



Combining indices of soil pollution due to heavy metals

 **Satyendra Nath Chakrabartty**

Indian Ports Association, Indian Statistical Institute, India.

Email: chakrabarttysatyendra3139@gmail.com



ABSTRACT

Article History

Received: 13 August 2026

Revised: 16 September 2026

Accepted: 29 September 2026

Published: 26 October 2026

Keywords

Geometric mean

Heavy metals

Lognormal distribution

Pollution indices

Progress path.

Soil contamination indices based on a single heavy metal (HM) fail to satisfy the interests of regulatory authorities, landowners and researchers who are more interested in a comprehensive index by aggregating such measures, facilitating evaluation of the effectiveness of remedial actions. Existing indices suffer from methodological limitations. Avoiding limitations of existing indices, the paper suggests a composite index of soil contamination by HMs (by multiplicative aggregation of individual contaminant indices) defined as ratio of pooled average concentration of a HM and the Standard for the HM. Logarithm of the index gives an additive model following a lognormal distribution. The index quantifies overall soil pollution status due to the concentration of different HMs, and satisfies desired properties like time reversal test, formation of chain indices and enables comparison across regions or time periods and undertaking parametric analysis. Novelty includes knowledge of the probability distribution of the index, assessment of progress or decline over the previous years and statistical inferences. The proposed approach with significant advantages can be computed for any finite number of HMs irrespective of sources of HMs and helps in better classification of regions with assessment of efficiency of classification. The proposed index with theoretical advantages is recommended.

Contribution/Originality: The study proposes a composite index of soil contamination by heavy metals (HMs) ($I_{HM_{Soil}}$) using multiplicative aggregation of individual contaminant indices ($p_i s$) with knowledge of the probability distribution of the index is a novelty.

1. INTRODUCTION

Various soil pollution indices are used to assess the concentration of individual Heavy Metals (HMs) from natural and anthropogenic sources and resultant degradation of soil (Mazurek et al., 2017). Model-based approaches like Ward's hierarchical cluster analysis (HCA), Principal component analysis (PCA), etc. attempt to find similarities among homogeneous clusters of indices (Murtagh & Legendre, 2014). However, the authors found that two algorithms of the Ward hierarchical clustering method produced different results even when applied to same distance matrix. Risk assessment of HMs in soil was investigated by Yang et al. (2023) considering single-factor contaminant indices, Nemerow index, spatial distributions of HMs, and used the model given by United States Environmental Protection Agency (2011) irrespective of sources of HMs from local as well as distant sources of emissions.

However, an index based on a single specific HM fails to satisfy the interests of regulatory authorities, landowners and researchers who are more interested in aggregated measure of all HMs obtained from the sampling units, facilitating evaluation of the effectiveness of remedial actions, statistical inferences, including estimation of population parameters, etc. (Yoo et al., 2018). There is no agreed method for evaluation of overall contamination of

soil (Xie et al., 2021). Thus, comprehensive assessments are required where individual indices are combined to get a composite index (CI) reflecting overall pollution status of HMs. Clearly, CI is multidimensional, where component indices with different probability distributions of HM concentrations for different locations, different sections of soil follow distributions other than normal distribution (Zhao et al., 2022). Empirical relationships of HMs and soil degradation through correlation and regression analysis without checking assumptions of such statistical techniques may not provide comprehensive information (Kowalska et al., 2016).

The recently proposed Biogeochemical Index (BGI) is taken as ratio of the concentration of a particular HM in the organic horizon (C_O) and concentration in the mineral horizon (C_A) (Mazurek et al., 2017). $BGI > 1$ indicates high accumulation of trace elements like Pb, Zn, Cu, Cd, Ni, etc. in soil. The quality of topsoil is often evaluated by BGI along with the Nemerow pollution index (NPI) and Potential ecological risk index (RI) (Salikhov, Petrov, & Shuralev, 2024). Despite advantages, BGI is associated with problems because of its focus on total concentration at the top layer of soil ignoring density of soil particles separately for O and A horizons, actual bioavailable fraction of HMs, etc.

HMs like lead, mercury and arsenic are classified as highly hazardous substances as their accumulation in bones and parenchymal organs like liver, kidneys and heart may lead to gradual degeneration and reduce responsiveness of the immune system, contributing to neurological disorders and even cancer (Pourret & Hursthouse, 2019; Skugoreva, Ashikhmina, Fokina, & Lyalina, 2016).

However, toxicity and HM concentrations are different concepts. A fixed toxicity coefficient for each HM in Risk Index (RI) often fails to explain variable toxicity of HMs in different environmental conditions. Thus, RI fails to identify the major drivers of contamination and initiate appropriate mitigation strategies. Migration of HMs in farmland, forest and urban soils are due to natural causes and human activities like fossil fuel combustion, urban and agricultural runoff, domestic/industrial waste water effluents, atmospheric deposition, etc. HMs are transferred to food chain, adversely affecting food quality and ultimately the health of animals and humans (Duskaev, Miroshnikov, Sizova, Lebedev, & Notova, 2014). Monitoring of “soil-plant-animal” system for macronutrients, micronutrients and HMs helped to identify bio-geocoenotic pathologies among cows (Nogoibaev, Nogoibaeva, Omurbaeva, Akunova, & Kyrbasheva, 2025). The authors proposed an innovative ecological centre dedicated to safeguarding the health of both humans and animals.

After addressing methodological limitations of the existing indices, the paper suggests a comprehensive index of soil contamination by HMs ($I_{HM_{Soil}}$) using multiplicative aggregation of individual contaminant indices (p_i) defined as $\frac{\text{Concentration of a HM}}{\text{Standard for the HM}}$ where the numerator is the pooled average of HM concentration across samples. The index $I_{HM_{Soil}}$ quantifies overall soil pollution status with respect to HMs, and satisfies desired properties enabling comparison of the indices across regions or time periods, and undertaking parametric analysis.

2. LITERATURE REVIEW

Major problems in combining different soil pollution indices are:

- Different emphasis on different aspects. While popular indices try to measure concentration of HM in soil, other indices attempt to find sources of pollution and pathways.
- Different scoring methods result in non-uniform score ranges of the indices. Cautions are needed in interpreting I_{geo} defined as $\log_2[\frac{C_{Sample}}{1.5 \times C_{Background}}]$ since $I_{geo} < 0$ till $C_{Sample} < 1.5 \times C_{Background}$. Nemerow index, Synthetic Pollution Index consider extreme values emerging from the data, which often are unreliable outliers.
- Different distributions of HM concentrations resulting from different selection of locations, sections of soil, etc. are unknown and are not normal (Zhao et al., 2022).

- Non-meaningful arithmetic aggregation due to dissimilar probability distributions of individual p_i 's and lack of knowledge about the distribution of PI_{Sum}
- Scales of PI_{Sum} , $PI_{Average}$ and PI_{Vector} are not precise and not standardized.
- Classification of soils into different numbers of classes based on different indices produced different results without reporting efficiency of classification. The number of classes was different for different pollution indices. While five classes are usually reported for Single-factor Pollution Index (PI), seven classes were suggested by Sawadogo et al. (2023) for I_{geo} and four classes for CF. Complicated methods using laser-induced breakdown spectroscopy (LIBS) in machine learning (ML) were suggested for large-scale operations of environmental monitoring (Kantemiris, Xenogiannopoulou, Vallas, & Oikonomou, 2025). Even unpolluted regions characterise by $PLI < 1$ may have ($I_{geo} > 0$), implying high contamination (Andem, Okorafor, Oku, & Ugwumba, 2015).
- Use of subjective rating scales giving rise to ordinal data and aggregation using weighted sum like Analytic Hierarchy Process (AHP) by Zhao et al. (2022); HM pollution index (HPI) do not help in meaningful comparisons. Methodological limitations of HPI were illustrated (Satyendra Nath Chakrabarty, 2025). PCA is sensitive to outliers and small samples. Sample-specific PCA weights vary with time and space. Gong et al. (2022) used PCA without verification of normality of the variables. Cluster analysis on accumulation patterns of 10 HMs in 47 soil samples by Ikenaka et al. (2010) failed to produce mutually exclusive clusters. A multi-index framework considered by Brahma, Sarmah, and Sarma (2025) comprising indices like EF, I_{geo} , CF, PLI, PERI and P_N found highest and average concentrations of individual HMs and showed seasonal variations and directional differences of several HMs by techniques like correlation analysis, PCA, cluster analysis, ANOVA, etc. However, each such technique is based on certain assumptions and has limitations. For example, correlation may not confirm causation, ANOVA assumes normal distribution of each index and homogeneity of variance, which are rarely found in skewed data on soil contaminants. Kowalska, Mazurek, Gąsiorek, and Zaleski (2018) compared indices like I_{geo} , PI, PI_{Sum} , $PI_{Nemerow}$, PLI, PI_{Ave} , PI_{Vector} , EF, CF, C_j , PIN, RI, BGI, CSI, MERMQ, MEC, C_{Deg} , mCd, ExF, etc. and suggested a comprehensive pollution index in soils at farmland, forest and urban areas. Commonly used pollution indices based on concentration of HMs in soil and their limitations are shown in Table 1.

Table 1. Indices of heavy metal pollution (Illustrative list).

Index	Purpose	Observations and Limitations
Single-factor contaminant index (PI) $p_i = \frac{C_i}{S_i}$ where C_i and S_i denote respectively concentration and National standard of i-th HM	To compare concentration of HMs in soil with respective standards.	$-p_i + p_j$ is not admissible since evaluation criterion of i^{th} HM and j^{th} HM are different ($i \neq j$). -Testing $H_0: P_i = P_j$ by Z-statistic requires independent samples collected via simple random sampling (SRS).
$PI_{Sum} = \sum_{i=1}^n p_i$ for n -number of HMs or $PI_{Average} = \frac{1}{n} \sum_{i=1}^n p_i$ (Qingjie, Jun, Yunchuan, Qingfei, & Liqiang, 2008)	To assess soil contamination by all analysed HMs	Meaningful addition requires similar probability distribution of each p_i facilitating derivation of distribution of PI_{Sum} . High impact of a HM gets masked due to substitution effect (Mariani & Ciommi, 2022)
$PI_{Vector} = \sqrt{\frac{1}{n} \sum_{i=1}^n p_i^2}$	Overall pollution level based on the vector modulus of single-factor indices.	Scale of PI_{Sum} , $PI_{Average}$ and PI_{Vector} are not precise and not standardized. Incorrect selection of GB can lead to misleading results.
Contamination factor (CF) of i^{th} HM is defined as $\frac{C_{Mean}}{C_{Background}}$	Assessment of soil	Ignores soil properties like pH, organic matter, cation exchange capacity, and intrinsic soil buffering may mislead ecological risk

where mean content of i^{th} HM is taken from ≥ 5 samples.	contamination by a single metal.	Sensitive to background concentration. May overestimate risks from tightly bound HMs which are not bioavailable.
Pollution load index (PLI) = $\sqrt[n]{\sum_{i=1}^n CF_i}$ (Tomlinson, Wilson, Harris, & Jeffrey, 1980)	Assessment of soil contamination level of multiple pollutants at a site	-Better aggregating procedure in absence of knowledge of probability distributions of CFs. But usual consideration of top five CFs gives only approximate PLI-value -Requires specific location-wise value of natural background concentrations. -Ignores synergistic or antagonistic toxic effects of multiple pollutants.
$PIN = \sum_{i=1}^n \frac{PI_{Class}^2 * C_n}{GB}$ where PI_{Class} : class of HM pollution, C_n : contamination by HM	To integrate all contaminations into a single value	-Ignores natural geochemical processes -Requires values of PI_{Class}
Nemerow pollution index P_N or Nemerow integrated pollution index (NIPI) = $\sqrt{\frac{P_{iMax}^2 + P_{iave}^2}{2}}$ where P_{iMax} and P_{iave} denote respectively the maximum and arithmetic mean of P_i s.	Assessment of overall degree of soil pollution including the contents of all analysed HMs	- P_N has 1 – 1 correspondence with Synthetic Pollution Index $SPI = \sqrt{\frac{P_{iMax} + P_{iMin}}{2}}$ (Gong et al., 2022). - P_N and SPI based on extreme values (unrealistic outliers) are influenced by data heterogeneity. - Consideration of P_{iMax} can overestimate the pollution level. -Computation of P_{iave}^2 is not meaningful (Chakrabartty, 2024). - Classification of soil by P_N or SPI needs to report efficiency of classification say by ratio of between and within group variances following F -distribution, Wilk's lambda assuming multivariate normal distribution, etc.
Geo-accumulation index $I_{geo} = \log_2 \left[\frac{C_{Sample}}{1.5 * C_{Background}} \right]$ where C_{Sample} denotes concentration of a particular HM in soil and $C_{Background}$ is same in the background (secondary data).	To assess pollution level for a single element	-Wrong choice of GB distorts the results. $I_{geo} < 0$ for $C_{Sample} < 1.4985 * C_{Background}$ -For a HM, $C_{Background} = C_{Sample} \Rightarrow I_{geo} = \log_2 \left[\frac{1}{1.5} \right] \Rightarrow -0.585$ against $I_{geo} = 0$ when $C_{Sample} = 1.5(C_{Background})$. - For a given $C_{Background}$ $\frac{I_{geo(1)}}{I_{geo(2)}} \neq \frac{C_{Sample(1)}}{C_{Sample(2)}}$ and $\frac{I_{geo(1)} - I_{geo(2)}}{\text{Increase in } I_{geo}}$ is not constant implying I_{geo} and C_{Sample} are not linearly related (Chakrabartty, 2025). - $C_{Background}$ may not be available for all areas (Dutta et al., 2021) - Classification based in I_{geo} ignoring external factors like soil pH, organic matter content, or redox potential may not be beyond question (Ahrivar, Das, Srivastava, & Kumar, 2023). - Aggregation of I_{geo} across different HMs, or multiple sites, is problematic. -Unpolluted regions ($PLI < 1$) may show high contamination in terms of I_{geo} values (Andem et al., 2015).
Improved Nemerow index $I_{IN} = \sqrt{\frac{I_{geoMax} + I_{geoAv}}{2}}$ (Santos-Francés, Martínez-Graña, Alonso Rojo, & García Sánchez, 2017)	Combining all HMs in a sample specially for mining-impacted areas	-Does not specify most suitable averaging procedure. Not suitable for all soil types. -Assumption of linearity between concentration environmental risks of pollutants may not hold for all HMs or pH changes in soil.
Ecological risk index (RI) $RI = \sum E_r$ where Ecological Risk (ER) $E_r = \text{Toxicity}$	Assessment of pollution level based on	-Toxicity coefficients of HM are arbitrarily taken as 1 for Zn; 5 for Cu, , Pb, Ni and 30 for Cd (Li, Ma, Liu, Sun, & Chen, 2015)

coefficient (T_r) * pollution factor (CF) where $CF = \frac{C_{Sample}}{C_{Crust}}$	accumulation and interaction effects of HMs in soil.	-Ignores chemical speciation of HMs (Wang, Wang, Yang, Yu, & Ni, 2023). -Better to use station-wise $PLI = \sqrt[n]{\prod CF_i}$ (Chakravarty & Patgiri, 2009).
Enrichment factor (EF) = $\frac{[\frac{C_n}{LV}]_{Sample}}{[\frac{C_n}{LV}]_{Background}}$ where LV denotes low variability of occurrence.	To quantify impact of anthropogenic activity on the concentration of HMs in soil.	-Common reference elements are Fe, Al, Ca, Ti, Sc or Mn. - $0.5 < EF < 1.5 \Rightarrow$ the HM is caused by natural processes. $EF > 1.5 \Rightarrow$ contamination by the HM is due to anthropogenic activities (Zhang & Liu, 2002). -Selection of appropriate GB is crucial.
Biogeochemical Index (BGI) = $\frac{HM \text{ content in Horizon O}(C_n O)}{HM \text{ content in Horizon A}(C_n A)}$ (Mazurek et al., 2017)	To determine ability of an O-horizon to absorb pollutants. BGI indicates vertical mobility of contaminants.	Ignores: -Density of soil particles separately for O and A horizons, and makes BGI an approximation only (Mazurek et al., 2017) -Actual bioavailable fraction of HMs, and focuses on total concentration, which may fail to reflect ecological risk in an accurate fashion. -Natural background levels in soil, creating difficulty in interpreting high concentration is due to pollution or natural geological causes. - Soil layers below the top layer and thus fail to reflect deeper concentration issues.

Note: In addition, rarely used indices include, among others: Contamination Security Index (CSI), Exposure factor (ExF), Multi-element contamination (MEC), degree of contamination (C_{Deg}), probability of toxicity (MERMQ), etc.

3. OBSERVATIONS

The soil pollution indices based on concentration of HMs have advantages and limitations. While soil pollution indices like EF , PI , PI_{Sum} , $PI_{Average}$, PI_{Vector} , I_{geo} , PIN , PLI are based on GB, indices like C_f , MEC , C_{Deg} , RI , $MERMQ$, CSI do not consider GB and indices like BGI , mCd and ExF consider analysed soil profile, avoiding parent materials (Kowalska et al., 2018). Most popular indices for evaluation of extent of HM contamination are single-factor contaminant index (PI) and I_{geo} (Sayadi, Shabani, & Ahmadpour, 2015). However, $C_{Sample} = C_{Background}$ indicates uncontaminated soil for which $I_{geo} = \log_2[\frac{1}{1.5}] = -0.585$. Parvez et al. (2023) recommended $I_{geo} = 0$ as a class indicating practically uncontaminated, negative I_{geo} represents class showing uncontaminated to moderately contaminated and $2 < I_{geo} < 3$ represents heavily contaminated soil. In fact I_{geo} becomes zero when $C_{Sample} = (1.5) \times C_{Background}$ and $2 < I_{geo} < 3$ implies $4 < \frac{C_{Sample}}{1.5 \times C_{Background}} < 8$ i.e., $6 < \frac{C_{Sample}}{C_{Background}} < 12$.

Similarly, environmental pollution based on CF was suggested as low for ($CF < 1$), moderate ($1 \leq CF \leq 3$), considerable ($3 \leq CF \leq 6$), or high ($CF > 6$) Zhao, Feng, Yang, Niu, and Shen (2012) and $PLI = \prod_{i=1}^n CF_i$ where $PLI > 1$ implies ongoing degradation and $PLI < 1$ indicates unpolluted regions (Mohiuddin, Ogawa, Zakir, Otomo, & Shikazono, 2011). For a hypothetical case where $CF_1 = 0.2$, $CF_2 = 1.2$, $CF_3 = 3.2$, PLI becomes $0.768 < 1$, which is wrongly interpreted as just baseline levels of pollutants.

Enrichment factor (EF) includes low occurrence of variable elements (Karim, Qureshi, & Mumtaz, 2015) and computation requires enrichment level of the sample and also of the reference value (Omotoso & Ojo, 2015). Significant correlations were observed between Potential Ecological Risk (PER) index and cadmium and cobalt concentration and also between I_{geo} and PER (El-Sharkawy et al., 2023).

4. PROPOSED METHOD

Combine the basic contaminant indices $p_i = \frac{\text{Concentration of } i\text{-th HM}}{\text{Standard for the HM}}$ for the i -th HM for $i = 1, 2, \dots, n$ by multiplicative aggregation to get the index of concentration of HMs in soil for the t -th time period as

$$I_{HMSoil_t} = \prod_{i=1}^n p_{it} = \frac{C_{1t} \cdot C_{2t} \cdot C_{3t} \cdot \dots \cdot C_{nt}}{C_{10} \cdot C_{20} \cdot C_{30} \cdot \dots \cdot C_{n0}} \quad (1)$$

Where $C_{it} > 0$ denotes the average concentration of the i -th HM across samples (pooled average), and $C_{i0} > 0$ is the value of the evaluation criteria of the i -th HM. Clearly, lower values of $I_{HM_{Soil}_t}$ reflecting overall pollution status of HMs in soil at t -th time is desirable. $\frac{C_{it}}{C_{i0}} > 1$ implies concentration of i -th HM exceeds the evaluation criterion. In addition, individual HMs can be ranked on the basis of $\frac{C_{it}}{C_{i0}}$.

The index $I_{HM_{Soil}_t}$ as per Equation 1 is equivalent to the geometric mean of the ratios $\frac{C_{it}}{C_{i0}}$ i.e.

$$\sqrt[n]{\frac{C_{1t} \cdot C_{2t} \cdot C_{3t} \cdot \dots \cdot C_{nt}}{C_{10} \cdot C_{20} \cdot C_{30} \cdot \dots \cdot C_{n0}}} \text{ and can cover all types of data expressed in percentages or unit-free ratios, irrespective of their}$$

score ranges, sample sizes, correlations and probability distributions.

5. PROPERTIES AND BENEFITS

The Equation 1 implies $\log I_{HM_{Soil}_t} = \sum_{i=1}^n \log(C_{it}) - \sum_{i=1}^n \log(C_{i0})$ i.e., an additive model following lognormal distribution since $\log GM$ tends to follow lognormal distribution for large data (Alf & Grossberg, 1979). $\log(X)$ following a lognormal distribution implies X follows normal and vice versa (Motulsky, Head, & Clarke, 2025). Lognormal distributions are common when components of a composite index are combined multiplicatively (Parkin & Robinson, 1992). This facilitates testing equality of means of $I_{HM_{Soil}}$ at two regions or one region at two different years by t -tests on the logarithms of the observations. Procedure for estimation of population GM and standard error of GM can be derived since $\log(GSD \text{ of } X_1, X_2, \dots, X_n) = \text{usual SD of } (\log X_1, \log X_2, \dots, \log X_n)$. Similarly, testing equality of p_{it} and p_{jt} for $i \neq j$ or p_{it} and $p_{i(t-1)}$ can be undertaken.

- Substitutability among the component sub-indices is reduced significantly.
- Not influenced by outliers and thus produce no bias for regions (Polluted or unpolluted).
- Relative importance of each p_i can be easily found by $\frac{\Delta p_i}{\Delta I_{HM_{Soil}}} \times 100$. The HMs can be ranked with respect to $\frac{\Delta p_i}{\Delta I_{HM_{Soil}}}$ with priorities to the HMs with high ranks. Effective mitigation measures for HM with 1st rank is expected to yield maximum result in reducing $I_{HM_{Soil}}$.
- Progress or decline of the index $I_{HM_{Soil}_t}$ over the previous year can be found by;

$$\frac{I_{HM_{Soil}_t}}{I_{HM_{Soil}(t-1)}} = \frac{C_{1t} \cdot C_{2t} \cdot C_{3t} \cdot \dots \cdot C_{nt}}{C_{1(t-1)} \cdot C_{2(t-1)} \cdot \dots \cdot C_{n(t-1)}} \quad (2)$$

Where $\frac{I_{HM_{Soil}_t}}{I_{HM_{Soil}(t-1)}} > 0$ indicates progress and quantification of percentage progress is given by;

$$\frac{I_{HM_{Soil}_t} - I_{HM_{Soil}(t-1)}}{I_{HM_{Soil}(t-1)}} \times 100 \quad (3)$$

- For the proposed index, $I_{HM_{Soil}_{t0}} * I_{HM_{Soil}_{0t}} = 1 \Rightarrow$ Time reversal test and $I_{HM_{Soil}_{20}} = I_{HM_{Soil}_{21}} * I_{HM_{Soil}_{10}} \Rightarrow$ formation of chain indices
- Facilitates plotting the progress path of improvement or decline for a region over successive time periods, which in turn, helps to assess impact of various mitigation measures across time. Regions can also be compared with respect to such paths in a multivariate setup.
- The graph of progress/decline of $I_{HM_t(t-1)}$, using the Chain indices for a region helps to assess impact of various mitigation measures across time.
- Precise classification of soil contamination by HMs into finite number of mutually exclusive classes with the proposed index $I_{HM_{Soil}}$ demands finding cut-off score (X_0) such that regions with $I_{HM_{Soil}}$ scores $\leq X_0$ are unpolluted or safe and regions with $I_{HM_{Soil}}$ scores $> X_0$ are taken to be suffering from contaminations from HMs. The regions with $I_{HM_{Soil}} > X_0$ can be further categorised into a number of classes based on degree of contaminations. Davies-Bouldin Index (DBI) (Davies & Bouldin, 1979) where lower values indicate efficiency of classification. Computation of DBI for K -classes requires calculation of diameter of a class by $Diam_{C_i} =$

$$\sqrt{\frac{\sum_{x \in C_i} (x - C_i)^2}{n_i}} \text{ where } n_i: \text{ Number of regions in the } i\text{-th class and } C_i: \text{ Centroid (or mean) of the } i\text{-th class}$$

followed by $DBI_K = \frac{1}{K} \sum_{j=1}^K \text{Max}[\frac{Diam_{C_i} + Diam_{C_j}}{\|C_i - C_j\|}]$ where $i \neq j$. Maximum value of DBI_K is 1.

6. DISCUSSION

Measuring multidimensional similarities or inequalities of a set of chosen indices of soil pollution requires a composite multivariate index. The proposed approach using the concept of geometric mean facilitates assessment of multidimensional inequalities with respect to the standards. Multidimensional similarities or inequalities could also be assessed by Generalised variance (GV) proposed by Wilks (1960) as determinant of the sample variance - covariance matrix of the indices ($|S|$) giving one-dimensional measure of the total volume or "scatter" of the points in a multi-dimensional space. Lower values are desirable both for $I_{HM_{Soil}}$ and GV. However, GV approach, measuring structural, multi-dimensional spread, assumes high substitutability between dimensions and is sensitive to outliers and correlation structure.

The index $I_{HM_{Soil}}$ is simple and can be applied even for skewed data, irrespective of correlations between pairs of indices. It reflects overall pollution status of soil due to concentration of different HMs, satisfies desired properties like time reversal test, enables computation of chain indices and better comparison and classification across regions or time periods by parametric analysis. The index shows strong responsiveness to changes in soil contamination. In addition, individual HMs can be ranked based on $\frac{C_{it}}{C_{io}}$. Reduced substitutability among the component sub-indices by multiplicative aggregation makes it more appealing in benchmarking exercises (Greco, Ishizaka, Matarazzo, & Torrisi, 2018). Novelty includes knowledge of the probability distribution of the index $I_{HM_{Soil}_t}$, assessment of progress or decline over the previous years and statistical inferences.

However, collection of samples is crucial for overall assessment of HM pollution in soil influencing environmental monitoring, agricultural safety, and policymaking (Zhou et al., 2023). The proposed index could be biased if sample size is small (Feng, Wang, & Tu, 2013). Inclusion of unpolluted or less polluted sites may unnecessarily lower the value of the proposed index. Stratified sampling for selection of regions followed by simple random sampling with replacement (SRSWOR) for selection of sites is recommended for a robust sampling design with increased precision, reduced sampling error and unbiased estimation of mean and variance.

A similar approach may be adopted to evaluate separate indices for edaphic properties and xenobiont behaviour in soil and combine them to assess the health of soil in a region. Impact of $I_{HM_{Soil}}$ on vegetables, fruits, fish, shellfish, crustaceans, etc. consumed by human beings can be found by regressing the composite index of HM in soil, plants, and fish (I_{SPF}) on $I_{HM_{Soil}}$.

7. CONCLUSIONS

The proposed multiplicative aggregation with significant advantages can be computed for any finite number of HMs irrespective of sources of HMs and help in better classification of regions with assessment of efficiency of classification. Empirical investigations with real-life data may be undertaken to highlight salient features of the proposed method and assess the impact of $I_{HM_{Soil}}$ on human health risk assessment, including Carcinogenic risks.

Funding: This study received no specific financial support.

Institutional Review Board Statement: Not applicable.

Transparency: The author states that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.

Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.

REFERENCES

- Ahirvar, B. P., Das, P., Srivastava, V., & Kumar, M. (2023). Perspectives of heavy metal pollution indices for soil, sediment, and water pollution evaluation: An insight. *Total Environment Research Themes*, 6, 100039. <https://doi.org/10.1016/j.totert.2023.100039>
- Alf, E. F., & Grossberg, J. M. (1979). The geometric mean: Confidence limits and significance tests. *Perception & Psychophysics*, 26(5), 419–421. <https://doi.org/10.3758/BF03204171>
- Andem, A. B., Okorafor, K. A., Oku, E. E., & Ugwumba, A. A. (2015). Evaluation and characterization of trace metals contamination in the surface sediment using pollution load Index (PLI) and geo-accumulation index (Igeo) of ona River, Western Nigeria. *International Journal of Scientific and Technology Research*, 4(1), 29–34.
- Brahma, S., Sarmah, P., & Sarma, P. C. (2025). XRF-based assessment of topsoil heavy metal contamination and ecological risk near a cement factory in Bokajan, Assam, India. *Soil and Sediment Contamination: An International Journal*, 1–31. <https://doi.org/10.1080/15320383.2025.2599974>
- Chakrabarty, S. N. (2024). Heavy metal concentration in soil, plant products, commercial fish: Composite index. *Journal of Energy and Environmental Sciences*, 2(2), 1–10.
- Chakrabarty, S. N. (2025). Comprehensive index for pollution status of heavy metals in soil. *Interdisciplinary Environmental Review*, 24(4), 309–322. <https://doi.org/10.1504/IER.2025.150086>
- Chakravarty, M., & Patgiri, A. D. (2009). Metal pollution assessment in sediments of the Dikrong river, NE India. *Journal of Human Ecology*, 27(1), 63–67.
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Duskaev, G., Miroshnikov, S., Sizova, E., Lebedev, S., & Notova, S. (2014). The influence of heavy metals on the organism of animals and the environment. *Animal Husbandry and Forage Production*, 3(86), 7–11.
- Dutta, N., Dutta, S., Bhupenchandra, I., Karmakar, R., Das, K., Singh, L. K., . . . Sarmah, T. (2021). Assessment of heavy metal status and identification of source in soils under intensive vegetable growing areas of Brahmaputra valley, North East India. *Environmental Monitoring and Assessment*, 193(6), 376. <https://doi.org/10.1007/s10661-021-09168-x>
- El-Sharkawy, M., Li, J., Kamal, N., Mahmoud, E., Omara, A. E.-D., & Du, D. (2023). Assessing and predicting soil quality in heavy metal-contaminated soils: Statistical and ANN-based techniques. *Journal of Soil Science and Plant Nutrition*, 23(4), 6510–6526. <https://doi.org/10.1007/s42729-023-01507-w>
- Feng, C., Wang, H., & Tu, X. M. (2013). Geometric mean of nonnegative random variable. *Communications in Statistics-Theory and Methods*, 42(15), 2714–2717. <https://doi.org/10.1080/03610926.2011.615637>
- Gong, C., Wang, S., Wang, D., Lu, H., Dong, H., Liu, J., . . . Wang, L. (2022). Ecological and human health risk assessment of heavy metal (loid) s in agricultural soil in hotbed chives hometown of Tangchang, Southwest China. *Scientific Reports*, 12(1), 8563. <https://doi.org/10.1038/s41598-022-11397-0>
- Greco, S., Ishizaka, A., Matarazzo, B., & Torrisi, G. (2018). Stochastic multi-attribute acceptability analysis (SMAA): An application to the ranking of Italian regions. *Regional Studies*, 52(4), 585–600.
- Ikenaka, Y., Nakayama, S. M., Muzandu, K., Choongo, K., Teraoka, H., Mizuno, N., & Ishizuka, M. (2010). Heavy metal contamination of soil and sediment in Zambia. *African Journal of Environmental Science and Technology*, 4(11), 729–739.
- Kantemiris, G., Xenogiannopoulou, E., Vallas, A., & Oikonomou, P. (2025). Classification of soil contamination by heavy metals (Cr, Ni, Pb, Zn) in wildfire-affected areas using laser-induced breakdown spectroscopy and machine learning. *Environmental Science and Pollution Research*, 32(5), 2359–2373. <https://doi.org/10.1007/s11356-024-35825-y>
- Karim, Z., Qureshi, B. A., & Mumtaz, M. (2015). Geochemical baseline determination and pollution assessment of heavy metals in urban soils of Karachi, Pakistan. *Ecological Indicators*, 48, 358–364. <https://doi.org/10.1016/j.ecolind.2014.08.032>
- Kowalska, J., Mazurek, R., Gąsiorek, M., Setlak, M., Zaleski, T., & Waroszewski, J. (2016). Soil pollution indices conditioned by medieval metallurgical activity—A case study from Krakow (Poland). *Environmental Pollution*, 218, 1023–1036. <https://doi.org/10.1016/j.envpol.2016.08.053>

- Kowalska, J. B., Mazurek, R., Gąsiorek, M., & Zaleski, T. (2018). Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination—A review. *Environmental Geochemistry and Health*, 40(6), 2395–2420. <https://doi.org/10.1007/s10653-018-0106-z>
- Li, Y.-M., Ma, J.-H., Liu, D.-X., Sun, Y.-L., & Chen, Y.-F. (2015). Assessment of heavy metal pollution and potential ecological risks of urban soils in Kaifeng City. *Environmental Science*, 36(3), 1037–1044.
- Mariani, F., & Ciommi, M. (2022). Aggregating composite indicators through the geometric mean: A penalization approach. *Computation*, 10(4), 64. <https://doi.org/10.3390/computation10040064>
- Mazurek, R., Kowalska, J., Gąsiorek, M., Zadrozny, P., Józefowska, A., Zaleski, T., . . . Orłowska, K. (2017). Assessment of heavy metals contamination in surface layers of Roztocze National Park forest soils (SE Poland) by indices of pollution. *Chemosphere*, 168, 839–850. <https://doi.org/10.1016/j.chemosphere.2016.10.126>
- Mohiuddin, K., Ogawa, Y., Zakir, H., Otomo, K., & Shikazono, N. (2011). Heavy metals contamination in water and sediments of an urban river in a developing country. *International Journal of Environmental Science & Technology*, 8(4), 723–736. <https://doi.org/10.1007/BF03326257>
- Motulsky, H. J., Head, T., & Clarke, P. B. (2025). Analyzing lognormal data: A nonmathematical practical guide. *Pharmacological Reviews*, 77(3), 100049. <https://doi.org/10.1016/j.pharmr.2025.100049>
- Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification*, 31(3), 274–295. <https://doi.org/10.1007/s00357-014-9161-z>
- Nogoibaev, M., Nogoibaeva, R., Omurbaeva, A., Akunova, S., & Kyrbasheva, M. (2025). Climate change in Kyrgyzstan and its impact on the emergence of biogeocoenotic pathologies in animals. *Bulletin of the Kyrgyz National Agrarian University*, 23(1), 70–78. <https://doi.org/10.63621/bknau.1.2025.70>
- Omotoso, O., & Ojo, O. (2015). Assessment of some heavy metals contamination in the soil of river Niger floodplain at Jebba, central Nigeria. *Water Utility Journal*, 9, 71–80.
- Parkin, T. B., & Robinson, J. A. (1992). Analysis of lognormal data. In B. A. Stewart (Ed.), *Advances in Soil Science*. In (Vol. 20, pp. 193–235). New York: Springer-Verlag.
- Parvez, M. S., Nawshin, S., Sultana, S., Hossain, M. S., Khan, R. M. H., Habib, M. A., . . . Khan, R. (2023). Evaluation of heavy metal contamination in soil samples around Rampal, Bangladesh. *ACS Omega*, 8(18), 15990–15999. <https://doi.org/10.1021/acsomega.2c07681>
- Pourret, O., & Hursthouse, A. (2019). It's time to replace the term “heavy metals” with “potentially toxic elements” when reporting environmental research. *International Journal of Environmental Research and Public Health*, 16(22), 4446. <https://doi.org/10.3390/ijerph16224446>
- Qingjie, G., Jun, D., Yunchuan, X., Qingfei, W., & Liqiang, Y. (2008). Calculating pollution indices by heavy metals in ecological geochemistry assessment and a case study in parks of Beijing. *Journal of China University of Geosciences*, 19(3), 230–241. [https://doi.org/10.1016/S1002-0705\(08\)60042-4](https://doi.org/10.1016/S1002-0705(08)60042-4)
- Salikhov, D., Petrov, S., & Shuralev, E. (2024). Assessment of heavy metal contamination of soils in the Republic of Tatarstan using complex pollution indices NPI, RI, BGI. *E3S Web of Conferences*, 539, 01026. <https://doi.org/10.1051/e3sconf/202453901026>
- Santos-Francés, F., Martínez-Graña, A., Alonso Rojo, P., & García Sánchez, A. (2017). Geochemical background and baseline values determination and spatial distribution of heavy metal pollution in soils of the Andes mountain range (Cajamarca-Huancavelica, Peru). *International Journal of Environmental Research and Public Health*, 14(8), 859. <https://doi.org/10.3390/ijerph14080859>
- Sawadogo, Y. Z., Bambara, T. L., Derra, M., Zongo, I., Kaboré, K., & Zougmore, F. (2023). Evaluation of heavy metal pollution and physico-chemical parameters in agricultural soils of Bouly, Burkina-Faso. *Journal of Materials Physics and Chemistry*, 11(2), 38–47.
- Sayadi, M. H., Shabani, M., & Ahmadvpour, N. (2015). Pollution index and ecological risk of heavy metals in the surface soils of Amir-Abad Area in Birjand City, Iran. *Health Scope*, 4(1), e21137. <https://doi.org/10.17795/jhealthscope-21137>

- Skugoreva, S. G., Ashikhmina, T. Y., Fokina, A. I., & Lyalina, E. I. (2016). Chemical bases of toxic action of heavy metals (review). *Theoretical and Applied Ecology*, 1, 4–13.
- Tomlinson, D. L., Wilson, J. G., Harris, C., & Jeffrey, D. (1980). Problems in the assessment of heavy-metal levels in estuaries and the formation of a pollution index. *Helgoländer Meeresuntersuchungen*, 33(1), 566-575. <https://doi.org/10.1007/BF02414780>
- United States Environmental Protection Agency. (2011). *Exposure factors handbook 2011 edition (Final); Office of emergency and remedial response*. Washington, DC: US Environmental Protection Agency.
- Wang, F., Wang, F., Yang, H., Yu, J., & Ni, R. (2023). Ecological risk assessment based on soil adsorption capacity for heavy metals in Taihu basin, China. *Environmental Pollution*, 316, 120608. <https://doi.org/10.1016/j.envpol.2022.120608>
- Wilks, S. S. (1960). Multidimensional statistical scatter. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling*. In (pp. 486–503). Stanford, CA: Stanford University Press.
- Xie, W., Yang, Y. D., Hou, J. Y., Kan, G. Q., Li, G. C., & Zhao, X. H. (2021). Application of multiple evaluation methods to soil environmental quality evaluation of main walnut producing areas in Tianjin. *Geophysics and Geochemical Exploration*, 45(01), 207-214.
- Yang, D., Zhu, H., Liu, J., Zhang, Y., Wu, S., Xiong, J., & Wang, F. (2023). Risk assessment of heavy metals in soils from four different industrial plants in a medium-sized city in North China. *Toxics*, 11(3), 217. <https://doi.org/10.3390/toxics11030217>
- Yoo, J.-C., Beiyuan, J., Wang, L., Tsang, D. C. W., Baek, K., Bolan, N. S., . . . Li, X.-D. (2018). A combination of ferric nitrate/EDDS-enhanced washing and sludge-derived biochar stabilization of metal-contaminated soils. *Science of the Total Environment*, 616, 572-582. <https://doi.org/10.1016/j.scitotenv.2017.10.310>
- Zhang, J., & Liu, C. (2002). Riverine composition and estuarine geochemistry of particulate metals in China—weathering features, anthropogenic impact and chemical fluxes. *Estuarine, Coastal and Shelf Science*, 54(6), 1051-1070. <https://doi.org/10.1006/ecss.2001.0879>
- Zhao, H., Wu, Y., Lan, X., Yang, Y., Wu, X., & Du, L. (2022). Comprehensive assessment of harmful heavy metals in contaminated soil in order to score pollution level. *Scientific Reports*, 12(1), 3552. <https://doi.org/10.1038/s41598-022-07602-9>
- Zhao, S., Feng, C., Yang, Y., Niu, J., & Shen, Z. (2012). Risk assessment of sedimentary metals in the Yangtze Estuary: New evidence of the relationships between two typical index methods. *Journal of Hazardous Materials*, 241, 164-172. <https://doi.org/10.1016/j.jhazmat.2012.09.023>
- Zhou, H., Chen, Y., Yue, X., Ren, D., Liu, Y., & Yang, K. (2023). Identification and hazard analysis of heavy metal sources in agricultural soils in ancient mining areas: A quantitative method based on the receptor model and risk assessment. *Journal of Hazardous Materials*, 445, 130528. <https://doi.org/10.1016/j.jhazmat.2022.130528>