Chemosphere 366 (2024) 143472

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Evaluation of the geochemical background of soil in a hyper-arid zone using a multivariate statistical methodology: The case of the city of Antofagasta in the Atacama Desert

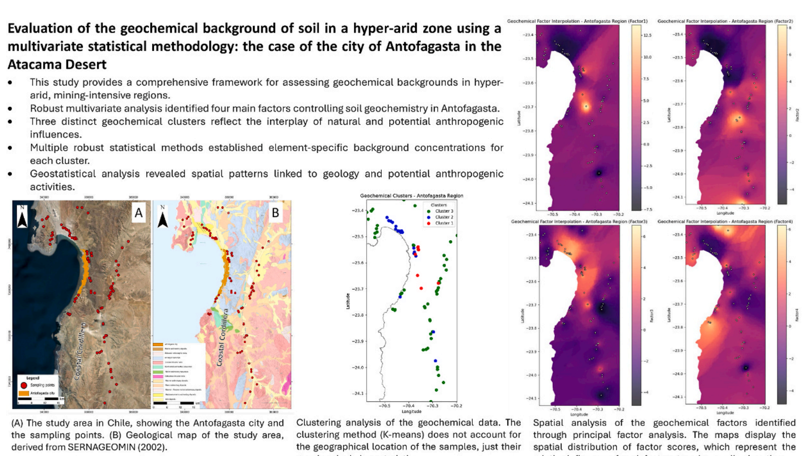
Brian F. Keith [a], Elizabeth J. Lam [b,*], Ítalo L. Montofré [c,d], Vicente Zetola [e], Javier Urrutia [f], Christian Herrera [g], Jaume Bech [h]

[a] Department of Computing & Systems Engineering, Universidad Católica del Norte, Antofagasta, 1270709, Chile
[b] Chemical Engineering Department, Universidad Católica del Norte, Antofagasta, 1270709, Chile
[c] Mining and Metallurgical Engineering Department, Universidad Católica del Norte, Antofagasta, 1270709, Chile
[d] Mining Business School, ENM, Universidad Católica del Norte, Antofagasta, Chile
[e] Construction Management Department, Universidad Católica del Norte, 1270709, Antofagasta, Chile
[f] Center for Research and Development of Water Ecosystems, Universidad Bernardo O'Higgins, Santiago, Chile
[g] Departamento de Ciencias Geológicas, Universidad Católica del Norte, Antofagasta, 1270709, Chile
[h] Soil Science Laboratory, Faculty of Biology, Universidad de Barcelona, Barcelona, Spain

HIGHLIGHTS

- Study provides a comprehensive framework for assessing geochemical backgrounds in hyper-arid, mining-intensive regions.
- Robust multivariate analysis identified four main factors controlling soil geochemistry in Antofagasta.
- Three distinct geochemical clusters reflect the interplay of natural and potential anthropogenic influences.
- Multiple robust statistical methods established element-specific background concentrations for each cluster.
- Geostatistical analysis revealed spatial patterns linked to geology and potential anthropogenic activities.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling editor: Petra Petra Krystek

Keywords:
Geochemical background concentration

ABSTRACT

This study investigates the geochemical background and factors influencing the variability of 19 environmentally relevant elements in the soils of Antofagasta, Chile, a region known for its extensive mining activities. Employing robust multivariate statistical techniques on a dataset of 94 soil samples, we identified four main factors

* Corresponding author.
E-mail addresses: brian.keith@ucn.cl (B.F. Keith), elam@ucn.cl (E.J. Lam), imontofre@ucn.cl (Í.L. Montofré), vzetola@ucn.cl (V. Zetola), javier.urrutia@ucn.cl (J. Urrutia), cherrera@ucn.cl (C. Herrera), jaumebechborras@gmail.com (J. Bech).

https://doi.org/10.1016/j.chemosphere.2024.143472
Received 16 June 2024; Received in revised form 1 October 2024; Accepted 3 October 2024
Available online 4 October 2024

Mining industries
Arid zone
Environmental risks
Cluster analysis
Geostatistics

explaining 70% of the total variance in elemental concentrations. These four factors reflect the influence of Jurassic volcanic rocks, intrusive rocks, marine sediments, and mafic to intermediate intrusive rocks. Cluster analysis revealed three distinct geochemical populations, each reflecting a unique combination of natural and anthropogenic influences. We established background concentrations for each element within these clusters using robust statistical methods. Geostatistical analysis, employing inverse distance weighted interpolation, produced factor distribution maps that, when integrated with geological data, provided insights into lithological and anthropogenic influences on soil geochemistry. Our findings highlight the complex interplay between natural geological processes, the region's unique arid climate, and anthropogenic activities in shaping the geochemical landscape of Antofagasta. This study contributes to the understanding of geochemical backgrounds in mining-intensive, arid regions and provides a methodological framework applicable to similar environments worldwide.

1. Introduction

The mining industry has been a key driver of economic growth in Chile, with the country being a global leader in the production of Ag, Au, B, Cu, I, Li, Mo, and Re (Lam et al., 2016; Smart, 2020). The Antofagasta region, located in the northern part of Chile, is particularly known for its extensive copper mining activities. While the mining sector has significantly contributed to the economic development of the region, it has also raised concerns about its environmental impact, especially on soil quality (Castro and Sánchez, 2003; Lagos and Blanco, 2010; Lam et al., 2018).

The city of Antofagasta is located in the Cordillera de la Costa of the Antofagasta Region. This area corresponds to one of the most arid parts of the Atacama Desert. The average precipitation ranges between 0.5 and 8 mm/year, and can even go without precipitation for 4 or 5 consecutive years and only a few sporadic short storms. Temperatures on the coast are mild and stable most of the year due to the softening effect of the sea. The average monthly values range between 13.5 °C and 20.1 °C in the city of Antofagasta (DGA, 2009). The extreme dryness of the Cordillera de la Costa of the Antofagasta Region is due to the permanent anticyclone conditions on the coasts of northern Chile and the cold Humboldt current (Klohn, 1972; Rutllant, 1985). On the other hand, the Andes Mountains are an impressive mountain barrier that prevents the arrival of humidity from the Atlantic Ocean through the Amazon Basin (Garreaud, 2009). Part of this humidity reaches the western slope of the Andes in the summer months, in the so-called "Altiplano Winter", but not the Central Depression.

The Antofagasta region, characterized by its hyperarid to arid climate, is notorious for its anomalous concentrations of As and other elements present in its water, soil, and rock (Pérez-Carrera and Fernández Cirelli, 2010). These geochemical anomalies are the result of geological processes that have given rise to important mineral deposits in the area (Oyarzun et al., 2006). The soils of Antofagasta show high contents of metals such as Cu, Mo, As, and other elements associated with mineralization (Lam Esquenazi et al., 2019). The high natural geochemical concentrations found in this region strongly contrast with the background geochemical levels of other non-mining regions in the country. Therefore, Antofagasta is considered a highly anomalous region from the perspective of environmental geochemistry. The detailed study of these anomalies is crucial for differentiating the natural geochemistry from possible anthropogenic impacts on the soils (Oyarzun et al., 2006; Tapia et al., 2018).

The intensive mining operations in this region have led to an increased risk of soil contamination, which can have detrimental effects on the ecosystem and human health (Castro and Sánchez, 2003; Lam et al., 2018). To assess the extent of soil contamination and develop effective management strategies, it is crucial to establish the geochemical background values of metals and metalloids in the region (Lam et al., 2020).

Geochemical background values play a crucial role in environmental assessments, particularly in distinguishing between natural and anthropogenic sources of elements in the environment. Determining background concentrations of various metals in northern Chile is important due to the region's unique hyper-arid conditions, particularly in the Atacama Desert. The scarcity of precipitation and limited natural water sources results in metal concentrations in soils behaving differently than in more temperate regions (Arens et al., 2021). Unlike areas with regular rainfall and runoff, where metals can be redistributed over time, the Atacama's hyper-arid core experiences minimal change, leading to potential accumulation of metals on the surface over extended periods. Establishing background levels in such an environment requires accounting for high variability across different locations, as the geochemistry of one site may not accurately represent another. This complexity is further compounded by wind-driven metal transport, which can impact local measurements.

The Antofagasta region, part of the Atacama Desert, is one of the driest places on Earth. Its hyper-arid conditions lead to minimal natural weathering processes and slower dispersal of contaminants. Over a century of mining activity has resulted in large volumes of waste containing potentially toxic elements (PTEs) such as arsenic, copper, lead, zinc, and cadmium (Reyes et al., 2020, 2021). These contaminants, often present in tailings and abandoned sites, have accumulated in surrounding soils. Differentiating between anthropogenic contamination and naturally occurring (geogenic) metal concentrations poses a significant challenge. The reduced precipitation in the region makes it difficult to establish a clear baseline, as natural geochemical processes that would otherwise distribute or dilute these metals are minimal. This leads to heightened concerns over human exposure, particularly in areas with expanding urban development (Reyes et al., 2020; Lam et al., 2023).

The slow breakdown of waste material in this arid environment contributes to prolonged exposure risks, complicating efforts to assess and remediate polluted sites (Tyler, 2020). Establishing accurate background levels is important for identifying the extent of contamination and developing regulatory standards to protect human health and the environment. The presence of communities living near contaminated soils further amplifies these environmental and health risks (García et al., 2024).

The concept of geochemical background, however, has varying definitions in the literature. Traditionally, as proposed by Matschullat et al. (2000), these values represent the natural concentration of elements in the soil, unaffected by anthropogenic activities. In contrast, the ISO 19258 standard defines "background content" as the "content of a substance in a soil resulting from both natural geological and pedological processes and including diffuse source inputs" (ISO-19258, 2018). This definition notably includes diffuse source inputs, which may encompass some anthropogenic influences.

The distinction between these definitions is significant when interpreting and comparing background values across different studies and regulatory frameworks. For the purposes of this study, we adhere to the ISO 19258 standard definition, recognizing that in many areas, especially those with long histories of human habitation and land use, it may be challenging to find soils completely unaffected by human activities. This approach allows for a more practical assessment of soil conditions in the context of real-world environments where diffuse anthropogenic inputs may be present.

Various methods, such as direct, statistical, and integrated approaches, can be employed to estimate these background values (Gałuszka, 2007; Lam et al., 2020). The choice of method and definition can significantly impact the resulting background values and their interpretation. In the context of Chile, the establishment of accurate geochemical background values is particularly crucial. While Chile has made progress in soil quality regulation with the introduction of NCh3628 in 2019, "Soil quality - Sampling - Part 1: Guidance on the design of sampling programmes", the country still lacks a comprehensive environmental quality standard for soils that establishes specific contaminant thresholds (Lam et al., 2023). The NCh3628 standard provides guidance on soil sampling procedures, which is a fundamental step towards soil quality assessment. However, it does not set specific limit values for soil contaminants or define background levels. This gap underscores the importance of establishing reliable geochemical background values to complement the existing sampling guidelines. These values are essential to support the development of appropriate soil quality guidelines and to accurately assess the environmental impact of mining activities in the region. By providing a baseline against which to measure potential contamination, geochemical background values serve as a fundamental tool in environmental management and policy-making, especially in areas with significant mining activities like the Antofagasta region. The integration of robust sampling procedures (as outlined in NCh3628) with well-established background values would significantly enhance Chile's capacity for comprehensive soil quality assessment and management.

Despite the crucial role of geochemical background values in environmental management, there is a scarcity of information regarding the geochemical background of the Antofagasta Region. This paucity of data impedes the accurate assessment of soil contamination levels and the identification of pollution sources, which are vital for developing effective environmental management strategies (Lam et al., 2020). However, a valuable resource is available in the form of a database compiled by the Antofagasta Regional Government (Project BIP 30359875-0, 2019) for the assessment of regional environmental risks, which was based on a previous evaluation by CENMA (2014). In particular, the database contains geochemical background levels (GBL) for 32 analytes measured in surface samples (0–20 cm) from pristine sites in the region. This study focused on the determination of metals and metalloids, considering that mining is the predominant economic activity in the Antofagasta region and the primary source of metal concentrations in soils. The availability of this database provides a foundation for assessing the natural geochemical background and distinguishing it from anthropogenic influences, which is essential for implementing targeted environmental management measures in this mining-intensive area. To address this knowledge gap, the present study focuses on two main objectives.

1. Determine the background concentrations of environmentally relevant elements in the soils around the city of Antofagasta, Chile.
2. Identify the source structures of elements to differentiate between natural and anthropogenic sources around the city of Antofagasta, Chile.

By establishing the geochemical background values, this research aims to contribute to the development of effective environmental management strategies and support the ongoing efforts to establish soil quality standards in Chile. The findings of this study will provide valuable insights into the natural and anthropogenic factors influencing the elemental composition of soils in the Antofagasta region by determining and facilitating a better understanding of the environmental impact of mining activities in the area. Furthermore, the results of this research will serve as a foundation for future studies on soil contamination and environmental management in the region, ultimately promoting sustainable mining practices and protecting the environment and human health.

2. Materials and methods

2.1. Area of study and geological background

The study area corresponds to the Antofagasta commune in the Antofagasta region, Chile (23° 39′ 0″ South, 70° 24′ 0″ West) (Fig. 1A). Note that communes in Chile correspond to the smallest administrative subdivision of Chile (Biskupovic, 2015). In this case, the commune includes the city itself and its surroundings. The region has a total population of 607,534 inhabitants and covers an area of 126,049 km^2 (National Institute of Statistics, 2018). The Antofagasta region is characterized by the development of resource exploitation operations and mining activity. Human occupation and intervention in various cities of the region have transformed the soil matrix from its natural origin. This results in high soil heterogeneity, where zones with varied levels of intervention in metal concentrations could be identified. Thus, the potential for soil contamination is facilitated, which can pose hazardous effects on human health and the ecosystem.

The Cordillera de la Costa (Fig. 1) represents the Upper Jurassic–Lower Upper Cretaceous Volcanic Arc. This period's volcanic activity is predominantly represented by the La Negra Formation, which reaches a maximum thickness of 8 km (González and Niemeyer, 2005). The plutonic rocks in the 18°S to 25°S segment exhibit a broad compositional range, from gabbros to granites (Scheuber and Gonzalez, 1999). This volcanic formation is intruded by several dikes, including basaltic, andesitic, and dacitic types. The volcanic rocks mainly consist of andesitic and basaltic andesite, interspersed with lesser amounts of continental and marine volcaniclastics (Charrier et al., 2007, 2015) and underlain by silica-rich pyroclastic rocks. Most of these plutonic rocks and dikes display calc-alkaline composition (Oliveros et al., 2007). The volcanic rocks were affected by different hydrothermal events during the Jurassic and Cretaceous, which involved seawater (Kojima et al., 2009). It should also be noted that within the intrusive bodies of the Middle-Upper Jurassic, there are vein-type deposits that contain copper, gold, and silver (Boric et al., 1990).

The Atacama Fault System in northern Chile, extending about 1000 km from 20°S to 30°S (González et al., 2006), controls pull-apart basins with Oligocene–Miocene sediments overlying Mesozoic rocks and Plio–Quaternary sediments (Kay et al., 1991; Muñoz and Stern, 1989). It has formed several fore-arc basins in the northern Cordillera de la Costa, notably at Salar Grande (20.5° - 20.8°S), which contains mainly halite saline deposits.

The sediments adjacent to the Cordillera de la Costa formed from various geological periods and resulted from diverse geological processes. These sediments include Quaternary colluvial, alluvial, eolian, and littoral deposits; Neogene marine, littoral, alluvial, colluvial, and eolian deposits; Lower Cretaceous continental sedimentary rocks, mainly composed of carbonate minerals; and Jurassic volcanic and intrusive rocks (Fig. 1B). The sedimentary sequence is predominantly composed of gravel supported by a matrix consisting primarily of medium-grained sandstone. The gravel clasts are typically composed of andesitic volcanic rocks and contain fragments of marine shells. This gravel is interbedded with thin layers of sandstone, siltstone, and clay, varying in spatial distribution throughout the sequence. At the southern area of Cerro Morro Moreno, metadioritic intrusive rocks are found, containing small ultramafic pyroxenite inclusions. These inclusions, covering less than 1 km^2 (González and Niemeyer, 2005), are too small to be represented at the scale of Fig. 1.

2.2. Sampling methodology, geochemical analysis and quality control

The geochemical data used in this study were obtained from a comprehensive sampling campaign conducted by the Chilean Ministry of the Environment in collaboration with the Regional Government of Antofagasta (Project BIP 30359875-0, 2019). This database encompasses a total of 303 soil samples, of which a total of 94 soil samples were

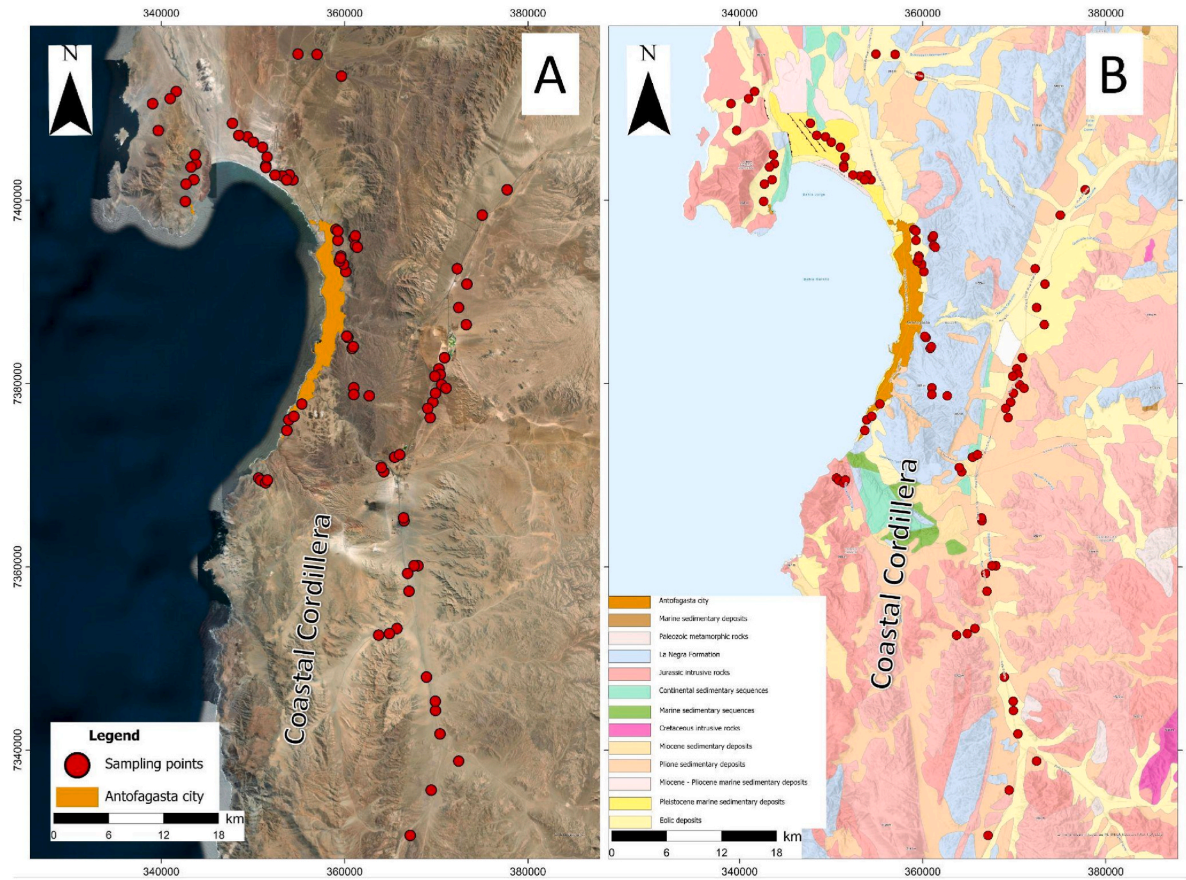


Fig. 1. (A) The study area in Chile, showing the Antofagasta city and the sampling points. (B) Geological map of the study area, derived from Servicio Nacional de Geología y MineríaSERNAGEOMIN (2002).

collected from strategically selected locations within the Antofagasta commune, aiming to represent the diverse geological units present in the area.

The selection of sampling sites was guided by the Geological Map of Chile (scale 1:1,000,000) provided by the National Geology and Mining Service (González and Niemeyer, 2005) and detailed geological maps at scales ranging from 1:100,000 to 1:250,000, available on the SERNAGEOMIN website. The aforementioned samples from the Antofagasta commune, which are the focus of this study, contain geochemical background levels (GBL) for 32 analytes, including essential elements and potentially toxic metals, measured in surface soil samples (0–20 cm) collected from pristine sites.

The chemical analysis of the soil samples was performed using Inductively Coupled Plasma Mass Spectrometry (ICP-MS). The analyses were conducted by a certified laboratory that adheres to a comprehensive set of quality assurance and quality control (QA/QC) protocols, including the use of certified reference materials (CRMs), matrix spikes, and method blanks, as well as appropriate instrument calibration and performance monitoring procedures. The original reference (Project BIP 30359875-0, 2019) contains annexes with the reports associated with each sample.

The suite of 32 chemical elements analyzed in this study encompasses a wide range of geochemically and environmentally significant parameters, including aluminum (Al), antimony (Sb), arsenic (As), barium (Ba), beryllium (Be), bismuth (Bi), boron (B), cadmium (Cd), calcium (Ca), chromium (Cr), cobalt (Co), copper (Cu), iron (Fe), lead (Pb), lithium (Li), magnesium (Mg), manganese (Mn), mercury (Hg), molybdenum (Mo), nickel (Ni), phosphorus (P), potassium (K), selenium (Se), silver (Ag), sodium (Na), strontium (Sr), thallium (Tl), tin (Sn), titanium (Ti), uranium (U), vanadium (V), and zinc (Zn).

By leveraging the geochemical data from the database, specifically the 94 samples from the Antofagasta commune, this study establishes a robust and locally relevant baseline for assessing soil geochemistry and potential contamination in the study area. The use of this dataset ensures that future evaluations of soil contamination levels and the delineation of natural versus anthropogenic elemental sources are grounded in reliable data that accurately capture the geochemical conditions of the area under study.

2.3. *Data analysis methodology*

The data analysis methodology employed in this study consists of four main components: Exploratory Data Analysis, Robust Data Analysis, Clustering Analysis, and Geostatistical Analysis.

2.3.1. *Exploratory Data Analysis*

Exploratory Data Analysis (Zuur et al., 2010) refers to the process of conducting preliminary investigations to extract general information from the data, discover patterns, and identify anomalies such as multimodality (e.g., using scatter plots, histograms, and box plots, among others), and validate previous assumptions about the data population using statistical summaries and graphs.

This phase also includes the process of cleaning the data set and removing samples that have issues. In particular, considering the nature of geochemical data, values below the detection limits (BDL) were inspected and excluded following a conservative criterion to preserve as much data as possible (Reimann et al., 2005). Thus, elements whose data proportion exceeds 55% have been excluded from subsequent stages of the study.

2.3.2. Robust Data Analysis

Robust Data Analysis (Filzmoser et al., 2009) techniques were employed to address the limitations of standard statistical methods when applied directly to raw geochemical data, which can lead to imprecise or misleading results, hindering the accurate interpretation of geochemical associations (Filzmoser et al., 2009). To overcome this issue, various data transformations were utilized to approximate normality and ensure homogeneity of variance (Reimann et al., 2002). These transformations are part of a family of log-ratio transformations, which include the additive log-ratio, centered log-ratio, and isometric log-ratio transformations (Aitchison, 1982).

Among these transformations, the isometric log-ratio transformation (Egozcue et al., 2003) was selected for this study due to its desirable properties and robustness in the multivariate analysis of geochemical data containing outliers (Reimann et al., 2011). The isometric log-ratio transformation is an isomorphism that preserves the Euclidean geometry of the sample space, enabling the application of standard multivariate methods to the transformed data. Thus, in this study, the isometric log-ratio transformation was chosen as it yielded more stable results compared to other log-ratio transformations. The transformed data served as the basis for subsequent statistical analyses, ensuring a more accurate and reliable interpretation of geochemical associations in the presence of compositional constraints and outliers.

As geochemical compositions presumably always contain outliers and closure effects (Liu et al., 2016), the isometric log-ratio transformation was applied to reveal the compositional structure of the database. Transformations are employed to open compositional databases (Egozcue et al., 2003). This transformation is widely accepted by various researchers. Subsequently, Robust Factor Analysis was applied to reduce dimensionality and identify structures between factors, along with identifying sources of groups of elements that can be relatively easily interpreted (Reimann et al., 2002). The term "robust" in this case refers to a group of techniques tolerant to the presence of outliers in the sampled dataset. This is done based on recommendations suggested by Reimann et al., 2014.

Thus, moving forward, multivariate outlier values are detected in the main dataset for the transformed data. The detection of potential multivariate outliers is carried out using the robust Mahalanobis distance (Maronna & Zamar, 2002; Olive, 2004; Rousseeuw and Driessen, 1999). Two R software packages were utilized for these statistical analyses: "Compositions" (Van Den Boogaart et al., 2011) and "RobCompositions" (Templ et al., 2011).

In particular, three robust methods were employed to detect outliers: the Olive-Hawkins RMVN method, the OGK method, and the Fast MCD method. For each method, the robust Mahalanobis distance was calculated using the *covrob(.)*, *covOGK(.)*, and *cov.mcd(.)* functions from the *rrcov* package, respectively. Data points with a robust Mahalanobis distance greater than a specified threshold, defined as the square root of the 97.5% quantile of the chi-squared distribution with degrees of freedom equal to the number of variables, were considered outliers.

The outliers identified by each method were stored in separate vectors, and the final set of outliers was obtained by taking the intersection of the outliers identified by all three methods. This conservative approach ensured that only data points consistently classified as outliers by all three methods were considered true anomalies. The identified outliers were then removed from the dataset, and the remaining data points were used for further analysis. By combining multiple robust methods and considering only the intersection of their results, the risk of removing valid data points was minimized while ensuring that true anomalies were detected and excluded from the study.

2.3.3. Clustering analysis

Clustering Analysis (Templ et al., 2008) was conducted to classify similar groups in the geochemical database. The first step in the clustering analysis process was to determine the optimal number of clusters. Based on this optimal number of clusters, the final clustering of the data (excluding anomalous points) was obtained using the K-means algorithm.

In particular, the optimal number of clusters for the geochemical dataset was determined using the *NbClust* package in R (Charrad et al., 2014). This package provides 26 indices for determining the optimal number of clusters and proposes the best clustering scheme from the different results obtained by varying the number of clusters, the clustering method, and the distance measures. The *NbClust* function was applied to the transformed data, with a minimum of 2 and a maximum of 10 clusters, using the k-means clustering method and Euclidean distance. The majority rule was used to determine the optimal number of clusters, which was found to be 3.

To further validate the optimal number of clusters, the gap statistic method (Tibshirani et al., 2001). This method compares the total within-cluster variation for different values of k (number of clusters) with their expected values under a null reference distribution of the data. The optimal number of clusters is determined by the smallest value of k that yields a gap statistic greater than the gap statistic for $k+1$ minus the standard error. The gap statistic also suggested 3 as the optimal number of clusters for the geochemical dataset.

Next, K-means clustering (MacQueen, 1967) was then performed on the transformed data using the *kmeans* function in R, with the number of clusters set to 3, a maximum of 20 iterations, and 100 random sets to ensure stable results. As mentioned before, the anomalous data points identified in the previous step were not included in the clustering process. The resulting clusters were designated as Cluster 1, Cluster 2, and Cluster 3. Then, the average concentrations of the elements in each cluster and the anomalous data points (considered as a separate "noise" cluster) were calculated. These average concentrations provide insights into the geochemical characteristics of each cluster and the anomalous data points (Table 2).

Finally, we compute three potential background values based on different criteria for identifying samples with unusually high concentrations of elements (Desaules, 2012; Redon et al., 2013). The first method, known as the "MAD" method, calculates the median + 2MAD (median absolute deviation) (Reimann et al., 2005; Teng et al., 2010; Tume et al., 2018). In the second method, the upper whisker of a Tukey boxplot is used, and this value is calculated by the third quartile +1.5 IQR (interquartile range) (Tume et al., 2019; McIlwaine et al., 2014). The third method involves using the 95th percentile as a threshold (Ander et al., 2013; Johnson et al., 2012). These methods provide us with robust estimates of the background values for the identified geochemical clusters.

2.3.4. Geostatistical analysis

Geostatistical Analysis (Kumar and Sinha, 2018) was employed to analyze the spatial distribution of the geochemical factors identified through principal factor analysis. This process involved creating interpolated surfaces using the Inverse Distance Weighting (IDW) method to visualize the spatial patterns of each factor across the Antofagasta commune and identify both the structure and geochemical anomalies (Khadka, 2023).

The IDW interpolation was performed using the KDTree method from the *scipy* library in Python (Virtanen et al., 2020). A grid of points covering the extent of the Antofagasta region was created, and the factor values were interpolated at each grid point using the IDW algorithm. The number of nearest neighbors (k) and the power parameter (p) were set to 10 and 2, respectively, to control the influence of nearby points on the interpolated values.

The resulting interpolated surfaces were then clipped to the boundaries of the Antofagasta commune using the *geopandas* library (Shreemathi et al., 2024) and the provided shapefile of the region. This ensured that the geostatistical analysis focused specifically on the area of interest. The clipped interpolated surfaces were visualized using the *matplotlib* library (Hunter, 2007), with each factor plotted separately. The plots included the Antofagasta region boundary, the interpolated surface displayed as a heatmap, and the original sampling points overlaid on the map. A color bar was added to represent the range of factor values.

These geostatistical maps provided valuable insights into the spatial distribution of the geochemical factors within the Antofagasta commune. They allowed for the identification of areas with higher or lower factor values, potentially indicating variations in the underlying geological or anthropogenic processes influencing the geochemical composition of the soils.

By combining the results of the principal factor analysis with the geostatistical maps, a comprehensive understanding of the geochemical patterns and their spatial distribution within the Antofagasta commune was achieved. This information can be further interpreted in the context of geological features and potential contamination sources to guide environmental management and decision-making in the region.

3. Results

3.1. Discarded elements

The exclusion of elements with insufficient data (<50%) regarding the detection level is imperative. In this initial step, the following elements were excluded: Ag, B, Be, Cd, Hg, Mo, Sb, Se, Sn, Tl, and U. Out of the elements present in the raw data table (32 elements), a refined selection is made, retaining 21 elements for further analysis.

This selection process is crucial for ensuring the reliability and robustness of the subsequent statistical analyses. Elements with inadequate data may introduce bias or uncertainty into the results, potentially compromising the overall integrity of the study. By focusing on the subset of elements with sufficient and reliable data, the analysis can yield more accurate insights into the environmental characteristics of the Antofagasta commune.

3.2. Preliminary factor analysis

The next crucial step involves a preliminary principal factor analysis (PFA). Through this analysis, it is determined that the appropriate number of factors is 5, sufficient to account for approximately 70% of the variance in the data. This strategic selection ensures a balance between capturing significant variability in the data and avoiding unnecessary complexity.

In the results of this PFA, two elements, namely Bi and Ca, are identified and subsequently excluded from further consideration due to their low communalities (<0.5). Communalities represent the proportion of variance in each variable that can be explained by the extracted factors. In this case, their exclusion is warranted by their limited contribution to the overall variance structure.

However, despite having a communality value below 0.5, Pb is retained in the analysis due to its paramount significance in soil pollution studies. The decision to preserve Pb is informed by its known environmental relevance and potential impact on contamination assessments. This deliberate choice ensures that the analysis maintains a comprehensive perspective on potential sources and implications of pollution in the Antofagasta commune.

With the exclusion of Bi and Ca the retention of Pb, the dataset is refined to a final set of 19 variables. These variables, characterized by their relevance to environmental factors and pollution dynamics, will be further explored in subsequent analyses to unravel the underlying patterns of soil geochemistry in the studied area.

3.3. Anomaly detection

Following this preliminary analysis and the exclusion of less important variables, the three anomaly detection methods mentioned before were employed to identify outliers within the remaining dataset. The identification of outliers is crucial for maintaining the integrity of the subsequent statistical analyses, ensuring that unusual or extreme values do not unduly influence the overall interpretation of the results.

The analysis reveals the presence of 16 anomalous data points out of 94 total samples, which are subsequently removed from further consideration. It is worth noting that the exact number of outliers identified may vary slightly depending on the specific method used for data normalization (e.g., ranging from 15 to 20 discarded data points). However, the decision to discard these outliers is made with care to maintain the consistency and reliability of the results.

This approach to handling outliers is necessary for obtaining a clearer understanding of the underlying environmental dynamics in the Antofagasta commune. The subsequent analyses, which include techniques such as cluster analysis and spatial mapping, will be conducted on a refined dataset that is free from the influence of extreme values. This ensures that the interpretations and conclusions drawn from the study are robust, reliable, and reflective of the true environmental conditions in the area under investigation.

3.4. Principal factor analysis

The final principal factor analysis, conducted after the removal of anomalous data, is presented in Table 1. The loadings represent the weights assigned to each element within the identified factors.

The interpretation of these factors, along with their respective loadings, forms a foundation for understanding the dominant sources and influencing factors contributing to the variability observed in the soil samples. This knowledge is crucial for informing environmental management strategies and regulatory decisions in the studied area.

We now present the analysis and interpretation of each factor.

- **Factor 1:** The first factor appears to represent a combination of elements associated with both natural geological processes and potential anthropogenic influences. High positive correlations among Cu, Fe, and Zn suggest a common source, possibly related to mineral deposits or industrial activities. The presence of Pb may indicate additional anthropogenic contributions. Mg and Mn could be linked to natural geological formations.
- **Factor 2:** The second factor seems to capture elements associated with different geological and geochemical processes. Al often reflects the influence of silicate minerals, while As may have both natural and anthropogenic sources. Ba is commonly associated with mineral deposits, and K is indicative of feldspar and mica minerals. Li is often found in igneous rocks, particularly in pegmatites and lithium-rich brines, and its presence in this factor may suggest the influence of

Table 1
Factor Loadings for the PFA. The highest absolute loading for each element is bolded for clarity.

Element	Factor 1	Factor 2	Factor 3	Factor 4
Al	0.507	**0.586**		0.560
As		**0.912**	−0.129	
Ba	0.190	**0.586**		−0.148
Cr		−0.262	0.150	**0.787**
Co	**0.865**	0.165	−0.191	0.391
Cu	**0.730**	0.254	−0.158	0.283
Fe	**0.896**	0.211	−0.309	0.112
Pb	**0.427**	0.257		
Li	0.531	**0.791**	−0.141	
Mg	**0.530**	0.393	0.443	0.363
Mn	**0.971**			
Ni	0.274	0.246	−0.304	**0.813**
P		−0.193	**0.836**	
K	0.158	**0.905**		0.229
Na		0.443	**0.577**	0.272
Sr	−0.277		**0.792**	−0.186
Ti	0.119	0.383	**−0.671**	
V	**0.821**	0.148	−0.375	0.138
Zn	**0.890**		0.156	

Table 2
Average concentration (mg kg^{-1}) of each element in the extracted clusters. The highest concentration for each element is bolded for clarity (excluding the noise cluster).

Element	C1 Avg	C2 Avg	C3 Avg	Noise
	mg kg^{-1}			
Al	13,902.38	10,444.24	13,892.86	15,701.50
As	**13.24**	9.41	11.70	13.76
Ba	**44.90**	37.49	39.42	46.87
Bi	**36.83**	34.04	33.44	42.74
Ca	36,283.62	54,147.95	**71,960.35**	128,773.44
Cr	23.65	20.90	**24.59**	26.73
Co	**12.36**	8.73	11.53	13.54
Cu	70.54	49.76	**127.22**	231.79
Fe	**30,657.75**	24,342.76	25,821.35	28,797.75
Pb	**12.38**	9.00	11.36	15.58
Li	20.90	15.53	**21.86**	29.29
Mg	13,896.25	11,420.43	**17,225.41**	22,492.88
Mn	**830.25**	471.52	635.69	934.42
Ni	12.14	8.27	**12.96**	13.97
P	1871.38	**2485.86**	1346.41	2494.62
K	**1868.00**	1333.71	1843.18	2282.12
Na	2853.50	2227.24	**2912.57**	3371.62
Sr	193.52	**355.99**	204.91	307.95
Ti	513.62	**690.35**	638.91	736.74
V	**87.65**	80.96	79.06	86.20
Zn	**219.44**	75.83	98.63	218.62

lithium-bearing geological formations or the occurrence of geothermal activity. This factor may represent the natural variability in the geological composition of the soil.

- **Factor 3:** The third factor is characterized by P, Na, Sr, and Ti. The negative correlation of Ti with this factor suggests an inverse relationship. P could be associated with organic matter or fertilizers, while Na and Sr may have both natural and anthropogenic origins. The negative correlation with Ti may indicate a distinct geological source for Ti.
- **Factor 4:** The fourth factor includes Cr and Ni. These elements are commonly associated with intrusive rocks and certain industrial activities. The positive correlation between Cr and Ni suggests a shared source, potentially related to natural geological processes or human activities.

3.5. Clustering analysis

For the clustering analysis, it was determined that the optimal division of the dataset was into three distinct clusters, with anomalous data considered as an additional "noise" cluster. The purpose of this clustering analysis is to reveal patterns and associations within the geochemical composition of the Antofagasta commune.

Table 2 presents the average concentrations of various elements across the different samples in the Antofagasta commune. The clustering results showcase the average concentrations for each element within the identified clusters. These results also provide valuable insights into the distribution and variability of elemental concentrations in the study area. Despite not being considered during the cluster constructions, the elements Ca and Bi were included in the description of each cluster, as they have potential utility in providing additional insights into the environmental context. The spatial distribution of the three extracted clusters, excluding the anomalous data points, is shown in Fig. 2.

The interpretation of these clusters is essential for understanding the spatial distribution of elements and potential sources of variation within the studied area. The average values derived from each cluster can serve as background concentrations, offering a reference point for future environmental assessments. This information is valuable for identifying areas with similar geochemical characteristics, which may be influenced by common geological formations, soil types, or anthropogenic activities. Thus, the clusters provide a basis for further investigation into the factors influencing elemental concentrations and support the development of targeted environmental management strategies based on the distinct characteristics of each group.

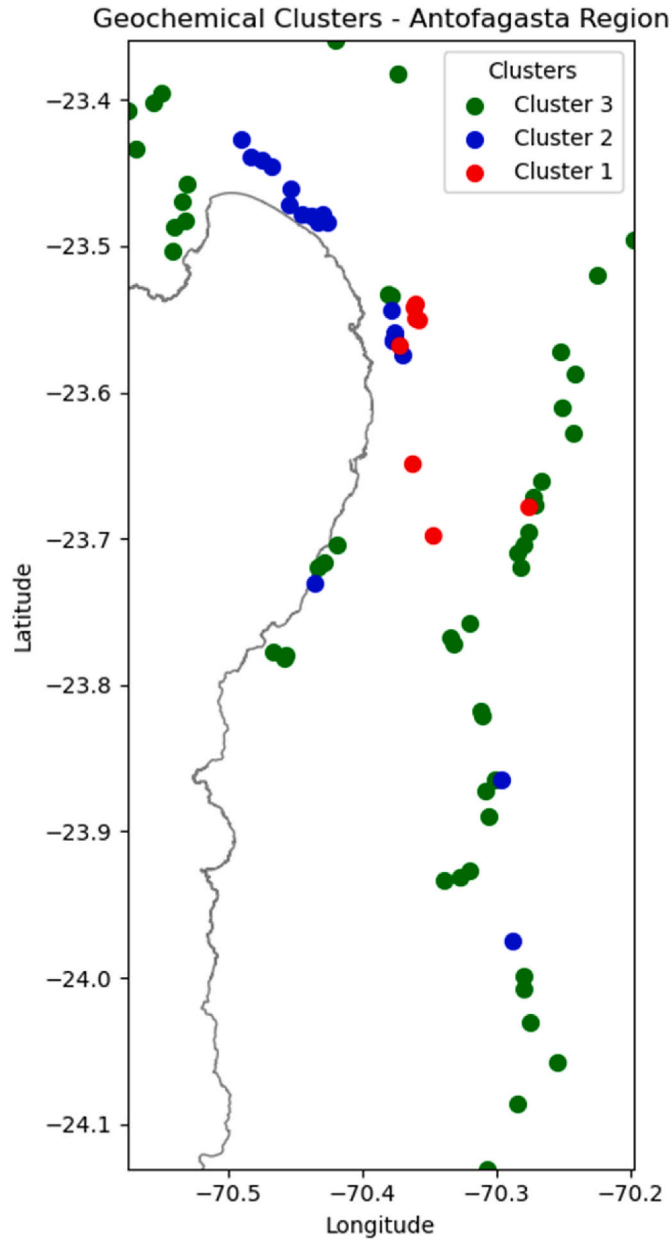


Fig. 2. Clustering analysis of the geochemical data of the Antofagasta commune. There are three clusters. The K-means clustering method does not account for the geographical location of the samples, just their geochemical characteristics.

Similar to the aforementioned factors, the extracted clusters can be interpreted to provide insights into the distribution of the observed elements in the soil samples.

We now present the analysis and interpretation of each cluster.

- **Cluster 1:** The first cluster exhibits heightened concentrations of Al, As, Ba, Bi, Co, Fe, Pb, Mn, K, V, and Zn. The intricate amalgamation of these elements implies a complex interplay between geological and anthropogenic factors in the Commune of Antofagasta. The presence of metals such as Pb and Mn may signal historical industrial activities, while elements like Al and K likely originate from geological sources. The diverse concentration levels within this cluster underscore the necessity for a nuanced understanding of the intricate balance between natural geological processes and past human-induced activities shaping the region's geochemical profile.

- **Cluster 2:** The second cluster is distinguished by elevated levels of P, Sr, and Ti. This grouping suggests a composite influence of both natural geological processes and potential anthropogenic sources. While agricultural practices, including fertilizers, are not a prominent factor, the geological origins of Sr and Ti concentrations add a distinct layer to this cluster. The unique composition of Cluster 2 underscores the importance of unraveling specific elemental

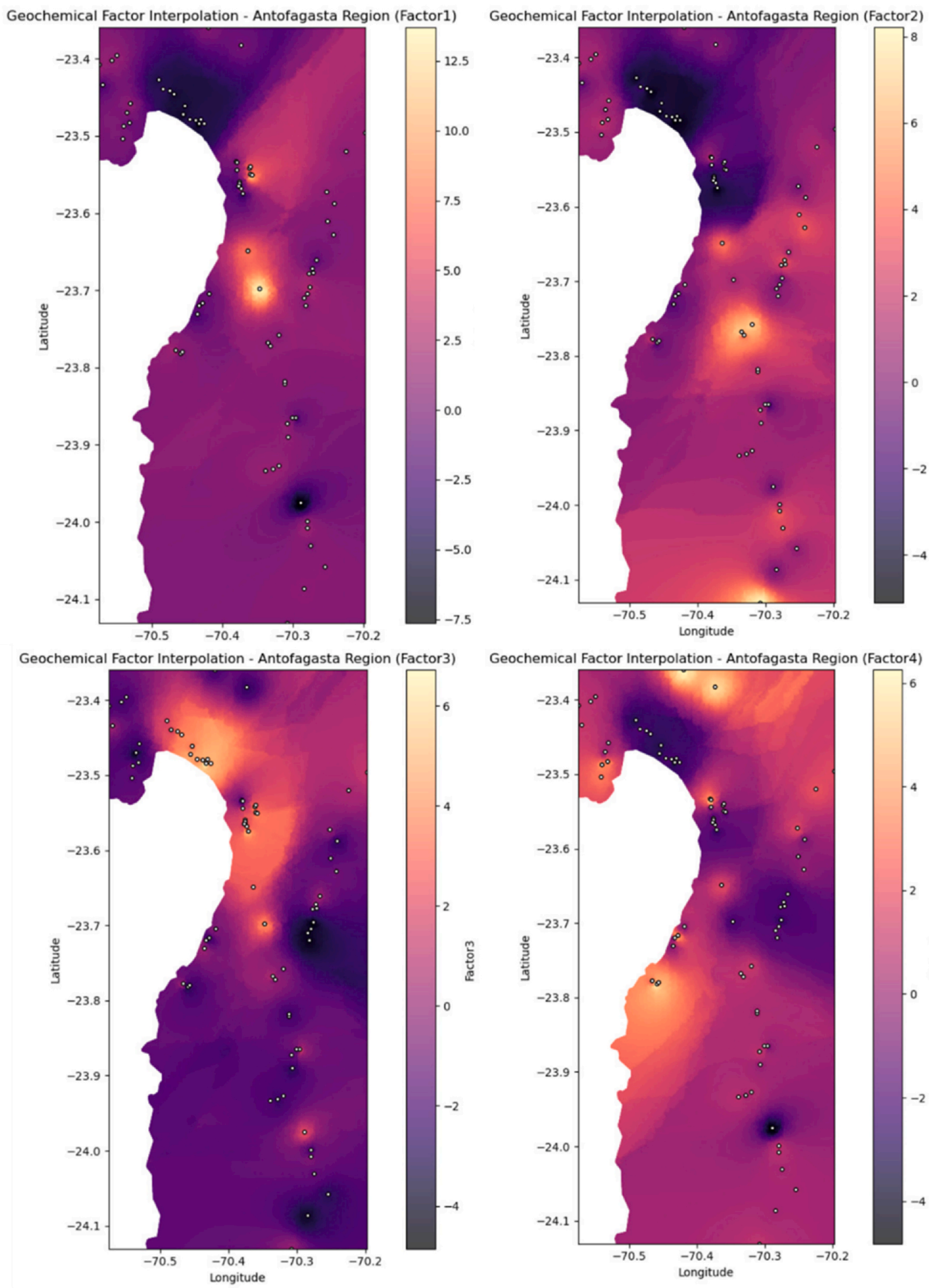


Fig. 3. Spatial analysis of the geochemical factors identified through principal factor analysis. The maps display the spatial distribution of factor scores, which represent the relative influence of each factor at each sampling location.

associations to comprehend the nuanced environmental dynamics at play in the Commune of Antofagasta.

- **Cluster 3:** The third cluster encompasses Ca, Cr, Cu, Li, Mg, Na, and Ni. The presence of these elements in this cluster hints at a dynamic interplay between geological processes and potential anthropogenic influences. Common soil constituents like Ca and Mg likely have geological origins, while elements such as Cr, Cu, and Ni may be linked to both natural geological processes and industrial sources. The inclusion of Li points towards geological formations, emphasizing the heterogeneous nature of the soil composition in the Commune of Antofagasta. Cluster 3 underscores the intricate balance between natural geological processes and potential anthropogenic factors shaping the geochemical landscape of the region.
- **Noise Cluster:** The noise cluster with all the anomalous data exhibits variability in elemental concentrations, potentially indicating heterogeneity or a mix of influences, which makes sense given its definition. It includes higher concentrations of certain elements like Ca, Mg, and Sr. This variability may result from a combination of natural geological processes and anthropogenic activities, highlighting the complexity of environmental conditions in these areas.

3.6. Geological analysis

The supplementary materials contain 19 dot concentration maps showing the spatial distribution of the chemical elements analyzed in this investigation. Based on the spatial analysis of the geochemical factors (Fig. 3) and considering the geological context, the following processes can be inferred.

- The spatial distribution of Factor 1 can be primarily attributed to Jurassic volcanic arc rocks, which are known to contain high concentrations of Al, Fe, Mn, and Zn. Additionally, Cu Co and V could be due to the region's potential mineralizations associated with these volcanic rocks. The spatial interpolation map of Factor 1 shows higher values in areas that align with the known distribution of these volcanic rocks, suggesting geological control (For a more detailed view of the spatial distribution of the elements of factor 1, please refer to the dot concentration maps for Al, Fe, Mn, Zn, Cu, Co, and V in the supplementary material).
- Factor 2 is likely explained by intrusive rocks and higher concentrations of feldspars indicated by the elevated levels of K, Al, and Na (dot concentration maps of K, Al, and Na of the supplementary material). The spatial distribution map for Factor 2 shows lower values in coastal areas with marine sediments and higher values inland where intrusive rocks are more common. This suggests that the geochemical signature of Factor 2 is influenced by the presence of intrusive rock bodies that are rich in feldspars.
- The elevated values of Factor 3 in coastal areas suggest a strong association with marine sediments. This is supported by the higher concentrations of elements like Sr, typically enriched in marine environments (dot concentration maps of Sr of the supplementary material). The spatial map of Factor 3 clearly shows higher values along the coast, aligning with the distribution of marine sediments.
- Intrusive rocks might influence Factor 4, given the association with elements such as Cr, Ni, and Mg. These rocks are known for their high concentrations of these elements. The spatial distribution map for Factor 4 shows certain localized high values that could correspond to areas where ultramafic rocks or related geological formations are present (dot concentration maps of Cr, Ni, and Mg of the supplementary material).

The spatial distribution maps for Factors 1 and 2, while showing some localized high values, are based on a robust dataset. These high values, potentially due to interpolation issues, should be interpreted with caution. More data points and further geological investigation are needed to confirm whether these are due to geological phenomena or interpolation effects.

The high concentrations of Al, Fe, Mn, and Zn suggest significant influence from Jurassic volcanic rocks. Cu, Co, and V indicate possible hydrothermal processes or secondary mineralization related to volcanic rocks that are subsequently observed in the soil samples (Herrera et al., 2023). The high concentrations of Al, K, and Na indicate intrusive rocks rich in feldspars. This is consistent with the region's geological history, which includes significant intrusive activity. The low values in coastal sediments and higher values inland suggest a clear differentiation between marine sediment influence and terrestrial geological formations. The elevated Sr values point to marine sediment influence, corroborated by the spatial distribution map showing higher values along the coastline. This factor highlights marine processes' geochemical contribution in shaping coastal regions' sediment composition. The association with Cr, Ni, and Mg suggests ultramafic rock influence, possibly from mantle-derived materials or ultramafic intrusions. The map shows specific areas with high concentrations, likely pointing to localized geological formations with ultramafic characteristics.

3.7. Geochemical background estimates

The final geochemical background concentrations for the elements in the Antofagasta commune were determined using the clustering results and three estimation methods mentioned before (MAD, upper whisker, and 95th percentile methods). The background concentrations were calculated separately for each of the three clusters identified through the *k*-means clustering analysis. We show the average for all three methods in Table 3. In the supplementary material, we provide Table S1 which contains all the results for the three estimation methods.

Cluster 1 generally exhibits the lowest background concentrations for most elements, with notable exceptions being Cu, Pb, and Sr. The median + 2MAD estimates for Cluster 1 range from 10.75 mg kg^{-1} for As to 62,227.50 mg kg^{-1} for Fe. The Tukey boxplot upper whisker values are slightly higher, ranging from 15.74 mg kg^{-1} for As to 76,345.63 mg kg^{-1} for Fe. The 95th percentile values provide the most conservative estimates, with As at 14.98 mg kg^{-1} and Fe at 62,598.30 mg kg^{-1}.

Cluster 2 displays higher background concentrations compared to Cluster 1 for several elements, including Ca, Mn, P, and Sr. The median + 2MAD estimates for Cluster 2 range from 6.10 mg kg^{-1} for Pb to 167,394.00 mg kg^{-1} for Ca. The Tukey boxplot upper whisker values span from 9.70 mg kg^{-1} for As to 328,000.50 mg kg^{-1} for Ca. The 95th

Table 3
Average background concentration (mg kg^{-1}) of each element in the extracted clusters based on the MAD, Upper Whisker, and 95th Percentile methods.

Element	C1 Background	C2 Background	C3 Background
	mg kg^{-1}		
Al	23,217.74	14,067.50	19,851.90
As	13.82	9.10	24.72
Ba	55.39	40.55	76.30
Bi	56.93	65.28	72.18
Ca	59,708.63	23,9427.80	77,701.07
Cr	31.94	32.87	35.09
Co	28.59	11.02	14.54
Cu	158.98	56.38	121.05
Fe	67,057.14	24,290.50	35,654.77
Pb	26.40	12.00	21.52
Li	50.23	19.88	33.47
Mg	38,084.41	29,600.67	22,603.30
Mn	2740.01	657.33	638.60
Ni	17.46	10.60	19.62
P	5408.09	10,747.67	1595.40
K	2954.80	1497.17	2949.33
Na	6815.41	5343.33	6164.50
Sr	199.15	1853.67	310.63
Ti	924.37	453.33	1177.83
V	173.72	80.57	114.35
Zn	675.92	135.50	96.12

percentile values range from 8.4 mg kg^{-1} for Ni to 222,889.00 mg kg^{-1} for Ca.

Cluster 3 exhibits the highest background concentrations for most elements among the three clusters. The median + 2MAD estimates for Cluster 3 range from 12.50 mg kg^{-1} for Co to 36,342.00 mg kg^{-1} for Ca. The Tukey boxplot upper whisker values extend from 17.55 mg kg^{-1} for Co to 88,178.00 mg kg^{-1} for Ca. The 95th percentile values range from 13.58 mg kg^{-1} for Co to 108,583.20 mg kg^{-1} for Ca.

The three estimation methods provide a range of background concentration values for each element within the clusters. The median + 2MAD method offers a balance between robustness and sensitivity to outliers, while the Tukey boxplot upper whisker and the 95th percentile methods provide more conservative estimates.

4. Discussion

The geochemical complexity of the Antofagasta commune is determined by the interaction of its unique geological environment and the presence of mineral deposits, extreme aridity, and the influence of anthropogenic activities.

4.1. Discussion on local geology and its impact on soil geochemistry

The Jurassic volcanic rocks of the region, represented by the La Negra Formation, play a significant role in the area's geochemical background. These volcanic rocks, which reach a maximum thickness of 8 km and are mainly composed of andesites and basaltic andesites interbedded with lesser amounts of continental and marine volcaniclastic rocks (Charrier et al., 2007), host various mineral deposits, such as copper, gold, and silver (Boric et al., 1990). The high concentrations of elements like Cu, Co, and V in the soil samples can be attributed to the presence of these mineral deposits and the associated hydrothermal processes that have enriched these elements in the volcanic rocks (Herrera et al., 2023). These hydrothermal processes, which involved seawater during the Jurassic and Cretaceous, affected the volcanic rocks and contributed to the formation of manto-type deposits containing copper, gold, and silver (Boric et al., 1990; Kojima et al., 2009).

The intrusive rocks of the region, particularly those of the Middle-Upper Jurassic, also contribute to the geochemical diversity of the Antofagasta commune. These intrusive bodies are characterized by a wide compositional range, from gabbros to granites (Scheuber and Gonzalez, 1999), and exhibit high concentrations of elements such as K, Al, and Na, indicative of the presence of feldspars. Additionally, within these intrusive bodies, there are vein-type deposits that contain copper, gold, and silver mineralizations (Boric et al., 1990). The spatial distribution of these elements, revealed by the factor analysis and geostatistical maps, aligns with the known distribution of the intrusive rocks in the area.

In addition to the geological influences, anthropogenic factors have played a significant role in shaping the geochemical landscape of the Antofagasta commune. The region has a long history of mining activities, especially related to copper extraction. The transport of minerals along road networks for several decades has likely contributed to the dispersion of metals and other elements into the surrounding environment (Lagos and Blanco, 2010). The presence of small-scale mining operations in the region can also introduce additional elements into the soil through the exposure of mineralized rocks, generation of mine tailings, and the use of processing chemicals (Lam et al., 2016).

Enrichment factors suggest that Fe could originate from any outcrop present in the area, while Mn could come from metamorphic, volcanic, and intrusive rocks. Therefore, it is inferred that both elements are naturally present in the soil of Antofagasta and their concentrations are comparable to the world average for these elements at the continental crust level (Fe 3.9% or 39,176.00 mg kg^{-1} and Mn 0.077% or 774.00 mg kg^{-1}; Rudnick and Gao, 2003). However, Tapia et al. (2018) consider that As could originate from soils previously enriched or contaminated with this element, while Mn and Fe are not considered contaminants and could originate from the rocks present in the region.

The integration of these findings with the geochemical analysis and background concentration estimates presented in this study contributes to a more comprehensive understanding of the geochemical complexity of the region. It highlights the importance of distinguishing between natural geological sources and anthropogenic contaminants when assessing soil composition and potential environmental and health risks in the Antofagasta commune.

The spatial distribution maps of the geochemical factors, albeit with some limitations, offer valuable insights into the spatial patterns and their alignment with the region's geological features. However, the spatial variability of certain elements near the port and their possible relationship to mineral transport and storage activities highlights the need to consider both natural and anthropogenic sources when interpreting these patterns.

4.2. Effect of arid climate on soil contamination

In arid regions, it is well known that climatic aridity facilitates the accumulation of dust from both the weathering of local geological formations and the contribution of dry, anthropogenic residues carried by the wind (Custodio, 1992). In the Atacama Desert, materials deposited on the soil surface tend to remain in place and not leach (Amundson, 2003). This occurs due to the hyperaridity, and the extremely low precipitation recorded in the Atacama Desert, meaning that the arid landscape is not subject to regular water erosion, except for the few rivers, some of which are intermittent, that are fed by precipitation in the highlands (Arenas-Díaz et al., 2022). Specifically in the coastal area of Antofagasta, this phenomenon is associated with the significant amount of marine aerosol that can accumulate over decades before being washed away by rainfall and recharging the aquifers (Herrera and Custodio, 2014).

The hyper-arid climate of the city of Antofagasta has a significant impact on soil contamination, as the dry conditions and low precipitation limit the natural leaching of pollutants. Due to the limited availability of water, contaminants do not dissolve or move to deeper layers of the soil, resulting in higher concentrations at the surface and increasing the risk of exposure and wind-driven displacement, particularly in coastal areas. This condition promotes the accumulation of heavy metals, primarily originating from mining and industrial activities, which do not dilute or disperse easily (Golan et al., 2024).

Research conducted in the urban center of Antofagasta shows the buildup of fine particulate matter on the façades of buildings and houses, which remains indefinitely due to the near absence of rainfall (Palme, 2014). This phenomenon is a clear indication of what could also occur in the suburban areas of Antofagasta, where fine particulate matter would accumulate on the surface of various soils and persist for extended periods due to the lack of precipitation. The concentration of heavy metals accumulating on the soil surface would be directly linked to the region's arid climate.

Studies continue to provide evidence of the significant influence of anthropogenic mining activities on the geochemical composition of the Antofagasta region (Queirolo et al., 2000). Tapia et al. (2018) examined the distribution and enrichment of various elements, including Cu and As, in river sediments across the Northern Atacama Region. Their findings indicate substantial enrichment of these elements in areas affected by mining activities, with enrichment factors for Cu reaching up to 55 times the background levels in some locations. Moreover, they observed that the redistribution of these elements extends far beyond the immediate vicinity of mining operations, affecting downstream watersheds. These results align with earlier observations by Flynn et al. (2002), who reported Cu levels of up to 499 mg L-1 in aqueous extracts of soil and sediment samples from the Antofagasta lowlands, with elevated concentrations extending up to 20 km away from a Cu smelter. The persistence of these geochemical signatures over time, as evidenced by these studies, underscores the long-term impact of mining activities on the region's environment and highlights the ongoing need for comprehensive environmental monitoring and protection measures.

4.3. Limitations

The current study provides valuable insights into the geochemical characteristics and spatial distribution of elements in the Antofagasta commune. However, the analysis has relied on a limited number of sampling points, which may not fully capture the spatial variability of the geochemical factors across the entire study area. Additionally, while the IDW method used for spatial interpolation is robust and widely used, it has some limitations, such as sensitivity to outliers and extreme values, lack of uncertainty quantification for the interpolated values, smoothing effect that may obscure local variations, and inability to extrapolate beyond the extent of the sampling points.

To address these limitations and further enhance the analysis, future studies could consider increasing the sampling density, exploring more advanced geostatistical methods, incorporating additional environmental and anthropogenic variables, and conducting temporal studies and validation using independent datasets or ground-truthing techniques. These approaches would provide a more comprehensive understanding of the geochemical complexity of the Antofagasta commune and improve the robustness and reliability of the findings.

In general, the integration of geochemical, geological, and anthropogenic contamination data is essential for developing effective environmental management strategies in the region. Future studies addressing the spatial and temporal variability of contamination, as well as the mechanisms of element transport and dispersion, would provide a strong foundation for decision-making and the protection of the environment and public health in the Antofagasta commune.

5. Conclusions

This study aimed to unravel the geochemical complexities of the Antofagasta commune through a comprehensive analysis of soil samples. Multivariate statistical approaches, including principal factor analysis and cluster analysis, provided valuable insights into the variability of environmental elements and the factors influencing their distribution. Through the application of robust multivariate statistical methods and geostatistical analysis, we have identified four principal factors and three distinct geochemical clusters that characterize the region's soil composition.

Our findings reveal that the geochemical background of Antofagasta is primarily determined by its diverse geological setting, which includes Jurassic volcanic and intrusive rocks, and Cenozoic marine sediments. However, the long history of mining activities in the region has left a discernible imprint on the soil geochemistry, particularly evident in the elevated concentrations of elements such as Cu, As, and Pb in certain areas.

The establishment of element-specific background concentrations for each identified cluster provides a nuanced understanding of the region's geochemical baseline. This information is invaluable for distinguishing between natural elemental enrichment and anthropogenic contamination, a critical distinction in a region with significant mining activities.

The spatial distribution maps of geochemical factors, generated through geostatistical analysis, offer a visual representation of the complex patterns of elemental distribution across the commune. These maps, when interpreted in conjunction with geological data, provide a powerful tool for identifying areas of potential environmental concern and guiding targeted sampling and remediation efforts.

The methodological approach employed in this study, combining robust statistical techniques with spatial analysis, offers a comprehensive framework for geochemical characterization that can be adapted to other mining-intensive regions worldwide. This approach is particularly valuable in areas with complex geological histories and significant anthropogenic influences, where distinguishing between natural and human-induced geochemical signatures is challenging.

While this study provides a solid foundation for understanding the geochemical landscape of Antofagasta, it also highlights areas for future research. These include the need for higher-resolution sampling in areas of particular interest, the incorporation of additional environmental variables, and the exploration of temporal changes in soil geochemistry.

In conclusion, this research not only enhances our understanding of Antofagasta's geochemical makeup but also provides a robust scientific basis for informed decision-making in environmental management, urban planning, and the development of region-specific soil quality guidelines. By offering a more nuanced perspective on what constitutes 'background' levels in a mining-intensive region, this study contributes to the broader field of environmental geochemistry and supports the development of sustainable mining practices in Chile and beyond.

CRediT authorship contribution statement

Brian F. Keith: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Elizabeth J. Lam:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Conceptualization. **Ítalo L. Montofré:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Vicente Zetola:** Writing – original draft, Validation, Methodology. **Javier Urrutia:** Writing – review & editing, Visualization, Validation, Software. **Christian Herrera:** Writing – review & editing, Validation, Software, Methodology. **Jaume Bech:** Writing – review & editing, Writing – original draft, Validation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We would like to thank our colleague Dr. Joseline Tapia Zamora at Universidad Católica del Norte for their valuable feedback throughout this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemosphere.2024.143472.

References

Aitchison, J., 1982. The statistical analysis of compositional data. J. Roy. Stat. Soc. B 44 (2), 139–160.
Amundson, R., 2003. Soil Formation. In: Treatise on Geochemistry, vol. 5. Elsevier Science, pp. 1–35.
Ander, E.L., Johnson, C.C., Cave, M.R., Palumbo-Roe, B., Nathanail, C.P., Lark, R.M., 2013. Methodology for the determination of normal background concentrations of contaminants in English soil. Sci. Total Environ. 454, 604–618.
Arenas-Díaz, F., Fuentes, B., Reyers, M., Fiedler, S., Böhm, C., Campos, E., Shao, Y., Bol, R., 2022. Dust and aerosols in the Atacama Desert. Earth Sci. Rev. 226, 103925.
Arens, F.L., Airo, A., Feige, J., Sager, C., Wiechert, U., Schulze-Makuch, D., 2021. Geochemical proxies for water-soil interactions in the hyperarid Atacama Desert, Chile. Catena 206, 105531.
Biskupovic, C., 2015. From rural haven to civil political project: utopian ideals and environmental protection in the precordillera, Santiago, Chile. Journal of Political Ecology 22 (1), 183–198.

Boric, R., Díaz, F., Maksaev, V., 1990. Geología y yacimientos metalíferos de la Región de Antofagasta, vol. 40. Servicio Nacional de Geología y Minería, Boletín, p. 246. Santiago.
Castro, S.H., Sánchez, M., 2003. Environmental viewpoint on small-scale copper, gold and silver mining in Chile. J. Clean. Prod. 11 (2), 207–213.
CENMA, 2014. Informe final Versión 5. Diagnóstico regional de suelos abandonados con potential presencia de contaminantes. Contrato Nº 618775–3-LP13. (Available at:
Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. Determining the number of clusters using NbClust package. MSDM 2014, 1.
Charrier, R., Pinto, L., Rodríguez, M.P., 2007. Tectonostratigraphic Evolution of the Andean Orogen in Chile.
Charrier, R., Ramos, V.A., Tapia, F., Sagripanti, L., 2015. Tectono-stratigraphic evolution of the andean orogen between 31 and 37 S (Chile and western Argentina). Geological Society, London, Special Publications 399 (1), 13–61.
Custodio, E., 1992. Coastal aquifer salinization as a consequence of aridity: the case of Amurga phonolitic massif, Gran Canaria Island. In: Proceedings No. 12, Saltwater Intrusion Meeting Study and Modelling of Salt Water Intrusion into Aquifers. Centre Internacional de Mètodes Numèrics en Enginyería- Universitat Politècnica de Catalunya, Barcelona, pp. 81–98.
Desaules, A., 2012. Critical evaluation of soil contamination assessment methods for trace metals. Sci. Total Environ. 426, 120–131.
Dirección General de Aguas, 2009. Levantamiento Hidrogeológico para el desarrollo de nuevas fuentes de agua en áreas prioritarias de la Zona Norte de Chile, regiones XV, I, II y III: etapa 2. Informe Final. Parte IX. Sistema Piloto II Región Salares El Laco Y Aguas Calientes 2, Laguna Tuyajto Y Pampas Puntas Negras, Las Tecas Y Colorada.
Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., et al., 2003. Isometric logratio transformations for compositional data analysis. Math. Geol. 35, 279–300. https://doi.org/10.1023/A:1023818214614.
Filzmoser, P., Hron, K., Reimann, C., 2009. Principal component analysis for compositional data with outliers. Environmetrics: The Official Journal of the International Environmetrics Society 20 (6), 621–632.
Flynn, H.C., Mc Mahon, V., Diaz, G.C., Demergasso, C.S., Corbisier, P., Meharg, A.A., Paton, G.I., 2002. Assessment of bioavailable arsenic and copper in soils and sediments from the Antofagasta region of northern Chile. Sci. Total Environ. 286 (1–3), 51–59.
Gałuszka, A., 2007. A review of geochemical background concepts and an example using data from Poland. Environ Geol 52, 861–870. https://doi.org/10.1007/s00254-006-0528-2.
García, D., Tapia, J., Aguilera, J., Vega, C., Zúñiga, P., Lavin, P., et al., 2024. Contamination of urban soils in a historical mining town of northern Chile. Environ. Earth Sci. 83 (6), 182.
Garreaud, R.D., 2009. The Andes climate and weather. Advances in geosciences 22, 3–11.
Golan, R., Gavrieli, I., Katzir, R., Sharabi, G., Nachshon, U., 2024. Soil contamination in arid environments and assessment of remediation applying surface evaporation capacitor model. A Case Study from the Judean Desert, Israel. EGUsphere [preprint]. https://doi.org/10.5194/egusphere-2024-1014.
González, G., Niemeyer, H., 2005. Cartas Antofagasta y punta tetas, región Antofagasta. Servicio Nacional de Geología y Minería, Carta Geológica de Chile, Serie Geología Básica 89 (35), 1.
González, L.G., Dunai, T., Carrizo, D., Allmendinger, R., 2006. Young displacements on the Atacama Fault System, northern Chile from field observations and cosmogenic 21Ne concentrations. Tectonics 25 (3).
Herrera, C., Custodio, E., 2014. Origin of waters from small springs located at the northern coast of Chile, in the vicinity of Antofagasta. Andean Geol. 41, 314–341. https://doi.org/10.5027/andgeoV41n2-a03.
Herrera, C., Godfrey, L., Urrutia, J., Custodio, E., Gamboa, C., Jodar, J., Lam, E., Fuentes, J., 2023. Origin of old saline groundwater in the deep coastal formations of the Atacama Desert region: consideration of lithium, boron, strontium and uranium isotopes contents. J. Hydrol. 624, 129919.
Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9 (3), 90–95.
ISO-19258, 2018. Soil Quality — Guidance on the Determination of Background Values. International Organization for Standardization, Genève, Switzerland, p. 26.
Johnson, C.C., Ander, E.L., Cave, M.R., Palumbo-Roe, B., 2012. Normal Background Concentrations (NBCs) of Contaminants in English Soils: Final Project Report.
Kay, S.M., Mpodozis, C., Ramos, V.A., Munizaga, F., 1991. Magma Source Variations for Mid-late Tertiary Magmatic Rocks Associated with a Shallowing Subduction Zone and a Thickening Crust in the Central Andes (28 to 33 S), vol. 265. Geological Society of America Special Paper, pp. 113–137.
Khadka, D.R., 2023. Geostatistics and inverse distance weighing (IDW) multivariate interpolation methods for regional geochemical stream sediment survey for base metal prospecting around Wapsa, Solukhumbu District, Nepal. J. Nepal Geol. Soc. 61–74.
Klohn, W., 1972. Hidrografía de las zonas desérticas de Chile.
Kojima, S., Trista, D., Hayashi, K., 2009. Genetic aspects of the manto-type copper deposits based on geochemical studies of North Chilean deposits. Resour. Geol. 59, 87–98. https://doi.org/10.1111/j.1751-3928.2008.00081.x.
Kumar, N., Sinha, N.K., 2018. Geostatistics: principles and applications in spatial mapping of soil properties. Geospatial Technologies in Land Resources Mapping, Monitoring and Management, pp. 143–159.
Lagos, G., Blanco, E., 2010. Mining and development in the region of Antofagasta. Resour. Pol. 35 (4), 265–275.
Lam Esquenazi, E.J., Keith Norambuena, B.F., Montofre Bacigalupo, Í.L., Gálvez Estay, M.E., 2019. Necessity of intervention policies for tailings identified in the Antofagasta Region, Chile. Rev. Int. Contam. Ambient. 35 (3), 515–539.
Lam, E.J., Gálvez, M.E., Cánovas, M., Montofré, I.L., Rivero, D., Faz, A., 2016. Evaluation of metal mobility from copper mine tailings in northern Chile. Environ. Sci. Pollut. Control Ser. 23, 11901–11915.
Lam, E.J., Keith, B.F., Montofré, Í.L., Gálvez, M.E., 2018. Copper uptake by Adesmia atacamensis in a mine tailing in an arid environment. Air Soil. Water Res. 11, 1178622118812462.
Lam, E.J., Montofré, I.L., Álvarez, F.A., Gaete, N.F., Poblete, D.A., Rojas, R.J., 2020. Methodology to prioritize chilean tailings selection, according to their potential risks. Int. J. Environ. Res. Publ. Health 17 (11), 3948.
Lam, E.J., Urrutia, J., Bech, J., Herrera, C., Montofré, Í.L., Zetola, V., et al., 2023. Heavy metal pollution index calculation in geochemistry assessment: a case study on Playa Las Petroleras. Environ. Geochem. Health 45 (2), 409–426.
Liu, Y., Cheng, Q., Zhou, K., Xia, Q., Wang, X., 2016. Multivariate analysis for geochemical process identification using stream sediment geochemical data: a perspective from compositional data. Geochem. J. 50 (4), 293–314.
MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press.
Maronna, R.A., Zamar, R.H., 2002. Robust multivariate estimates for high dimensional data sets. Technometrics 44 (4), 307–317.
Matschullat, J., Ottenstein, R., Reimann, C., 2000. Geochemical background—can we calculate it? Environ. Geol. 39, 990–1000 [CrossRef].
McIlwaine, R., Cox, S.F., Doherty, R., Palmer, S., Ofterdinger, U., McKinley, J.M., 2014. Comparison of methods used to calculate typical threshold values for potentially toxic elements in soil. Environ. Geochem. Health 36, 953–971.
Muñoz, J.B., Stern, C.R., 1989. Alkaline magmatism within the segment 38°–39° S of the plio-quaternary volcanic belt of the southern south American continental margin. J. Geophys. Res. Solid Earth 94 (B4), 4545–4560.
Nacional Institute of Statistics, 2018. Estimaciones y proyecciones de la población de Chile 1992–2050. Instituto Nacional de Estadistica. Subdepartamento de Demografía y Estadísticas Vitales.
Olive, D.J., 2004. A resistant estimator of multivariate location and dispersion. Comput. Stat. Data Anal. 46 (1), 93–102.
Oliveros, V., Morata, D., Aguirre, L., Féraud, G., Fornari, M., 2007. Magmatismo asociado a subducción del Jurásico a Cretácico Inferior en la Cordillera de la Costa del norte de Chile (18° 30'-24° S): geoquímica y petrogenesis. Revista geológica de Chile 34 (2), 209–232.
Oyarzun, R., Guevara, S., Oyarzún, J., Lillo, J., Maturana, H., Higueras, P., 2006. The As-contaminated Elqui river basin: a long lasting perspective (1975–1995) covering the initiation and development of Au–Cu–As mining in the high Andes of northern Chile. Environ. Geochem. Health 28, 431–443.
Palme, M., 2014. Urban sustainability in arid climates: challenges for Antofagasta, Chile. World Sustainability Forum 2014 – Conference Proceedings Paper.
Pérez-Carrera, A., Fernández Cirelli, A., 2010. Arsenic and water quality challenges in South America. Water and Sustainability in Arid Regions: Bridging the Gap Between Physical and Social Sciences 275–293.
Project BIP 30359875-0, 2019. Environmental risk assessment Antofagasta. Chapter 3. Annex E - Background Statistics. Source. https://mma.gob.cl/wp-content/uploads/2019/12/Diagnostico-Riesgo-Ambiental-Region-Antofagasta-2.zip.
Queirolo, F., Stegen, S., Mondaca, J., Cortes, R., Rojas, R., Contreras, C., Munoz, L., Schwuger, M.J., Ostapczuk, P., 2000. Total arsenic, lead, cadmium, copper, and zinc in some salt rivers in northern Andes of Antofagasta, Chile. Sci. Total Environ. 255, 85–95.
Redon, P.O., Bur, T., Guiresse, M., Probst, J.L., Toiser, A., Revel, J.C., et al., 2013. Modelling trace metal background to evaluate anthropogenic contamination in arable soils of south-western France. Geoderma 206, 112–122.
Reimann, C., Filzmoser, P., Garrett, R.G., 2002. Factor analysis applied to regional geochemical data: problems and possibilities. Appl. Geochem. 17 (3), 185–206.
Reimann, C., Filzmoser, P., Garrett, R.G., 2005. Background and threshold: critical comparison of methods of determination. Sci. Total Environ. 346 (1–3), 1-1.
Reimann, C., Filzmoser, P., Garrett, R., Dutter, R., 2011. Statistical Data Analysis Explained: Applied Environmental Statistics with R. John Wiley & Sons.
Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P. (Eds.), 2014. Chemistry of Europe's Agricultural Soils, Part A: Methodology and Interpretation of the GEMAS Data Set. Schweizerbart Science Publishers.
Reyes, A., Thiombane, M., Panico, A., Daniele, L., Lima, A., Di Bonito, M., De Vivo, B., 2020. Source patterns of potentially toxic elements (PTEs) and mining activity contamination level in soils of Taltal city (northern Chile). Environ. Geochem. Health 42, 2573–2594.
Reyes, A., Cuevas, J., Fuentes, B., Fernández, E., Arce, W., Guerrero, M., Letelier, M.V., 2021. Distribution of potentially toxic elements in soils surrounding abandoned mining waste located in Taltal, Northern Chile. J. Geochem. Explor. 220, 106653.
Rousseeuw, P.J., Driessen, K.V., 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics 41 (3), 212–223.
Rutllant, J., 1985. Natural Desertification Mechanisms along the Arid West Coast of South America. Proceeding of the International Workshop on Sand Transportation and Desertification in Arid Lands, Khartoum, Sudan, pp. 235–252.
Scheuber, E., Gonzalez, G., 1999. Tectonics of the Jurassic-Early Cretaceous magmatic arc of the north Chilean Coastal Cordillera (22–26 S): a story of crustal deformation along a convergent plate boundary. Tectonics 18 (5), 895–910.
Servicio Nacional de Geología y Minería, SERNAGEOMIN, 2002. Mapa Geológico de Chile. Carta Geológica de Chile, No.75, Serie Geología Básica, Escala 1: 1.000.000.
Shreemathi, M., Senthilkumar, B., Sujithra, S.M., Praisoodi, A., Rithika, S., 2024. Mastering geospatial analysis with Python: understanding geopandas, GDAL, fiona, matplotlib, data integration, and GIS tools. In: Ethics, Machine Learning, and Python in Geospatial Analysis. IGI Global, pp. 120–149.

Smart, S., 2020. The political economy of Latin American conflicts over mining extractivism. Extr. Ind. Soc. 7 (2), 767–779.
Tapia, J., Davenport, J., Townley, B., Dorador, C., Schneider, B., Tolorza, V., von Tümpling, W., 2018. Sources, enrichment, and redistribution of As, Cd, Cu, Li, Mo, and Sb in the Northern Atacama Region, Chile: implications for arid watersheds affected by mining. J. Geochem. Explor. 185, 33–51.
Templ, M., Filzmoser, P., Reimann, C., 2008. Cluster analysis applied to regional geochemical data: problems and possibilities. Appl. Geochem. 23 (8), 2198–2213.
Templ, M., Hron, K., Filzmoser, P., 2011. Rob-compositions: robust estimation for compositional data. Manual and Package, Version 1.4.4.
Teng, Y., Ni, S., Wang, J., Zuo, R., Yang, J., 2010. A geochemical survey of trace elements in agricultural and non-agricultural topsoil in Dexing area, China. J. Geochem. Explor. 104 (3), 118–127.
Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. Roy. Stat. Soc. B 63 (2), 411–423.
Tume, P., Roca, N., Rubio, R., King, R., Bech, J., 2018. An assessment of the potentially hazardous element contamination in urban soils of Arica, Chile. J. Geochem. Explor. 184, 345–357.
Tume, P., González, E., Reyes, F., Fuentes, J.P., Roca, N., Bech, J., Medina, G., 2019. Sources analysis and health risk assessment of trace elements in urban soils of Hualpen, Chile. Catena 175, 304–316.
Tyler, S.W., 2020. Are arid regions always that appropriate for waste disposal? examples of complexity from yucca Mountain, Nevada. Geosciences 10 (1), 30.
Van Den Boogaart, K.G., Tolosana-Delgado, R., Bren, R., 2011. Compositions: compositional data analysis. R Package Version 1, 10–12. http://CRAN.Rprojector g/package=compositions.
Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17 (3), 261–272.
Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol. Evol. 1 (1), 3–14.