Geochemical trends in sedimentary environments using PCA approach

DEEPSHIKHA SRIVASTAVA[1], CHANDRA PRAKASH DUBEY[2,*], UPASANA SWAROOP BANERJI[2,3] and KUMAR BATUK JOSHI[2]

[1]*Central University of Karnataka, Kadaganchi Campus, Gulbarga, Karnataka 585 311, India.*
[2]*National Centre for Earth Science Studies, Thiruvananthapuram, Kerala 695 011, India.*
[3]*Ministry of Earth Sciences, New Delhi 110 003, India.*
Corresponding author. e-mail: p.dubey48@gmail.com

Investigating the geochemical composition of bulk sediments stands as a crucial method for unraveling the complexities of various sedimentary processes. However, the intricacies arising from extensive datasets and alterations in sediment due to diverse factors often impede the clear identification of underlying patterns in geochemical fluctuations. In addressing these, employing multivariate statistical analyses has proven to be an invaluable tool for elucidating intricate patterns within large dataset. In this study, we focus on the utilization of Principal Component Analysis (PCA), a multivariate statistical technique, to uncover the underlying sedimentary processes influencing distinct geochemical dataset. Specifically, our attention is directed towards the examination of geochemical data from the previously published geochemical data of metasediments from Shimla and Chail group (referred to as SCM) and the mudflat sediments of Diu Island (referred to as DMS). Our PCA outcomes reveal that the initial three principal components (PC1, PC2, and PC3) account for 52.51% and 79.30% of the total variance within the SCM and DMS geochemical data, respectively. Notably, the negative loading of SiO_2, alongside positive loadings of incompatible elements and those associated with mafic rocks on PC1 within the SCM dataset, indicates sediment origins ranging from felsic to intermediate sources. Additionally, the coexistence of Th, U, Zr, and Sc, exhibiting positive loadings in PC1 and PC2, suggests a significant influence of reworking and recycling from felsic to intermediate sources. In the context of the DMS dataset, PCA analysis highlights the dominant influence of *in-situ* productivity and mafic sediment sources along the positive axis of PC1. Conversely, the negative axis of PC1 is shaped by intermediate and potentially other sources. Further granularity in interpretation reveals the positive axis of PC2 being attributed to weathering proxies, while the dominance of plagioclase minerals in the clayey fraction controls the positive axis of PC3. Through this investigation, our study underscores the essential role of PCA-assisted geochemical data analysis in unraveling the intricate web of processes contributing to the variance observed within sedimentary systems. By effectively distilling the multifaceted factors driving geochemical variability, this approach emerges as a pivotal asset in enhancing our understanding of sedimentary dynamics.

1. Introduction

The realm of comprehensive geochemical and isotopic dataset concerning rocks and sediments has unfurled expansive avenues for delving into the gradual evolution of the continental crust (Haughton *et al.* 1991; Schwab 2003; Joshi *et al.* 2022a), the intricate genesis of rocks (Joshi *et al.* 2017, 2022b), as well as the nuanced reconstruction of paleoclimatic and paleogeographic conditions (Ramirez-Herrera *et al.* 2007; Tripathy *et al.* 2014; Shaji *et al.* 2022), and the estimation of provenance (Lipp *et al.* 2020; Hifzurrahman *et al.* 2023; Banerji *et al.* 2022a; Joshi *et al.* 2021a). The intricate geochemical characteristics of fine-grained sediments and sedimentary rocks are fundamentally shaped by their provenance rocks, followed by intricate interactions with fluids that metamorphose source-rock particles into solutes and/or nascent minerals via the mechanisms of weathering (Nesbitt 1979; Nesbitt *et al.* 1980; Taylor and McLennan 1985). Moreover, the compositions of these sedimentary assemblages are malleable, subject to modifications stemming from an array of factors encompassing hydrodynamic sorting, interactions with porewater fluids during the interment process, and the exchange of cations with ambient waters (Fedo *et al.* 1995; Nesbitt *et al.* 1996; Garzanti *et al.* 2009; Garzanti 2016; Lipp *et al.* 2020). Notwithstanding the profusion of expansive geochemical dataset, untangling the multifaceted contributions of these disparate factors to the formation of sediments and sedimentary rocks persists as a formidable enigma. The complexities intertwined with managing such colossal dataset have posed formidable challenges for practitioners in sedimentary geochemistry (Lipp *et al.* 2020). Consequently, in recent times, the acumen of harnessing multivariate statistical approaches (MSA), inclusive of methodologies like principal component analysis (PCA), discriminant analysis, linear discriminant analysis, factor analysis, singular value decomposition, and hierarchical cluster analysis, has emerged as a salient strategy to grapple with the intricacies inherent in navigating extensive data arrays. The PCA is a

multivariate statistical method used for the reduction of dimension and identification of pattern or trend in data (Reid and Spencer 2009) while discriminant analysis discriminates between two or more groups based on their characteristics (Chien and Lautz 2018). As per Braun *et al.* (2013), linear discriminant analysis is a method that identifies linear combination of features to discriminate between two or more groups. Factor analysis is a method that identifies and quantifies underlying factors which explain the observed correlation among a set of variables (Hoseinzade and Mokhtari 2017). Singular value decomposition is related to PCA and is applied to the covariance matrix of the data to find principal components (PCs; Chen *et al.* 2015). Hierarchical cluster analysis is a method of cluster analysis that builds a hierarchy of clusters (similar elements into clusters, Jiang *et al.* 2015).

The assimilation of these statistical techniques has not solely proffered glimpses into the art of deducing provenance (Ohta 2004; Pe-Piper *et al.* 2008; Tolosana-Delgado *et al.* 2018; Armstrong-Altrin 2020; Banerji *et al.* 2022b; McManus *et al.* 2020) but has also deftly illuminated the influence of hydrodynamic fractionation and sorting within the intricate tapestry of the sedimentary environment (Pe-Piper *et al.* 2008). Through systematic mathematical adjustments, the practice of multivariate analysis kindles the revelation of heightened variations within a more tractable dimension scape (Gazley *et al.* 2015). This paradigm duly embraces multiple determinants that contemporaneously impact the data variability (Borvka *et al.* 2005), thus presenting an upper hand over the univariate and bivariate methodologies susceptible to distortions stemming from repetitive statistical trials (Manly1997). Multivariate statistical analysis is a type of statistical analysis that deals with more than two variables. It is used to decipher the correlation among large and complex datasets by reducing the number of variables without any loss of crucial information (Nadiri *et al.* 2013). Multivariate statistical analysis has emerged as a highly advantageous approach across diverse geochemical investigations. Its widespread utilization

encompasses the examination of geochemical data derived from stream sediment, soil, and estuarine sediments, facilitating the detection of mineralization or contamination (Chork and Salminen 1993; Dominech *et al.* 2022; Paternie *et al.* 2023). In the realm of petrology, this methodology has proven pivotal in the discernment of diagenetic processes and the comprehension of how provenance impacts the overarching chemistry of rocks and sediments (Hakstege *et al.* 1992). Furthermore, the application of multivariate statistical analysis has enabled the characterization of fluvial deposits through the scrutiny of intricate and heterogeneous geochemical datasets (Helvoort *et al.* 2005). Notably, Garcia *et al.* (2020) successfully showcased the method's efficacy in differentiating depositional paleo-environments on the basis of geochemical data.

Within the scope of the present study, we have harnessed the PCA technique to reevaluate the previously published geochemical dataset originating from two distinct systems: The Shimla and Chail metasediments (SCM) situated in Himachal Himalayas, Himachal Pradesh (Joshi *et al.* 2021c), as well as the Diu Island mudflat sediments (DMS) located in outhern Saurashtra, Gujarat (Banerji *et al.* 2021a). The primary objective was to unravel the provenance intricacies by underpinning these two systems, one was sedimentary and the other metasedimentary in nature.

2. Principal component analysis (PCA)

In recent times, the remarkable advancement in computational capabilities has ushered in the widespread adoption of the PCA across diverse geoscientific investigations. The origins of PCA trace back to Pearson's groundwork in 1901, which was subsequently refined by Hotelling in 1933. The PCA is a multivariate statistical method used for the reduction of dimensions that is generally used in statistical analysis and machine learning. The prime objective of PCA is to transform high-dimensional data into a lower-dimensional representation, seizing all the valuable data while discarding the less relevant details (Hongyu *et al.* 2016). PCA assumes that the relationships between variables are linear and the data is normally distributed. However, the PCA also has some limitations as it converts the large dimension of data into a square matrix. There may be some data loss as well, as it sometimes reduces the interpretability and thus, it is not suitable for columns having many missing values (Lee 2010).

As a robust multivariate statistical exploratory tool, PCA empowers researchers to navigate data variability adeptly. Its prowess becomes particularly pronounced when grappling with extensive datasets, where intricate interdependencies among variables render interpretation and comprehension a formidable challenge. The core objective of PCA involves the transformation of a comprehensive set of potentially correlated variables into a more succinct set of uncorrelated variables known as PCs. These PCs efficiently encapsulate and preserve the pivotal information within the original dataset (Wishart *et al.* 2013; Sunkari and Abu 2019). Each PC epitomizes a linear amalgamation of the original variables, weighted in accordance with their contributions to elucidate the variance along a specific orthogonal dimension, sequenced in descending order (Geladi and Grahn 1996). Through this mechanism, PCA streamlines data representation while retaining its intrinsic characteristics, thereby simplifying the comprehension and visualization of intricate variable relationships.

Within PCA, the linear and non-linear relationship among samples manifests through scatter plots of scores where each data point corresponds to a distinct sample. Similar chemical attributes lead to the clustering of corresponding samples in proximity. On the converse, the association between different variables is expounded by loadings. These loadings intricately illustrate the fusion of concentration values for diverse chemical elements, forming the basis for the scores. The magnitude and sign of loadings unveil the significance of specific elements in shaping overall variance, thereby unveiling correlations between elements. Variables that coalesce signify positive correlations, while those positioned in diagonally opposing quadrants suggest negative correlations. The inaugural PC (PC1) accounts for the preeminent share of variance within the dataset, representing the primary axis of variability. Successively, the second PC (PC2) captures orthogonal variance, a pattern that endures through subsequent components (Geladi and Grahn 1996; Makvandi *et al.* 2016). Each PC furnishes a distinctive perspective on the data, collectively elucidating the majority of the dataset's variability.

The PCA emerges as most crucial tool in the realm of high-dimensional data (Ueki and Iwamori 2017; Corcoran *et al.* 2019; Henrichs *et al.* 2019), where the trend and patterns of the dataset often elude direct observation, causing the graphical

representation poorly understandable. By condensing the data's dimensionality and highlighting thet pivotal sources of variation, PCA empowers researchers to decode and interpret intricate datasets with greater manageability.

3. Geological background of metasediments and mudflat sediments

In the present study, the geochemical datasets from the metasediments of Lesser Himalaya and mudflat sediment of Diu Island have been studied through the statistical approach of PCA in order to decipher the sediment source and the dominant factor controlling the sedimentary system of the region. The published geochemical datasets of nearly 30 samples each for Shimla and Chail metasediment (hereafter SCM) from Lesser Himalayas (Joshi *et al.* 2021a) and Diu Island mudflat sediments (hereafter DMS) core (Banerji *et al.* 2021a) were studied and analysed through PCA approach. By embarking upon the PCA analysis on the metasediments (SCM) and mudflat sediments (DMS), we aim to investigate the plausible source and mechanism responsible for the intricate patterns and correlation in the geochemical variables.

3.1 *Shimla and Chail metasediments (SCM)*

The expansive Himalayan mountain range has been methodically subdivided into distinctive litho-tectonic units for the purpose of geological classification. Among these units are the Sub-Himalaya, Lesser Himalaya Sediments (LHS), Lesser Himalayan Crystalline sequence (LHCS), Higher Himalaya Crystalline Sequence (HHCS), and Tethyan Himalaya (Chambers *et al.* 2008; Bhargava *et al.* 2011; Law *et al.* 2013). Notably, Ahmad *et al.* (2000), employing isotopic markers, further refined the LHS into the relatively youthful outer zone and the elder inner zone. The outer Lesser Himalaya sediments were attributed to a dominant provenance from Meso- to Neo-proterozoic sources. Interestingly, both outer LHS and HHCS sediments exhibited congruent depositional ages and isotopic traits, hinting at a shared origin (Parrish and Hodges 1996; Ahmad *et al.* 2000; Richards *et al.* 2005). In contrast, the inner Lesser Himalaya sediments emerged as products largely sourced from Late-Archean to Paleoproterozoic origins.

Within this intricate geological landscape, the Chail, Shimla (Simla), and Blaini series exist as moderate to weakly metamorphosed strata, lying beneath the medium-grade Jutogh group of rocks. A stratigraphic marker in the form of stromatolite-bearing limestone horizons led Raha and Sastry (1982) to propose an upper Riphean age for the Shimla group, with $^{40}Ar/^{39}Ar$ mica dating yielding a maximum depositional age of 860 Ma (Frank 2001; Bhargava *et al.* 2011). The Shimla group exhibits a diverse composition comprising slates, greywackes, quartzites, and carbonates, with origins attributed to sedimentation via north-directed turbidity currents (Valdiya 1970; Srikantia and Sharma 1971, 1976; Sinha 1978; Joshi *et al.* 2021b). In contrast, the Chail Group (figure 1a) encompasses phyllites, phyllitic quartzite, psammitic and pelitic schists, orthoquartzites, arkose, chlorite schist, limestones, and meta-basic rocks, affiliating it with the Lesser Himalaya (Valdiya 1980). The outer LHS sediments, comprising formations like Chail, Tal, Krol, and others (figure 1), have been ascribed Neoproterozoic to Cambrian depositional ages (Richards *et al.* 2005).

3.2 *Diu Island mudflat sediments (DMS)*

The mudflat of Diu Island is located along the southern Saurashtra coast of western Gujarat (figure 1). Majority of the Saurashtra peninsula comprises of a basalts and its derivatives belonging to the Deccan Trap Formation of upper Cretaceous period (Bhonde and Bhatt 2009). Unlike the Deccan plateau of west-central India, the Saurashtra Deccan basalts can be differentiated based on its tholeiitic flood basalts thickness (Najafi *et al.* 1981), dominance of granophyre and rhyolite, volcano plutonic complexes (Naushad *et al.* 2019) and pervasive compositional variety (Melluso *et al.* 1995; Sheth *et al.* 2011, 2012a). The Saurashtra Deccan basalts are unconformably overlain by Gaj, Dwarka formations of Tertiary and Miliolites, Chaya formations, Katpur and Mahuva formations of Quaternary Period (Pandey *et al.* 2007). Gaj Formation consists of marly limestone which is rich in fossils (foraminifera, echinodermata, lamellibranch and gastropods) and is well exposed near Una–Veraval Road. Dwarka Formation is only exposed near Jafrabad, SW of Diu Island (Verma and Mathur 1979a). Miliolite are considered to be of both marine as well as aeolian and are distinguished based on their sedimentary structures and
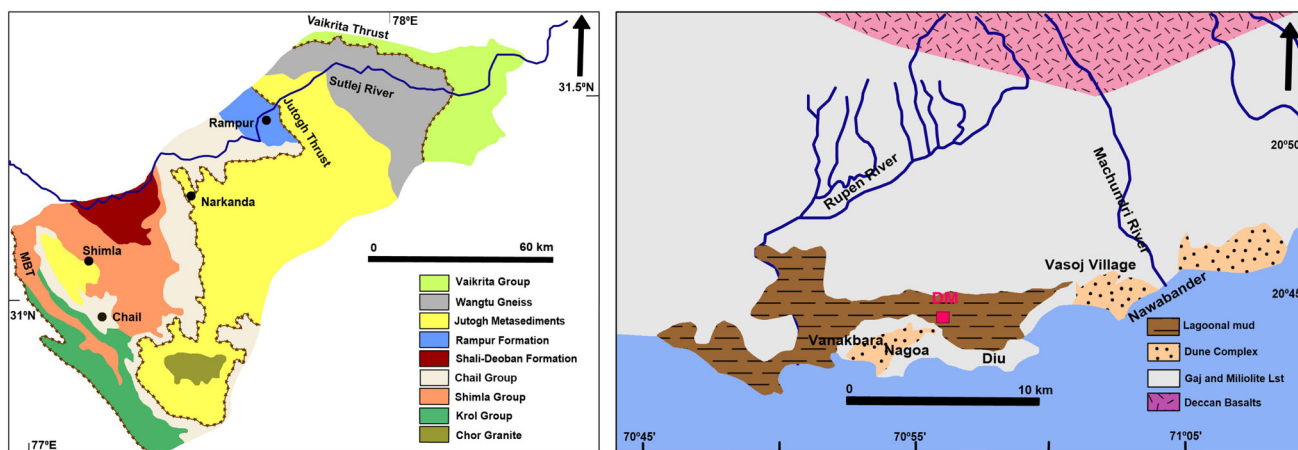
Figure 1. The geological map of the southern Saurashtra coast surrounding the active mudflat of Diu, Gujarat, modified after Banerji *et al.* (2019) and Pant and Juyal (1993), filled square indicates sampling site (DM) (Banerji *et al.* 2021a, b) and modified geological map of Sutlej section of Himachal Lesser Himalaya (Thakur 1992; Vannay and Grasemann 1998; Richards *et al.* 2005) sample locations shown as red square after Joshi *et al.* (2021a, b, c).

quantitative faunal characteristics (Verma 1982). They includes only pelletoid (and oolitic) calcarenites and associated micrites but devoid of megafossils (Verma and Moitra 1975). The miliolite exposures are found near the Machundri river section (Verma and Mathur 1979b). The coastal rocks such as the dead coral reefs, oyster beds and other highly fossiliferous limestone are included in the Chaya Formation. The age of the Chaya Formation ranges from the late Pleistocene to the Holocene (Gupta 1972; Gupta and Amin 1974). Katpur and Mahuva formations were deposited during the Holocene epoch wherein the former includes oxidized and pedoogenised tidal flat clays/silts while the latter comprises freshwater alluvium (sand and clays), coastal deposits, lime mud, calcareous sand with marine shells (Mathur *et al.* 1987; Bhatt 2003; Pandey *et al.* 2007).

4. Computational methodology

PCA is a robust statistical technique that leverages orthogonal transformations to transmute an assortment of potentially interrelated observations into an array of linearly uncorrelated variables. These newly formed uncorrelated variables, commonly referred to as PCs, act as crucial tools in rendering intricate high-dimensional datasets into easily recognisable 2D or 3D patterns. A series of systematic steps unfold when conducting PCA on a matrix characterized by n variables and m samples, encapsulating the following convoluted progression.

4.1 *Data preparation and mean centering*

Standardization emerges as a pivotal phase within the PCA, serving to grant equanimity to dissimilar variables with divergent scales in contributing to the analysis. Through standardization, parity is established, encompassing a uniform range and data variability for all variables. This process of standardization unfolds in two essential steps. Initially, data is standardized by aligning each variable onto a shared scale; subsequently, data is centered by adjusting it in relation to the means of each variable. This centering maneuver situates the data at the origin of the PCs. The method of standardization employed varies depending on data characteristics. For instance, integration of the median or median absolute deviation can prove instrumental in mitigating the sway of outliers within the dataset. Notably, when grappling with geochemical datasets, meticulous attention to analytical uncertainties is warranted for their proper integration into the analysis.

In certain scenarios, the raw data necessitates preprocessing via log ratios, particularly when grappling with data constrained by constants like percentages or parts per million (Aitchison1986). This transformation guarantees conformity to the imposed constraints, rendering the data amenable to PCA. Within the scope of this study, the preferred standardization methodology revolves around the mean and standard deviation for its simplicity and effectiveness. The standardization process is realized through equation (1), thereby ensuring data achieves the requisite

standardization, thereby priming it for subsequent PCA exploration.

$$Standardization = \left(\frac{Mean\,(data)}{Standard\ deviation\,(data)} \right).$$
(1)

The mean is a measure of central tendency that represents the average value of a set of numbers. It is obtained by summing up all the values in a dataset and then dividing that sum by the total number of data points. The mean is located between the median (the middle value in a sorted list of data) and the mode (the most frequently occurring value). The formula for calculating the mean of a dataset is as follows:

$$Mean = \sum_{i=1}^{n} \frac{X_i}{n},$$
(2)

where X_i is the ith element of the individual data points of variable X and X_m is the mean of X variables while n represents the number of elements.

The standard deviation is a statistical measure that quantifies the spread or dispersion of a dataset in relation to its mean value. It is computed as the square root of the variance, which represents the average squared deviation of each data point from the mean. By determining the distance of each data point from the mean, the standard deviation assesses how much the values in the dataset deviate from the average value. The formula to calculate the standard deviation is as follows:

$$Standard\ deviation = \sqrt{\sum_{i=1}^{n} \frac{(X_i - X_m)^2}{n-1}}.$$
(3)

Adjustment refers to a series of processes undertaken to enhance the classification, timing, valuation or coverage of data. It also involves adapting data to a specific recording or accounting basis and addressing any discrepancies in data quality during the assembly of dataset. To carry out data adjustment, we employ a specific formula or method (equation 4) that helps to modify the original data to better suit the intended analysis or reporting requirements. The adjustment process aims to ensure the accuracy and reliability of the data for further analysis and interpretation.

$$Adjusting\ data = \sum_{i=1}^{n} (X_i - X_m)$$
(4)

where the term X_i represents the ith element of the individual data points of the variable X. X_m denotes the mean of the X variable, which is the average value of all the data points in the dataset. The variable n represents the total number of elements in the dataset, reflecting the size of the data sample.

4.2 *Variance and covariance*

The scaled data obtained from section 4.1 is utilized to compute the covariance, which measures the relationship between two different datasets in terms of their positive and negative values. A positive covariance suggests that the variables tend to increase and decrease together, while a negative covariance indicates that the two variables vary in opposite directions. The covariance analysis also provides insights into the spatial relationship and variance of the dataset concerning different variables. The covariance between any two variables, X and Y can be calculated using the following equations:

$$\mathrm{Var}(X, X) = \sum_{i=1}^{n} \frac{(X_i - X_m)^2}{n-1}$$
(5)

and

$$\mathrm{Cov}(X, Y) = \sum_{i=1}^{n} \frac{\{(X_i - X_m)(Y_i - Y_m)\}}{n-1}$$
(6)

where the term X_i represents the ith element of the individual data points of the variable X. X_m denotes the mean of the X variable; Y_i represents the ith element of the individual data points of the variable Y, and Y_m denotes the mean of the Y variables; the variable n represents the total number of elements in the dataset, reflecting the size of the data sample.

4.3 *Eigen decomposition*

The covariance matrix provides the necessary information to calculate the eigenvalues and eigenvectors, which are essential in PCA. These eigenvalues and eigenvectors play a crucial role in representing the overall variability of the dataset. Eigenvalues and eigenvectors always come in pairs, where eigenvalues determine the magnitude or importance of each PC, and eigenvectors demonstrate the direction of the data with the largest variance in the dataset. The eigenvector associated with the highest eigenvalue corresponds to the first

PC, which accounts for the greatest possible variance in the dataset. Subsequent PCs have progressively lower variances, capturing less and less of the total variability in the data. Eigenvalues and eigenvectors can be calculated using the following procedure:

$$\det(A - \lambda I)X = 0 \qquad (7)$$

where A is a covariance matrix, I is the identity matrix, λ is the eigenvalue, and X is the eigenvector matrix.

4.4 *Selection of principal components*

After computing the eigenpairs (eigenvalues and eigenvectors), it is necessary to sort them based on the magnitude of their eigenvalues. This sorting process allows us to select the desired number of PCs with higher scores and loadings, which are more significant for dimensionality reduction. Typically, the eigenvectors with higher eigenvalues are chosen as the feature vectors, as they capture the most important information about the data. This selection can be accomplished by plotting the cumulative sum of the eigenvalues and identifying the point where the explained variance reaches a satisfactory level. Once the desired PCs (feature vectors) are identified, the transformed feature vector is multiplied with the transformed, adjusted data (the data centered around the means) to reconstruct the original data in the new lower-dimensional space. This transformation (equation 8) helps to retain the maximum relevant information about the original data while reducing its dimensionality, enabling easier visualization and analysis.

$$\begin{aligned} Final\ data = {}& Row\ feature\ vector \\ & \times\ Row\ data\ adjusted \end{aligned} \qquad (8)$$

where row feature vector is the eigenvectors transposed, and row data adjusted is the mean adjusted data of the original data transposed.

In this study, a MATLAB-based computational algorithm is developed to compute RQ-mode PCA, following the steps mentioned in flowchart diagrams (figures 2 and 3). R-mode PCA is primarily based on variables and is suitable for identifying associations between variables and a set of observations (elements). It processes the covariance matrix and creates new orthogonal linear combinations that preserve the variance of the original variables. These new combinations account for successively decreasing portions of the variance, allowing for dimensionality reduction. On the other hand, Q-mode PCA is primarily based on observations (samples) and is suitable for characterizing samples. It analyzes the covariance matrix to identify patterns and relationships among samples.

RQ-mode PCA is a method that calculates both variables and object loadings simultaneously, combining aspects of both R-mode and Q-mode PCA. For this particular study, RQ-mode PCA computation is chosen due to its detailed analysis of sedimentary processes. It allows for the characterization of different elements and their associations with the process under investigation. Furthermore, it enables the identification of different rock types based on geochemical datasets. Using RQ-mode PCA, the researchers can gain valuable insights into the complex relationships and patterns within the geochemical dataset, aiding in the understanding and interpretation of sedimentary processes and rock types.

5. Results and discussion

In the present study, the PCA was applied to two distinct sets of geochemical datasets, namely, the SCM and DMS, with the aim of understanding the dominant processes influencing these distinct sites. For the SCM dataset, a total of 30 geochemical data points, including major and trace elemental compositions, were analysed from the previous study (Joshi *et al.* 2021c). The major and trace elements retain important cues to sedimentary processes in the SCM locality. In case of the DMS dataset, major elements and selected trace elements were used to delineate the processes acting on the region on a temporal scale (Banerji *et al.* 2021a). The variations in elemental compositions in the DMS dataset are influenced by various geochemical proxies, such as *in-situ* productivity, paleo-weathering, and sediment source. The detailed implications of these geochemical proxies on a temporal scale are discussed in Banerji *et al.* (2021a).

The developed algorithm was subsequently implemented on the SCM and DMS datasets, and the eigenvalues of the PCs were calculated. The contributions of each PC to the total variance of the dataset were estimated, and these results are presented in the Supplementary file (tables S1 and S2). These contributions provide valuable insights into the importance of each PC in explaining the variability within the dataset and can help to identify the most significant factors or processes influencing the SCM and DMS localities.

This research helps us identify the major geochemical factors that influence sediment composition and indicate the sedimentary settings in which the sediments formed. We must normalize the data before doing the PCA in order to avoid any kind of error during the analysis and to make the contribution of each variable proportional to the analysis; otherwise, it might influence the geochemical trends.

5.1 *Shimla and Chail metasediments (SCM)*

In PCA of the SCM dataset, a scree plot (figure 4a) was generated, showing a total of 29 PCs. The

Figure 2. The process to carry out the PCA of any given dataset in general.

scores of the observations were depicted as symbols, while the loadings of the different elements were plotted in figure 4(b and c). From the scree plot, it was evident that the first nine PCs showed an elbow point, which collectively accounted for 85.53% of the total variance. The contributions of these nine PCs were as follows: PC1 (20.86%), PC2 (19.75%), PC3 (11.90%), PC4 (7.42%), PC5 (7.06%), PC6 (5.56%), PC7 (5.35%), PC8 (4.67%), and PC9 (2.96%).

Considering that most of the PCs had contributions <10%, we focused on explaining the total variability of the data using the first three PCs: PC1, PC2, and PC3. These three PCs together accounted for 52.51% of the total variation in the dataset. Additionally, for simplicity and to highlight the most significant relationships, only two combinations of PCs were taken into consideration: PC1 *vs.* PC2 and PC2 *vs.* PC3 (figure 4b and c). These plots reveal the major patterns and associations between variables in the dataset. By selecting the first three PCs and plotting these specific combinations, the researchers aimed to capture the most important information while reducing the complexity of the analysis.

Upon careful analysis of the contribution from PC1, it becomes clear that a significant number of major oxides and trace elements in the dataset demonstrate a positive correlation with PC1.
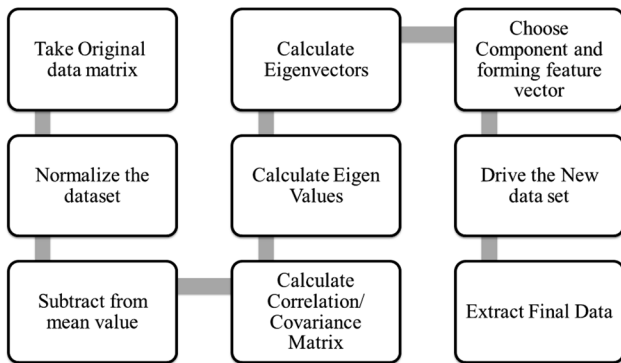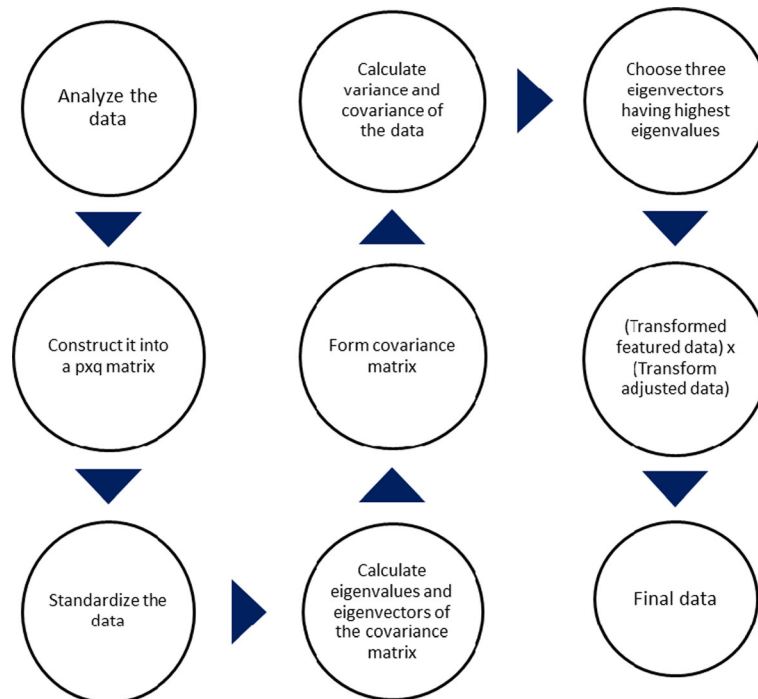

Figure 3. The process to carry out the PCA of the given dataset in Matlab (where *p* is the number of samples and *q* is the number of variables).
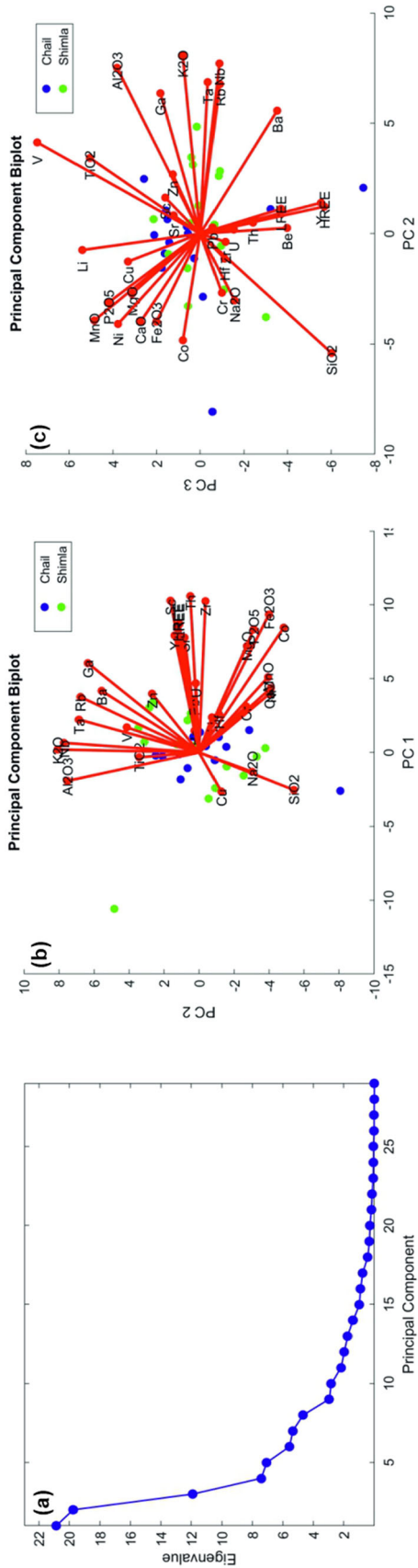
Figure 4. (**a**) Scree plot for eigenvalues against PC, (**b**) PC1 *vs.* PC2 biplot, and (**c**) PC2 *vs.* PC3 biplot for metasediments from Chail–Shimla group.
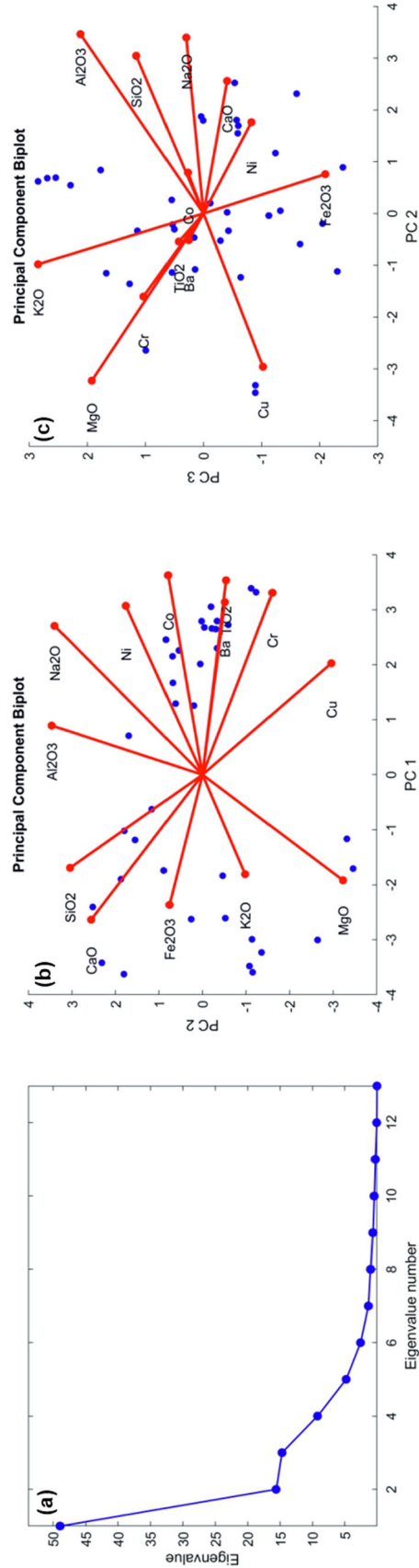

Figure 5. (**a**) Scree plot for eigenvalues against PC, (**b**) PC1 *vs.* PC2 biplot, and (**c**) PC2 *vs.* PC3 biplot for Diu mudflats.

Additionally, PC1 exhibits positive loadings for elements such as K, Rb, Th, Ba, and LREEs (Light Rare Earth Elements), which are typically considered incompatible elements and are indicative of rocks with felsic composition. Interestingly, SiO_2 shows a slight negative loading on PC1. This, in combination with the positive loadings for incompatible elements, might suggest an intermediate source for the rocks in the dataset. Furthermore, major oxides and trace elements that have a strong affinity with mafic to intermediate rocks exhibit positive loadings on both PC1 and PC3, while showing negative loadings on PC2 (as seen in figure 4b and c). These loading patterns provide valuable insights into the relationships and characteristics of different rock types present in the dataset, helping to identify their composition and potential sources.

Due to their higher compatibility, K_2O, Na_2O, and CaO are typically enriched in feldspars. The relative enrichment of K_2O and depletion of Na_2O along PC1 and PC2 suggest that K-feldspar is the primary repository of potassium and the predominant feldspar in the SCM, as compared to sodic plagioclase. This observation aligns with the higher K_2O/Na_2O ratios found in bulk rock geochemistry (Joshi *et al.* 2021b). Furthermore, the close association of Al_2O_3 and TiO_2 along the positive PC2 axis suggests that phyllosilicates are the main carriers of these elements in the SCM. The fact that both phyllosilicates and K-feldspar display positive loadings along PC2 and PC3 further supports their role as major reservoirs for K_2O, Al_2O_3, and TiO_2, as also noted by Joshi *et al.* (2021b) based on oxide correlations.

Studies have indicated that heavy minerals, such as zircon, apatite, and titanite, which possess higher partition coefficients for Rare Earth Elements (REEs), can influence the concentration of trace elements in the SCM (Armstrong-Altrin *et al.* 2012). The enrichment of Light Rare Earth Elements (LREEs) and Heavy Rare Earth Elements (HREEs) along both PC1 and PC2 suggests that these accessory minerals control the REE budget of the SCM. The distribution of least mobile incompatible elements, such as REEs, HFSEs (High Field Strength Elements), Th and Y, can reflect the provenance of the sediments and help differentiate between various lithologies (McLennan 1989; Cullers 1994; Taylor and McLennan 1995; Large *et al.* 2018). The enrichment of Th, U, Zr, and Sc along positive PC1 and PC2 suggests the influence of reworking and recycling of felsic to intermediate sources. The slight negative loading of SiO_2 with PC1, along with the positive loadings of MgO, Fe_2O_3, Co, Ni, Th, and U, might indicate the possible contribution of intermediate rocks as a source for the studied sediments. These findings shed light on the origin and composition of the SCM sediments, providing valuable insights into the processes that have shaped their geochemical characteristics.

5.2 *Diu mudflat sediments (DMS)*

In the PCA of the DMS dataset, a scree plot (figure 5a) displayed a total of 13 PCs. The scree plot revealed that the first three PCs showed an elbow point and collectively accounted for 79.30% of the total variance. The contributions of these three PCs were as follows: PC1 (48.94%), PC2 (15.64%), and PC3 (14.72%). Due to the comparable variance of PC2 and PC3, only two combinations of PCs were considered for further analysis: PC1 *vs.* PC2 and PC2 *vs.* PC3 (figure 5b and c). The biplot for these combinations illustrates the scores of the observations displayed as symbols, and the loadings of the different elements are plotted. Upon careful analysis of the contribution from PC1, it becomes evident that oxides account for most of the variations in this component compared to the other elements. Combining PC1 and PC2 accounts for a significant portion (64.58%) of the total variability in all datasets within this group. These findings provide valuable insights into the major factors contributing to the variations in the DMS dataset and allow for a better understanding of the geochemical characteristics of this region.

The positive correlation of Total Organic Carbon (TOC) and Cu with both PC1 and PC2 is significant in the PCA of the DMS dataset. In marine, coastal, and lacustrine sediments, TOC is widely regarded as a significant indicator of *in-situ* productivity (Tribovillard *et al.* 2006; Chandana *et al.* 2017; Banerji *et al.* 2019, 2021b). However, TOC is susceptible to degradation over time. On the other hand, Cu is delivered to the sediments through organometallic complexes and serves as an additional proxy for *in-situ* productivity. The close association between TOC and Cu, as well as their enrichment towards the positive axis of both PC1 and PC2, indicates that *in-situ* productivity has played a crucial role in shaping the geochemical variations observed in the DMS dataset. The positive correlation of TOC and Cu with these PCs suggests that variations in *in-situ*

productivity have had a significant impact on the geochemical composition of the sediments in the DMS region. These findings provide valuable insights into the environmental conditions and processes that have influenced the sedimentary characteristics of the studied area.

The fact that some elements like Cu, Ba, TiO_2, Co, and Ni are more abundant along the positive axis of PC1 suggests that similar lithologies are involved. Hayashi *et al.* (1997) found important minerals like olivine, pyroxene, hornblende, biotite, and ilmenite with TiO_2. The ferromagnesian trace elements Cr, Ni, and Co generally exhibit a similar behaviour during the magmatic processes, although weathering may result in their fractionation (Feng and Kerrich 1990). Nevertheless, they are more abundant in mafic igneous rocks and their associated weathering products (Armstrong-Altrin *et al.* 2004; Joshi *et al.* 2021c). The simultaneous enrichment of Ni and Cr in the floodplain sediments of the Cauvery River has been interpreted as a possible indication of a mafic origin (Singh and Rajamani 2001). Furthermore, the concurrent enrichment of Fe_2O_3, CaO, and SiO_2 along the negative axis of PC1 and PC2 demonstrated the possible influence of an intermediate rock source. The hinterland of the Saurashtra peninsula is comprised of Deccan basalts, trachyte, rhyolite, granophyre, and pitchstone dykes, which are associated with mafic dolerite dykes at Sirohi-Palitana (Chatterjee and Bhattacharji 2001) and Picritic dykes at Dedan (Krishnamacharlu 1972). In addition, scientists have found granophyre, rhyolite, and obsidian at Barda (Cucciniello *et al.* 2019) and a sequence of rhyolite, pitchstone, and basaltic andesite lava flow at Osham (Sheth *et al.* 2012). A combination of different types of rock and other sources (Banerji *et al.* 2021a) must have caused the intermediate and mafic rock signatures in the DMS geochemical data.

The enrichment of K_2O and MgO along the positive axis of PC2 and PC3 indicates enhanced weathering intensities. Climate plays a pivotal role in sediment weathering, while other factors, such as the nature of source rocks, microbes, and relief, also significantly influence the geochemical composition of sediments (Nesbitt and Young 1982; Taylor and McLennan 1985; McLennan *et al.* 1993; Joshi 2014; Madhavaraju *et al.* 2016). Notably, K_2O and MgO normalized with Al_2O_3 are extensively used as paleo-weathering proxies in sediment cores from coastal, marine, and lake environments (Banerji *et al.* 2017, 2019, 2021b; Bhushan *et al.*

2018). Furthermore, the positive axis of PC3 reveals enrichments in Al_2O_3, SiO_2, and Na_2O, suggesting a prevalence of clayey textures derived from terrestrial sources, particularly plagioclase minerals. These findings emphasize the significant contribution of sediment sources originating from the hinterland of the Saurashtra peninsula.

In summary, the control of the PC1 is mainly attributed to *in-situ* productivity and the mafic source, with a smaller contribution from the intermediate source. PC2 is influenced by weathering proxies, while PC3 is predominantly governed by the clayey fraction originating from plagioclase minerals found in the Saurashtra peninsula. These factors collectively shape the geochemical composition and variations observed in the sediments under study.

6. Conclusions

Geological processes govern the elemental assemblages derived from geochemical datasets, which pose a challenge due to the vast amount of data reflecting various geochemical processes. In our study investigating the provenance using high-dimensional bulk-sediment geochemical data from Lesser Himalayan rocks (Shimla and Chail groups) and Diu mudflats, we draw the following conclusions:

- In the metasediments of the LHS region, the examination of the eigenvectors of PCs reveals that accessory minerals play a crucial role in controlling the trace element budget of the metasediments. Additionally, the presence of reworking and recycling of felsic to intermediate sources is suggested for the studied sedimentary cover in the Lesser Himalayan region.
- In the Diu mudflats, the first three PCs indicate an intermediate to mafic source associated with processes involving olivine and pyroxene. The outcome of present work invokes the significance and applicability of PCA on the high-dimensional geochemical datasets in Geosciences.

Author statement

DS is responsible for the computation and preparation of the initial draft. CPD, USB, and KBJ are responsible for conceptualizing the work and modifying the draft to its final stage. USB and KBJ are responsible for data generation and visualization.

References

Ahmad T, Harris N and Bickle M 2000 Isotopic constraints on the structural relationships between the Lesser Himalayan Series and the High Himalayan Crystalline Series, Garhwal Himalaya; *Bull. Geol. Soc. Am.* **112** 467–477, https://doi.org/10.1130/0016-7606(2000)112%3c467:ICOTSR%3e2.0.CO;2.

Armstrong-Altrin J S 2020 Detrital zircon U–Pb geochronology and geochemistry of the Riachuelos and Palma Sola beach sediments, Veracruz State, Gulf of Mexico: A new insight on palaeoenvironment; *J. Palaeogeogr.* **9** 1–27.

Armstrong-Altrin J S, Il Lee Y and Kasper-Zubillaga J J 2012 Geochemistry of beach sands along the western Gulf of Mexico, Mexico: Implications for provenance; *Geochemistry* **72** 345–362.

Armstrong-Altrin J S, Lee Y I, Verma S P and Ramasamy S 2004 Geochemistry of sandstones from the Upper Miocene Kudankulam Formation, southern India: Implications for provenance, weathering, and tectonic setting; *J. Sedim. Res.* **74** 285–297, https://doi.org/10.1306/082803740285.

Banerji U S, Bhushan R and Joshi K B 2021a Hydroclimate variability during the last two millennia from the mudflats of Diu Island, western India; *Geol. J.* **56** 3584–3604, https://doi.org/10.1002/gj.4116.

Banerji U S, Shaji J and Arulbalaji P 2021b Mid-late Holocene evolutionary history and climate reconstruction of Vellayani lake, South India; *Quat. Int.* **599–600** 72–94, https://doi.org/10.1016/j.quaint.2021.03.018.

Banerji U S, Bhushan R and Jull A J T 2019 Signatures of global climatic events and forcing factors for the last two millennia on the active mudflats of Rohisa, southern Saurashtra, Gujarat, western India; *Quat. Int.* **507** 172–187, https://doi.org/10.1016/j.quaint.2019.02.015.

Banerji U S, Bhushan R and Jull A J T 2017 Mid−late Holocene monsoonal records from the partially active mudflat of Diu Island, southern Saurashtra, Gujarat, western India; *Quat. Int.* **443** 200–210, https://doi.org/10.1016/j.quaint.2016.09.060.

Banerji U S, Dubey C P, Goswami V and Joshi K B 2022a Geochemical indicators in provenance estimation; In: Geochemical Treasures and Petrogenetic Processes (eds) Armstrong-Altrin J S, Pandarinath K and Verma S K, Springer, Singapore. https://doi.org/10.1007/978-981-19-4782-7_5.

Banerji U S, Joshi K B, Pandey L and Dubey C P 2022b An outline of geochemical proxies used on marine sediments deposited during the Quaternary Period; In: *Stratigraphy and timescales,* Academic Press, Vol. **7**, pp. 1-35.

Bhargava O N, Frank W and Bertle R 2011 Late Cambrian deformation in the Lesser Himalaya; *J. Asian Earth Sci.* **40** 201–212, https://doi.org/10.1016/j.jseaes.2010.07.015.

Bhatt N 2003 The Late Quaternary bioclastic carbonate deposits of Saurashtra and Kachchh, Gujarat, western India: A review; *Proc. Indian Nat. Sci. Acad.* **69** 137–150.

Bhonde U and Bhatt N 2009 Joints as fingerprints of stress in the Quaternary carbonate deposits along coastal Saurashtra, western India; *J. Geol. Soc. India* **74** 703–710.

Bhushan R, Sati S P and Rana N 2018 High-resolution millennial and centennial scale Holocene monsoon variability in the Higher Central Himalayas; *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **489** 95–104, https://doi.org/10.1016/j.palaeo.2017.09.032.

Boruvka L, Vacek O and Jehlička J 2005 Principal component analysis as a tool to indicate the origin of potentially toxic elements in soils; *Geoderma* **128** 289–300.

Braun M, Hubay K, Magyari E, Veres D, Papp I and Bálint M 2013 Using linear discriminant analysis (LDA) of bulk lake sediment geochemical data to reconstruct late glacial climate changes in the South Carpathian Mountains; *Quat. Int.* **293** 114,122.

Chambers J A, Argles T W and Horstwood M S A 2008 Tectonic implications of Palaeoproterozoic anatexis and Late Miocene metamorphism in the Lesser Himalayan Sequence, Sutlej Valley, NW India; *J. Geol. Soc. London* **165** 725–737.

Chandana K R, Bhushan R and Jull A J T 2017 Evidence of poor bottom water ventilation during LGM in the Equatorial Indian Ocean; *Front. Earth Sci.* **5** 84, https://doi.org/10.3389/feart.2017.00084.

Chatterjee N and Bhattacharji S 2001 Origin of the felsic and basaltic dikes and flows in the Rajula–Palitana–Sihor area of the deccan traps, Saurashtra, India: A geochemical and geochronological study; *Int. Geol. Rev.* **43** 1094–1116, https://doi.org/10.1080/00206810109465063.

Chen Y, Zhang L and Zhao B 2015 Application of singular value decomposition (SVD) in extraction of gravity components indicating the deeply and shallowly buried granitic complex associated with tin polymetallic mineralization in the Gejiu tin ore field, Southwestern China; *J. Appl. Geophys.* **123** 63–70.

Chien N P and Lautz L K 2018 Discriminant analysis as a decision-making tool for geochemically fingerprinting sources of groundwater salinity; *Sci. Total Environ.* **618** 379–387.

Chork C Y and Salminen R 1993 Interpreting exploration geochemical data from Outokumpu, Finland: A MVE-robust factor analysis; *J. Geochem. Explor.* **48** 1–20.

Corcoran L, Simonetti A and Spano T L 2019 Multivariate analysis based on geochemical, isotopic, and mineralogical compositions of uranium-rich samples; *Minerals* **9**, https://doi.org/10.3390/min9090537.

Cucciniello C, Choudhary A K, Pande K and Sheth H 2019 Mineralogy, geochemistry and ^{40}Ar–^{39}Ar geochronology of the Barda and Alech complexes, Saurashtra, northwestern Deccan Traps: Early silicic magmas derived by flood basalt fractionation; *Geol. Mag.* **156** 1668–1690, https://doi.org/10.1017/S0016756818000924.

Cullers R L 1994 The controls on the major and trace element variation of shales, siltstones, and sandstones of Pennsylvanian–Permian age from uplifted continental blocks in Colorado to platform sediment in Kansas, USA; *Geochim. Cosmochim. Acta* **58** 4955–4972.

Dominech S, Albanese S, Guarino A and Yang S 2022 Assessment on the source of geochemical anomalies in the sediments of the Changjiang river (China), using a modified enrichment factor based on multivariate statistical analyses; *Environ. Pollut.* **313** 120126.

Fedo C M, Nesbitt H W and Young G M 1995 Unravelling the effects of potassium metasomatism in sedimentary rocks and paleosols, with implications for paleoweathering conditions and provenance; *Geology* **23** 921–924, https://doi.org/10.1130/0091-7613(1995)023%3c0921:UTEOPM%3e2.3.CO.

Feng R and Kerrich R 1990 Geochemistry of fine-grained clastic sediments in the Archean Abitibi greenstone belt, Canada: Implications for provenance and tectonic setting; *Geochim. Cosmochim. Acta* **54** 1061–1081, https://doi.org/10.1016/0016-7037(90)90439-R.

Frank W 2001 A review of the Proterozoic in the Himalaya and the northern Indian Shield; *J. Asian Earth Sci.* **19** 17–18.

Garcia R J L, da Silva Júnior J B and Abreu I M 2020 Application of PCA and HCA in geochemical parameters to distinguish depositional paleoenvironments from source rocks; *J. South Am. Earth Sci.* **103** 102734.

Garzanti E 2016 From static to dynamic provenance analysis-sedimentary petrology upgraded; *Sedim. Geol.* **336** 3–13, https://doi.org/10.1016/j.sedgeo.2015.07.010.

Garzanti E, Andò S and Vezzoli G 2009 Grain-size dependence of sediment composition and environmental bias in provenance studies; *Earth Planet Sci. Lett.* **277** 422–432.

Gazley M F, Collins K S and Roberston J 2015 Application of principal component analysis and cluster analysis to mineral exploration and mine geology; In*: AusIMM New Zealand branch annual conference, Dunedin, New Zealand,* pp. 131–139.

Geladi P and Grahn H 1996 *Multivariate Image Analysis*; John Wiley and Sons Inc., New York.

Gupta S K 1972 Chronology of the raised beaches and Inland Coral Reefs of the Saurashtra Coast; *J. Geol.* **80** 357–361, https://doi.org/10.1086/627738.

Gupta S K and Amin B S 1974 Io/U ages of corals from Saurashtra coast; *Mar. Geol.* **16** 79–83.

Hakstege A L, Kroonenberg S B and Van Wijck H 1992 Geochemistry of Holocene clays of the Rhine and Meuse rivers in the central-eastern Netherlands; *Geol. Mijnb.* **71** 301–315.

Haughton P D W, Todd S P and Morton A C 1991 Sedimentary provenance studies; *Geol. Soc. London, Spec. Publ.* **57** 1–11.

Hayashi K-I, Fujisawa H, Holland H D and Ohmoto H 1997 Geochemistry of ∼1.9 Ga sedimentary rocks from northeastern Labrador, Canada; *Geochim. Cosmochim. Acta* **61** 4115–4137.

Henrichs I A, Chew D M and O'Sullivan G J 2019 Trace element (Mn–Sr–Y–Th–REE) and U–Pb isotope systematics of metapelitic apatite during progressive greenschist- to amphibolite-facies Barrovian metamorphism; *Geochem. Geophys. Geosyst.* **20** 4103–4129.

Hifzurrahman, Nasipuri P and Joshi K B 2023 Geochemistry of Jutogh Metasediments, Lesser Himachal Himalaya, India, and their Implications in source area weathering, provenance, and tectonic setting during Paleoproterozoic Nuna Assembly; *J. Geol. Soc. India* **99** 897–905. https://doi.org/10.1007/s12594-023-2411-0.

Hoseinzade Z and Mokhtari A R 2017 A comparison study on detection of key geochemical variables and factors through three different types of factor analysis; *J. Afr. Earth Sci.* **134** 557–563.

Hongyu K, Sandanielo V L M and de Oliveira Junior G J 2016 Análise de componentes principais: Resumo teórico, aplicação e interpretação; *E&S Eng. Sci.* **5(1)** 83–90.

Jiang Y, Guo H, Jia Y, Cao Y and Hu C 2015 Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia; *Geochemistry* **75**(2) 197–205.

Joshi K B 2014 Microbes: Mini iron factories; *Indian J. Microbiol.* **54**, https://doi.org/10.1007/s12088-014-0497-1.

Joshi K B, Bhattacharjee J and Rai G 2017 The diversification of granitoids and plate tectonic implications at the archaean–Proterozoic boundary in the Bundelkhand Craton, central India; *Geol. Soc. Spec. Publ.* **449** 123–157, https://doi.org/10.1144/SP449.8.

Joshi K B, Banerji U S, Dubey C P and Oliveira E P 2021a Heavy minerals in provenance studies: An overview; *Arabian J. Geosci.* **14** 1–16.

Joshi K B, Ray S and Ahmad T 2021b Geochemistry of metasediments from Neoproterozoic Shimla and Chail Group of Outer Lesser Himalaya: Implications for provenance, tectonic setting and paleo-weathering conditions; *Geol. J.*, https://doi.org/10.1002/gj.4183.

Joshi K B, Ray S and Ahmad T 2021c Geochemistry of metasediments from Neoproterozoic Shimla and Chail Groups of Outer Lesser Himalaya: Implications for provenance, tectonic setting, and paleo-weathering conditions; *Geol. J.*, https://doi.org/10.1002/gj.4183.

Joshi K B, Banerji U S, Dubey C P and Oliveira E P 2022a Detrital zircons in crustal evolution: A perspective from the Indian subcontinent; *Lithosphere* **2022** 3099,822.

Joshi K B, Singh S K and Halla J 2022b Neodymium isotope constraints on the origin of TTGs and high-K granitoids in the Bundelkhand Craton, central India: Implications for Archaean crustal evolution; *Lithosphere* **2022** 6956845.

Krishnamacharlu T 1972 Petrology of picrodolerites from the Dedan cluster, Gujarat; *J. Geol. Soc. India* **13** 262–272.

Large R R, Mukherjee I and Zhukova I 2018 Role of uppermost crustal composition in the evolution of the Precambrian ocean–atmosphere system; *Earth Planet. Sci. Lett.* **487** 44–53.

Law R D, Stahr Iii D W and Francsis M K 2013 Deformation temperatures and flow vorticities near the base of the Greater Himalayan Series, Sutlej Valley and Shimla Klippe, NW India; *J. Struct. Geol.* **54** 21–53.

Lee S 2010 Drawbacks of principal component analysis; *arXiv preprint arXiv* 1005.1770.

Lipp A G, Shorttle O, Syvret F and Roberts G G 2020 Major element composition of sediments in terms of weathering and provenance: Implications for crustal recycling; *Geochem. Geophys. Geosyst.* **21**, https://doi.org/10.1029/2019GC008758.

Madhavaraju J, Tom M, Lee I L and Lee Y 2016 Provenance and tectonic settings of sands from Puerto Peñasco, Desemboque and Bahia Kino beaches, Gulf of California, Sonora, México; *J. South Am. Earth Sci.* **71** 262–275, https://doi.org/10.1016/j.jsames.2016.08.005.

Makvandi S, Ghasemzadeh-Barvarz M and Beaudoin G 2016 Principal component analysis of magnetite composition

from volcanogenic massive sulfide deposits: Case studies from the Izok Lake (Nunavut, Canada) and Halfmile Lake (New Brunswick, Canada) deposits; *Ore Geol. Rev.* **72** 60–85, https://doi.org/10.1016/j.oregeorev.2015.06.023.

Mathur U B, Verma K K and Mehra S 1987 Tertiary-Quaternary stratigraphy of Porbandar area, southern Saurahstra, Gujarat; *Visesa Prakasana-Bharatiya Bhuvaijñanika Sarveksana,* pp. 333–345.

McLennan S M 1989 Rare earth elements in sedimentary rocks: Influence of provenance and sedimentary processes; *Geochem. Mineral. REE Rev. Mineral.* **21** 169–200.

McLennan S M, Hemming S, McDaniel D K and Hanson G N 1993 Geochemical approaches to sedimentation, provenance, and tectonics; *Spec. Pap. Soc. Am.* **21**.

McManus C E, McMillan N J, Dowe J and Bell J 2020 Diamonds certify themselves: Multivariate statistical provenance analysis; *Minerals* **10** 916.

Melluso L, Beccaluva L and Brotzu P 1995 Constraints on the mantle sources of the Deccan Traps from the petrology and geochemistry of the basalts of Gujarat State; *J. Petrol.* **36** 1393–1432.

Nadiri A A, Moghaddam A A, Tsai F T and Fijani E 2013 Hydrogeochemical analysis for Tasuj plain aquifer, Iran; *J. Earth Syst. Sci.* **122** 1091–1105.

Najafi S J, Cox K G and Sukheswala R N 1981 Geology and geochemistry of the basalt flows (Deccan Traps) of the Mahad–Mahableshwar section, India; Mem. Soc. India 300–315.

Naushad M, Dongre A and Behera J R 2019 Mineralogy of a new occurrence of Lamprophyre dyke from the Saurashtra Peninsula of Gujarat, Northwest Deccan Trap, India; *J. Geol. Soc. India* **93** 629–637, https://doi.org/10.1007/s12594-019-1241-6.

Nesbitt H 1979 Mobility and fractionation of REE during wearhering of a granodiorite; *Nature* **279** 206–210.

Nesbitt H W, Markovics G and Price R C 1980 Chemical processes affecting alkalis and alkaline earths during continental weathering; *Geochim. Cosmochim. Acta* **44** 1659–1666, https://doi.org/10.1016/0016-7037(80)90218-5.

Nesbitt H W and Young G M 1982 Early Proterozoic climates and plate motions inferred from major element chemistry of lutites; *Nature* **299** 715–717.

Nesbitt H W, Young G M, McLennan S M and Keays R R 1996 Effects of chemical weathering and sorting on the petrogenesis of siliciclastic sediments, with implications for provenance studies; *J. Geol.* **104** 525–542, https://doi.org/10.1086/629850.

Ohta T 2004 Geochemistry of Jurassic to earliest Cretaceous deposits in the Nagato Basin, SW Japan: Implication of factor analysis to sorting effects and provenance signatures; *Sedim. Geol.* **171** 159–180.

Pandey D K, Bahadur T and Mathur U B 2007 Stratigraphic distribution and depositional environment of the Chaya Formation along the northwestern coast of Saurashtra Peninsula, western India; *J. Geol. Soc. India* **69** 1215–1230.

Pant R K and Juyal N 1993 Late Quaternary coastal instability and sea level changes: New evidences from Saurashtra Coast, western India; *Zeitschrift fir Geomorphologie N.F.* **37(1)** 29–40.

Parrish R R and Hodges V 1996 Isotopic constraints on the age and provenance of the Lesser and Greater Himalayan sequences, Nepalese Himalaya; *Geol. Soc. Am. Bull.* **108** 904–911.

Paternie E D E, Hakkou R and Nga L E 2023 Geochemistry and geostatistics for the assessment of trace elements contamination in soil and stream sediments in abandoned artisanal small-scale gold mining (Bétaré-Oya, Cameroon); *Appl. Geochem.* **150** 105592.

Pe-Piper G, Triantafyllidis S and Piper D J W 2008 Geochemical identification of clastic sediment provenance from known sources of similar geology: The Cretaceous Scotian Basin, Canada; *J. Sediment. Res.* **78** 595–607, https://doi.org/10.2110/jsr.2008.067.

Raha P K and Sastry M V A 1982 Stromatolites and Precambrian stratigraphy in India; *Prec. Res.* **18** 293–318.

Ramirez-Herrera M T, Cundy A and Kostoglodov V 2007 Sedimentary record of late-Holocene relative sea-level change and tectonic deformation from the Guerrero Seismic Gap, Mexican Pacific Coast; *Holocene* **17** 1211–1220, https://doi.org/10.1177/0959683607085127.

Reid M K and Spencer K L 2009 Use of principal components analysis (PCA) on estuarine sediment dataset: The effect of data pre-treatment; *Environ. Pollut.* **157(89)** 2275–2281.

Richards A, Argles T and Harris N 2005 Himalayan architecture constrained by isotopic tracers from clastic sediments; *Earth Planet. Sci. Lett.* **236** 773–796.

Schwab F L 2003 Sedimentary petrology; *Encycl. Phys. Sci. Technol.,* pp. 495–529.

Shaji J, Banerji U S and Maya K 2022 Holocene monsoon and sea-level variability from coastal lowlands of Kerala, SW India; *Quat. Int.*, https://doi.org/10.1016/j.quaint.2022.03.005.

Sheth H C, Choudhary A K and Bhattacharyya S 2011 The Chogat–Chamardi subvolcanic complex, Saurashtra, northwestern Deccan Traps: Geology, petrochemistry, and petrogenetic evolution; *J. Asian Earth Sci.* **41** 307–324, https://doi.org/10.1016/j.jseaes.2011.02.012.

Sheth H C, Choudhary A K and Cucciniello C 2012 Geology, petrochemistry, and genesis of the bimodal lavas of Osham Hill, Saurashtra, northwestern Deccan Traps; *J. Asian Earth Sci.* **43** 176–192, https://doi.org/10.1016/j.jseaes.2011.09.008.

Singh P and Rajamani V 2001 Geochemistry of the floodplain sediments of the Kaveri river, southern India; *J. Sedim. Res.* **71** 50–60, https://doi.org/10.1306/042800710050.

Sinha A K 1978 Para-autochthonous turbiditic-flyschoidal Simla and terrigeneous-carbonate shale formations of Himalaya: Their litho-petrography and tectonic setting; *Him. Geol.* **8** 425–455.

Srikantia S V and Sharma R P 1976 Geology of the Shali Belt and the adjoining areas; *Geol. Surv. India Memoir* **106**.

Srikantia S and Sharma R P 1971 Simla group-A reclassification of the 'Chail Series', 'Jaunsar Series' and 'Simla Slates' in the Simla Himalaya; *Geol. Soc. India* **12** 234–240.

Sunkari E D and Abu M 2019 Hydrochemistry with special reference to fluoride contamination in groundwater of the Bongo district, Upper East Region, Ghana; *Sustain. Water Resour. Manag.* **5** 1803–1814, https://doi.org/10.1007/s40899-019-00335-0.

Taylor S R and McLennan S M 1985 *The continental crust: Its composition and evolution;* Blackwell Scientific Publications, Oxford, 362p, https://doi.org/10.1002/gj.3350210116.

Taylor S R and McLennan S M 1995 The geochemical evolution of the continental crust; *Rev. Geophys.* **33** 241–265.

Thakur V C 1992 Geology of western Himalaya; *Phys. Chem. Earth* **19** 1–355.

Tolosana-Delgado R, von Eynatten H, Krippner A and Meinhold G 2018 A multivariate discrimination scheme of detrital garnet chemistry for use in sedimentary provenance analysis; *Sedim. Geol.* **375** 14–26, https://doi.org/10.1016/j.sedgeo.2017.11.003.

Tribovillard N, Algeo T J, Lyons T and Riboulleau A 2006 Trace metals as paleoredox and paleoproductivity proxies: An update; *Chem. Geol.* **232** 12–32, https://doi.org/10.1016/j.chemgeo.2006.02.012.

Tripathy G R, Singh S K and Ramaswamy V 2014 Major and trace element geochemistry of Bay of Bengal sediments: Implications to provenances and their controlling factors; *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **397** 20–30, https://doi.org/10.1016/j.palaeo.2013.04.012.

Ueki K and Iwamori H 2017 Geochemical differentiation processes for arc magma of the Sengan volcanic cluster, Northeastern Japan, constrained from principal component analysis; *Lithos* **290–291** 60–75, https://doi.org/10.1016/j.lithos.2017.08.001.

Valdiya K S 1970 Simla slates: The Precambrian Flysch of the Lesser Himalaya, its turbidites, sedimentary structures and paleocurrents; *Geol. Soc. Am. Bull.* **81** 451–468.

Valdiya K S 1980 *Geology of Kumaun Lesser Himalaya*; Wadia Institute of Himalayan Geology, Dehradun, India.

Van Helvoort P-J, Filzmoser P and van Gaans P F M 2005 Sequential factor analysis as a new approach to multivariate analysis of heterogeneous geochemical dataset: An application to a bulk chemical characterization of fluvial deposits (Rhine–Meuse delta, The Netherlands); *Appl. Geochem.* **20** 2233–2251.

Vannay J C and Grasemann B 1998 Inverted metamorphism in the High Himalaya of Himachal Pradesh (NW India): Phase equilibria versus thermobarometry; *Schweizerische Mineralogische und Petrographische Mitteilungen* **78** 107–132.

Verma K K and Mathur U B 1979a *Geol. Surv. India*, pp. 1–54.

Verma K K and Mathur U B 1979b Report on the studies on miliolite limestones and associated tertiary and quaternary deposits of Delvada coast, Junagarh district, Gujarat; Geological Survey of India, Western Region.

Verma K K and Moitra A K 1975 Report on the drilling carried out till 18.2.1975 to study the extension of the Narmada–Broach Fault in the coastal Saurashtra in connection with the selection of site for setting up of Atomic Power Station at Balana in Saurashtra, Gujarat state; *Geol. Soc. India.*

Wishart J, George T S and Brown L K 2013 Measuring variation in potato roots in both field and glasshouse: The search for useful yield predictors and a simple screen for root traits; *Plant Soil* **368** 231–249, https://doi.org/10.1007/s11104-012-1483-1.

Corresponding editor: Kalachand Sain