Soil Substrates and Heavy Metals Availability in Arid Alkaline Regions: A Case Study from Egypt. *Water* **2023**, 15, 2481. [\[CrossRef\]](#)

42. Wu, J.; Long, J.; Liu, L.; Li, J.; Liao, H.; Zhang, M.; Zhao, C.; Wu, Q. Risk Assessment and Source Identification of Toxic Metals in the Agricultural Soil around a Pb/Zn Mining and Smelting Area in Southwest China. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1838. [[CrossRef](#)] [[PubMed](#)]
43. Zhang, P.; Qin, C.; Hong, X.; Kang, G.; Qin, M.; Yang, D.; Pang, B.; Li, Y.; He, J.; Dick, R.P. Risk assessment and source analysis of soil heavy metal pollution from lower reaches of Yellow River irrigation in China. *Sci. Total Environ.* **2018**, *633*, 1136–1147. [[CrossRef](#)] [[PubMed](#)]
44. Karim, Z.; Qureshi, B.A.; Mumtaz, M.; Qureshi, S. Heavy metal content in urban soils as an indicator of anthropogenic and natural influences on landscape of Karachi—A multivariate spatio-temporal analysis. *Ecol. Indic.* **2014**, *42*, 20–31. [[CrossRef](#)]
45. Muhie, S.H. Novel approaches and practices to sustainable agriculture. *J. Agric. Food Res.* **2022**, *10*, 100446. [[CrossRef](#)]
46. Sahoo, P.K.; Dall'Agnol, R.; Salomão, G.N.; da Silva Ferreira Junior, J.; da Silva, M.S.; Martins, G.C.; e Souza Filho, P.W.M.; Powell, M.A.; Maurity, C.W.; Angelica, R.S.; et al. Source and background threshold values of potentially toxic elements in soils by multivariate statistics and GIS-based mapping: A high-density sampling survey in the Parauapebas basin, Brazilian Amazon. *Environ. Geochem. Health* **2019**, *42*, 255–282. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.