Indian Journal of Ecology (2023) 50(6): 2016-2023
DOI: https://doi.org/10.55362/IJE/2023/4168

Manuscript Number: 4168
NAAS Rating: 5.79

Selection of Significant Raster Images for Digital Soil Mapping Using Data Reduction Technique

S. Vishnu Shankar, R. Kumaraperumal[1], M. Radha, S.G. Patil, M. Athira[2] and M. Nivas Raj[1]

Department of Physical Science & Information Technology
[1]Department of Remote Sensing and Geographic Information System
[2]Department of Soil Science and Agricultural Chemistry
Tamil Nadu Agricultural University, Coimbatore-641 003, India
E-mail: s.vishnushankar55@gmail.com

Abstract: The study was conducted to select the significant environmental covariates for mapping the soil through the computer-assisted soil mapping method. A total of 340 soil profile information was intersected against the environmental covariates to reduce the dimension of the data by which the quality of the spatial soil predictions can be improved. Robust PCA is known for its supremacy in handling image drive data as it can even process the data extracted from raster image series, which are high in outliers. The selection of significant covariates for digital soil mapping was done through robust Principal Component Analysis (PCA) and conventional PCA. The scree plot indicates that four principal components are to be considered from both methods. The selected principal components of robust PCA cumulatively contribute 63.17% of the total information to the original dataset, whereas conventional PCA contributes 52.46% only. Contribution charts were employed for extracting the significant environmental covariates in which 26 out of 33 covariates are obtained from the selected principal components of robust PCA. Soil mapping would be efficient if the covariates are selected through this process and employed for the mapping process.

Keywords: Data reduction technique, Digital soil mapping, Environmental covariates, Principal component analysis, Robust

Soil is a life support component that is vital to the Earth's ecosystem. All living organisms depend on soil directly or indirectly, where plants are predominant. Therefore, information regarding its nature, types, properties, spatial distribution, and extent is important for efficiently utilizing and managing (Afshar et al 2018) the resources. Over a decade, the distribution, pattern, and type of soil vary due to biotic and abiotic factors from place to place and time to time. Hence, it is necessary to identify and classify the soil forms in different areas. Mapping of soils helps to locate and classify the existing soil types (Semy and Singh, 2021). In addition, it collects information regarding their location, nature, and properties. Generally, traditional soil mapping is done based on the information gathered from soil samples at regular intervals throughout the landscape. The maps are primarily in hard copy and are gradually converted to digital form after the Geographical Information System (GIS) advent. But they are time-consuming, expensive, laborious, and untrustworthy in producing spatial variability of soil maps (Abdel-Kader et al 2011). The possibility of human error is high and unavoidable in conventional cases.

Digital soil mapping (DSM) is a predictive soil mapping that collects the geographically referenced soil databases created on quantitative relationships between spatially related environmental data (Kumaraperumal et al 2022). It is a quantitative technique of surveying, mapping, and analyzing the associations between soil properties and environmental variables (Polisgowdar et al 2022, Wani et al 2022). It involves the generation of initial soil survey maps, refining, or updating the soil survey information, making specific soil interpretations, and measuring the risk (Zhang et al 2017). DSM is employed using the soil profile information as dependent and environmental covariates as independent variables. These data are collected once in a particular period, and the soil maps are predicted for the future using probability statistics or artificial intelligence methods. The choice of the independent variable for effective mapping remains a challenging one. Out of available covariates, effective mapping can be generated only if the significant covariates are identified and used in the mapping process. Environmental covariates are available in shape files and were converted into a raster format for mapping purposes in which the outliers are common. The common cause of outliers is faulty pixels like noise, occlusion, or alignment errors. It makes the conventional principal component analysis fail. So, robustness is the better option for handling these outlier datasets which paved the path for using robust principal component analysis in selecting the significant variable. These powerful dimensionality reduction techniques handle the outlier efficiently and analyze image data in a better way.

The modern glitches of estimation, optimization, image recognition, and signal processing are part of analyzing huge dimensional data. In computer vision applications, outliers naturally occur within a sample image which demands the need for structuring low-dimension approximations for the large-scale dataset. This versatile technique can be used even when the assumptions are violated which readily deals with pixels, raster images, facial recognition, image compression, and remote sensing data, etc. (Pinto da Costa and Cabral 2022). Robust PCA is insensitive to outliers and produces the same result as the conventional method without outliers (Sapra 2010). However, they leave the residuals associated with the outlier and help in identifying the influential points. So, the study attempts to use robust principal component analysis to discover the significant covariates. This study gives insight in selecting the important environmental covariates that DSM can use for future purposes.

MATERIAL AND METHODS

Study area: The Coimbatore district is located in the western Agro-climatic Zone of Southern India which is 411 meters above the mean sea level and geographically extended from 11°24'23" to 10°13'12" N latitude and 76°39'20" to 77°18'00" E longitude covering an area of about 4721.28 sq. km. The average annual rainfall is about 642.2mm with the mean maximum and minimum temperatures of the district being 32.7°C and 21.5°C, respectively. Sugarcane, cotton, turmeric, and oilseeds were the different crops cultivated in the study area. The area map and the corresponding soil profile sites, along with their elevation and physiographical information are given in Figure 1.

Data: The soil class information containing 340 soil profile samples i.e., a subgroup was collected with the help of the Department of Remote Sensing and GIS and the Department of Soil Science and Agricultural Chemistry, TNAU, Coimbatore in 2022. Further, thirty-three different environmental covariates (Table 1) depicting the Climate, Organisms, Relief, and Parent were developed from the DEM (Digital Elevation Model) and other remote sensing variables. The continuous spatial information on the climatic parameters was downloaded from WorldClim 2.1 including the temporal range of 1970-2000 interval at 30 arc seconds. The qualitative nature of the Agro-Climatic, Agro-Ecological zones, and the Western Ghats were rasterized to enable stacking with other covariates. Satellite data products (Landsat 8) along with their transformed NDVI images were utilized to impart the influence of the organisms on soil formation (NRSC 2012). The cloud-free Landsat 8 data product was downloaded from USGS earth explorer during March to May period to enhance the delineation of the soil attributes.

The relief parameters were primarily derived from Shuttle Radar Topography Mission (SRTM) DEM product by utilizing the hydro geomorphometric indexes of the SAGA GIS software. Existing soil information on the Geology, Geomorphology, and Land use and Land cover products obtained from the National Remote Sensing Centre (NRSC) at 1:50,000 map interval was rasterized (NRSC, 2016) using the raster tool feature available in ArcGIS 10.6 software and were implemented as covariates. The rasterized covariates

Table 1. Parameters of environmental covariates

Covariate	Parameter	Scale
Climate	Maximum annual temperature	ºC / 30 seconds
	Minimum annual temperature	ºC / 30 seconds
	Mean annual rainfall	mm/ 30 seconds
	Agro-Climatic Zone	30m
	Agro-Ecological Zone	30m
Organisms	Land Use & land cover map	1:50,000 scale
	Landsat 8 – Band 2 (Blue)	30m
	Landsat 8 – Band 3 (Green)	30m
	Landsat 8 – Band 4 (Red)	30m
	Landsat 8 – Band 5 (NIR)	30m
Relief	Elevation (SRTM DEM)	30m
	Hill shading	30m
	Aspect	30m
	Convergence index	30m
	General curvature	30m
	Longitudinal curvature	30m
	Slope length steepness (LS) factor	30m
	Maximum curvature	30m
	Mid slope position	30m
	Minimum curvature	30m
	Plan curvature	30m
	Profile curvature	30m
	Slope gradient	30m
	Tangential curvature	30m
	Terrain ruggedness index	30m
	Topographic wetness index	30m
	Total catchment area	30m
	Total curvature	30m
	Valley depth	30m
	Western ghats	30m
	Physiography	1:50,000 scale
Parent material	Geomorphology	1:50,000 scale
	Rock outcrop difference ratio	1:50,000 scale

are then resampled and reprojected to the common projection system. The covariate parameters are then stacked and intersected with the soil profile information in the R spatial environment. The extracted data is subjected to data reduction techniques from which the significant environmental parameters are sorted.

Variable screening: Dimensionality reduction techniques are the multivariate approaches used when the variables are highly correlated or when the number of variables is more than the number of observations. It is also used to identify the variable which significantly contributes to the results. Data processing with a low dimensional significant variable makes the analysis easier and more effective (Gowsar et al 2019).

Principal component analysis: Principal Component Analysis (PCA) is one of the dimensionality reduction techniques used to reduce the dimension of the datasets by removing the nsignificant variables without any information loss and for forming predictive models (Dinesh et al 2022). It helps in screening the significant environmental covariates, thus reducing the dimension of original variables and extracting a small number of latent factors, *i.e.,* Principal Components (PC). Principal components are a linear combination of *p* random variables explaining the variation produced by original data. The linear combination the of *p* variable is given by,

$$PCv = ev_1Y_1 + ev_2Y_{2+}\ldots + ev_pY_p \quad (1)$$

PCA computes the covariance matrix or correlation matrix for the given datasets, followed by calculating the eigenvectors and the eigenvalues for the covariance matrix.

$$\det(\lambda I - A) = 0 \text{ and } (\lambda I - A)\, v = 0I \quad (2)$$

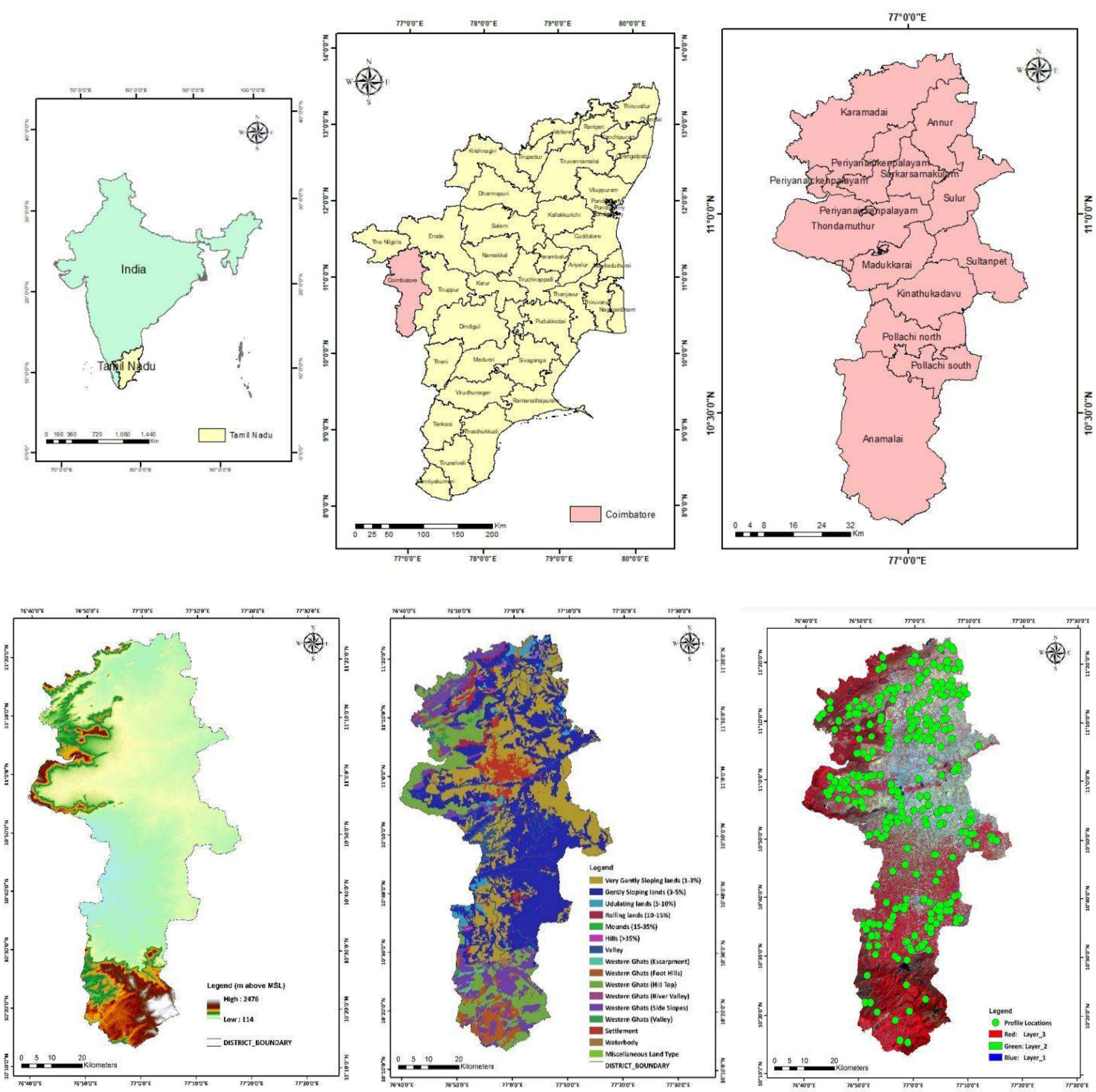

Fig. 1. Study area map

Where λ is an eigenvalue, *v* is an eigenvector, *A* is a square matrix, and *det* is the determinant of the matrix. However, the conventional PCA technique is not recommended for the study data, as they are susceptible to outliers and poorly handle the data extracted from raster images (Jolliffe and Cadima 2016).

Robust principal component analysis: Outliers are the observations that fluctuate significantly from actual observations. Outliers in data affect the conventional PCA analysis in terms of the correlation matrix, covariance estimations, and proportions of the total variance (Zahariah and Midi 2022). To overcome these problems, robust principal component analysis is replaced in place of the conventional method. Robust PCA helps reduce the dimensionality problem of explanatory variables and deals with multi-collinearity like the traditional way, but all these in the presence of outliers in data (Lu et al 2023). It recovers the low-rank matrix and principal components with high probability, avoiding the choice of the tuning parameter. Rousseeuw used the Minimum covariance determinant (MCD) method in robust principal component analysis (Kalina and Tichavský 2022). It reduces the number of iterations and analysis time for limited to low-dimensional data. It uses the intra-sample outlier process to account for pixel outliers. The Minimum Covariance Determinant (MCD) is also referred to as modified robust PCA. It is one of the popular and fastest methods in showing a high degree of robustness against outliers. In robust PCA, the location and scale parameters of conventional PCA are replaced and produced with high breakdown using MCD (Piccini et al 2019). Parameters μ and Σ are replaced with robust estimates μ (median) and Σ (MAD - Median Absolute Deviation) for calculating scale and location parameters. Initially, the data is split into groups, parameters are found using the MCD method, and finally, they are joined, which is the main modification in this procedure.

Let *h* represent the dimension of sub-clusters datasets containing n observations. The *h*-value determines the robustness of the estimator, and a minimum [((n + p + 1)) / 2] value should be taken as a lower bound. The MCD estimator attempts to find the minimum covariance determinant of the optimal h-subset containing these sub-clusters, and the distance between them is calculated by

$$d_i = \sqrt{(x-\hat{\mu}_0)^t \hat{\Sigma}_0^{-1}(x-\hat{\mu}_0)} \qquad (3)$$

The mean of the optimal h-subset gives the estimate of location parameter μMCD as

$$\hat{\mu}_{MCD} = \frac{\Sigma_{i=1}^{n} W(d_i^2)x_i}{\Sigma_{i=1}^{n} W(d_i^2)} \qquad (4)$$

The covariance matrix provides the estimate of scale parameter ΣMCD as

$$\hat{\Sigma}_{MCD} = C_1 \frac{1}{n} \Sigma_{i=1}^{n} W(d_i^2)(x_i \hat{\mu}_{MCD})(x_i - \hat{\mu}_{MCD})^t \qquad (5)$$

The breakdown point value of the MCD estimator is (n – h + 1)/n. Weighted estimators (Bulut and Zaman 2022) will be used to increase the estimator's efficiency.

RESULTS AND DISCUSSION

The respective profile information of each covariate, concerning the sample points is extracted using ArcGIS software. The data values obtained from stacked environmental covariates are used for the study. Even though all the parameters are expected to contribute some information to the map generation process, sorting out the important parameters with appropriate data reduction techniques is necessary (Samuel-Rosa et al 2015). Soil maps of the study area generated from the significant environmental covariates are highly efficient and informative compared to the map generated using all the covariates (Bhat et al 2020). Therefore, the environmental covariate layers, which significantly contribute to the map generation (Shafeeva et al 2022) process by providing necessary information, were sorted out using robust PCA techniques. Conventional PCA is worked for the data to show the effectiveness of robust PCA. The initial step of PCA starts with dimensioning the principal components using extracted data points. Each principal components contribute some percent of variation to the original data. The first few principal components account for a high percentage of variability, and the percentage of variation decreases as we move down. The next step is to select the minimum number of principal components which contribute a high percentage of the variation, i.e., cumulatively contributing to the original data set. The number of principal components is decided based on the knee point observed in the scree plot. Finally, the analysis is performed using R software.

Scree plot directs to pick four principal components as the knee point break at the fourth principal component (Greenacre et al 2022) (Table 2a, Fig. 2a). Those four principal components cumulatively contribute about 52.46% to the original dataset. Similarly, robust PCA is applied to the data. Figure 2b shows the scree plot formed using the eigenvalues of robust principal components. The selected four PCs cumulatively contribute about 63.17% variation to the total variation of original data. The percentage of variation contributed by PC1 is 30.84 which is the highest among PCs. It is followed by PC2, PC3, and PC4. These four principal components have eigenvalues greater than 1. Thus, the selected principal components are uncorrelated and free from multi-collinearity (Shankar et al 2019). Since robust PCA contributes more information with fewer principal components than the conventional method, the selected four principal components of the robust technique are taken for further analysis (Parra-González and Rodriguez-Valenzuela 2017). Generally, the selected four principal components are used for further computing purposes. But as this study aimed to determine the environmental parameters for mapping, the high contributing parameter of each principal component is found.

Table 2a. Results of conventional PCA

Principal components	Eigen value	Percentage of variance	Cumulative percentage of variance	Principal components	Eigen value	Percentage of variance	Cumulative percentage of variance
PC1	9.05	22.78	22.78	PC 18	0.24	1.32	96.04
PC2	6.18	15.01	37.79	PC 19	0.21	1.28	97.32
PC3	4.07	8.91	46.70	PC 20	0.20	0.77	98.09
PC4	2.13	5.76	52.46	PC 21	0.10	0.67	98.76
PC5	2.11	4.90	57.36	PC 22	0.06	0.54	99.30
PC6	1.34	4.74	62.10	PC 23	0.03	0.44	99.74
PC7	1.33	4.67	66.77	PC 24	0.02	0.23	99.97
PC8	1.01	4.56	71.33	PC 25	0.02	0.03	100.00
PC9	0.70	3.76	75.09	PC 26	0.00	0.00	100.00
PC10	0.73	3.56	78.65	PC 27	0.00	0.00	100.00
PC11	0.70	3.45	82.10	PC 28	0.00	0.00	100.00
PC12	0.61	3.33	85.43	PC 29	0.00	0.00	100.00
PC13	0.58	2.78	88.21	PC 30	0.00	0.00	100.00
PC14	0.51	2.01	90.22	PC 31	0.00	0.00	100.00
PC15	0.44	1.58	91.80	PC 32	0.00	0.00	100.00
PC 16	0.32	1.51	93.31	PC 33	0.00	0.00	100.00
PC 17	0.31	1.41	94.72				

Table 2b. Results of Robust PCA

Principal components	Eigen value	Percentage of variance	Cumulative percentage of variance	Principal components	Eigen value	Percentage of variance	Cumulative percentage of variance
PC1	6.90	30.84	30.84	PC 18	0.51	0.24	95.71
PC2	5.01	15.37	46.21	PC 19	0.41	0.98	96.69
PC3	3.01	10.47	56.68	PC 20	0.38	0.93	97.62
PC4	2.41	6.49	63.17	PC 21	0.30	0.74	98.36
PC5	1.53	5.47	68.64	PC 22	0.26	0.63	98.99
PC6	1.37	4.04	72.68	PC 23	0.25	0.61	99.60
PC7	1.30	3.82	76.50	PC 24	0.12	0.29	99.89
PC8	1.27	3.35	79.85	PC 25	0.07	0.10	99.99
PC9	0.99	2.95	82.80	PC 26	0.04	0.01	100.00
PC10	0.97	2.83	85.63	PC 27	0.03	0.00	100.00
PC11	0.96	2.69	88.32	PC 28	0.02	0.00	100.00
PC12	0.95	1.92	90.24	PC 29	0.01	0.00	100.00
PC13	0.91	1.85	92.09	PC 30	0.00	0.00	100.00
PC14	0.88	1.20	93.29	PC 31	0.00	0.00	100.00
PC15	0.77	0.89	94.18	PC 32	0.00	0.00	100.00
PC 16	0.73	0.75	94.93	PC 33	0.00	0.00	100.00
PC 17	0.64	0.54	95.47				

The significant parameter of selected principal components is screened using contributing chart developed from principal components. Contribution charts were produced between PC1-PC2 and PC3-PC4 and can be formed for two principal components simultaneously, containing bars and cut-off lines (Fig. 3a, 3b). The bars in the charts represent the contribution of each parameter. The cut-off line in the graph indicates that the bars above the line contribute highly, whereas the bars below the line contribute

Table 3. Major contributing covariates of four principal components

Contributing covariates of principal components PC1, PC2, PC3 and PC4 (26 variables)	
Satellite data- Blue	Longitudinal curvature
Satellite data- Green	LS factor
Satellite data- Red	Maximum curvature
Satellite data- NIR	Minimum curvature
Agro Climatic Zone	Physiography
Agro-Ecological Zone	Plan curvature
Western Ghats	Slope
Maximum Temperature	Tangential curvature
Minimum Temperature	Terrain Ruggedness Index
Rainfall	Topographic Wetness Index
Analytical Hill shading	Total curvature
Elevation	Valley depth
General curvature	Lithology

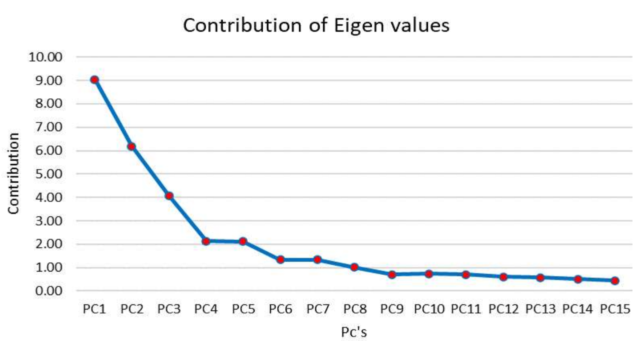

Fig. 2a. Scree plot of conventional PCA

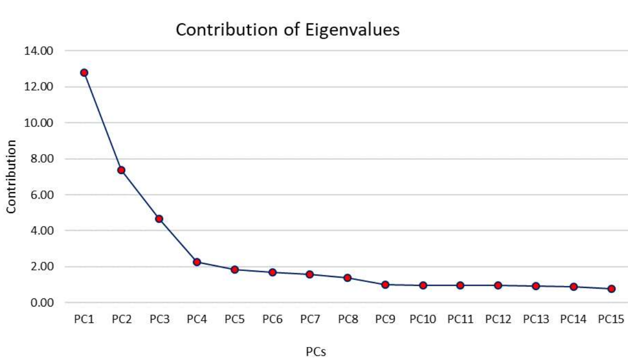

Fig. 2b. Scree plot of robust PCA

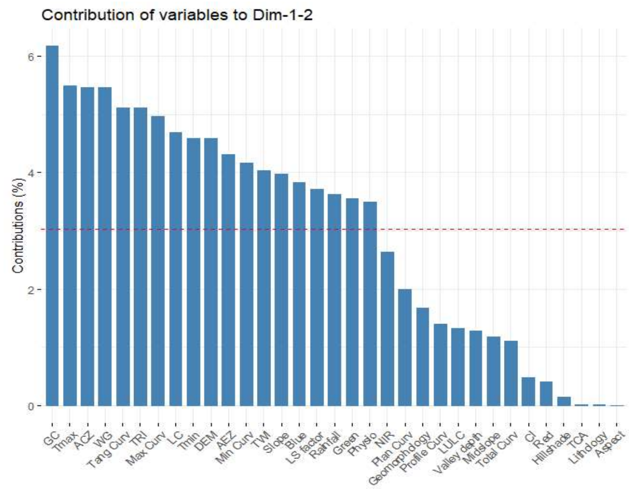

Fig. 3a. Contribution chart of PC1 and PC2

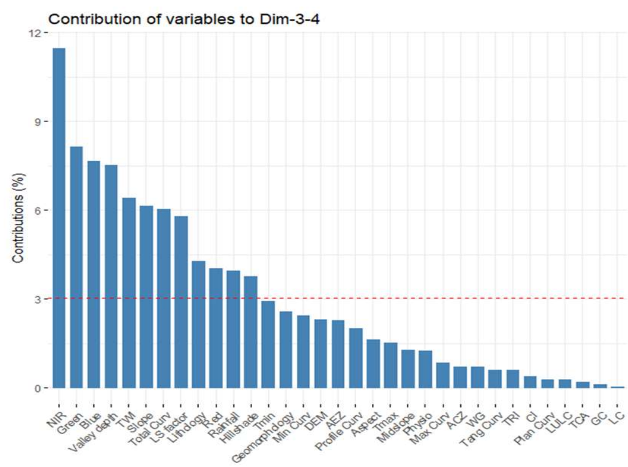

Fig. 3b. Contribution chart of PC3 and PC4

less (Luo et al 2022). If there are N parameters, then the expected average contribution of a PC is 1/N= n%. The cut-off line will fall on the n%, and the contributing layers of only two PCs can be determined. Selecting the contributing variables from two principal components, namely PC1 and PC2, given by

(n× eigenvalue of PC1) + (n× eigenvalue of PC2) (6)

The contribution chart of PC1-PC2 (Fig. 3a) shows that nineteen out of thirty-three layers have a high percentage and fall above the cut-off line. Similarly, the PC3-PC4 contribution chart (Fig. 3b) reveals those twelve out of thirty-three layers with a high contribution percentage. The two-contribution chart are combined and found that twenty-six layers out of thirty-three were highly contributing covariates of selected PCs. Table 3 gives the selected environmental covariates as the study's final independent variables and can be used for future mapping (Cavazzi et al 2013).

CONCLUSION

Robust PCA effectively reduces dimensionality when the datasets are derived from image or raster forms. They are extensions of the conventional method in terms of robustness and are effective in the presence of outliers. The robust PCA selects four principal components, cumulatively contributing to the variation of about 63.17%, whereas conventional PCA contributes only 52.46%. The scree plot aided the selection of appropriate principal components in both methods, and it was four principal components. Each principal component is composed of all environmental covariates, and the significant covariates in each principal component were found using contribution charts. Contribution charts were developed for PC1-PC2 and PC3-PC4. The cut-off line in the charts indicated the significantly contributing covariates. Environmental covariates like Agro Climatic Zone, Agro-Ecological Zone, Western Ghats, Maximum Temperature, Minimum Temperature, Rainfall, Analytical Hill shading, Elevation, General curvature, Longitudinal curvature, LS factor, Maximum curvature, minimum curvature, physiography, plan curvature, slope, tangential curvature, terrain ruggedness index, topographic wetness index, total curvature, valley depth, lithology, and green, red, blue, NIR wavelength are found to important participants of selected four principal components. Therefore, in the process of digital soil mapping, effective map generation can be achieved when the environmental covariates used as independent variables are chosen in the suggested way.

REFERENCES

Abdel-Kader and Fawzy Hassan 2011. Digital soil mapping at pilot sites in the northwest coast of Egypt: A multinomial logistic regression approach. *The Egyptian Journal of Remote Sensing and Space Science* **14(**1): 29-40.

Afshar FA, Ayoubi S and Jafari A 2018. The extrapolation of soil great groups using multinomial logistic regression at a regional scale in arid regions of Iran. *Geoderma* **315**: 36-48.

Bhat MA, Sahoo J and Sharma MK 2020. Vertical distribution of nutrients vis-a-vis soil properties in different geomorphic units of north-eastern Haryana, India. *Indian Journal of Ecology* **47**(1): 58-67.

Bulut H and Zaman T 2022. An improved class of robust ratio estimators by using the minimum covariance determinant estimation. *Communications in Statistics-Simulation and Computation* **51**(5): 2457-2463.

Cavazzi S, Corstanje R, Mayr T, Hannam J and Fealy R 2013. Are fine-resolution digital elevation models always the best choice in digital soil mapping? *Geoderma* **195**: 111-121.

Dinesh GK, Sharma DK, Jat SL, Sri KS, Bandyopadhyay KK, Bhatia A and Kumar RR 2022. Ecological relationship of earthworms with soil physicochemical properties and crop yields in conservation agriculture. *Indian Journal of Ecology* **49**(6): 2135-2139.

Gowsar SRN, Devi MN, Radha M, Kumaraperumal R and Shankar SV 2019. Characterization of groundwater quality using principal component analysis of Andipalayam and Neruperichel blocks of Tiruppur District. *International Archive of Applied Sciences and Technology* **10**(2): 76-83.

Greenacre M, Groenen PJ, Hastie T, d'Enza A, Markos A and Tuzhilina E 2022. Principal component analysis. *Nature Reviews Methods Primers* **2**(1): 100.

Jolliffe IT and Cadima J 2016. Principal component analysis: A review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences* **374**(2065): 20150202.

Kalina J and Tichavský J 2022. The minimum weighted covariance determinant estimator for high-dimensional data. *Advances in Data Analysis and Classification* **1**: 1-23.

Kumaraperumal R, Pazhanivelan S, Geethalakshmi V, Raj MN, Muthumanickam D, Kaliaperumal R, Shankar SV, Nair AM, Yadav AM and Kshatriya TVT 2022. Comparison of machine learning-based prediction of qualitative and quantitative digital soil-mapping approaches for Eastern Districts of Tamil Nadu, India. *Land* **11**(12): 2279.

Lu B, Hu Y, Yang D, Liu Y, Liao L, Yin Z, Xia T, Dong Z, Harris P, Brunsdon C and Comber L 2023. GW model: A software for geographically weighted models. *SoftwareX* **21:** 101291.

Luo C, Wang Y, Zhang X, Zhang W and Liu H 2022. Spatial prediction of soil organic matter content using multiyear synthetic images and partitioning algorithms. *Catena* **211**: 106023.

NRSC 2012. Lithology, Physiography and Soils of Tamil Nadu at 1:50,000 Scale, Natural Resources Census Project; *National Remote Sensing Centre, ISRO in Collaboration with Institute of Remote Sensing and Tamil Nadu Agricultural University: Hyderabad, India*.

NRSC 2016. Land Use/Land Cover Database on 1:50,000 Scale, Natural Resources Census Project, LUCMD, LRUMG, RSAA; *National Remote Sensing Centre, ISRO: Hyderabad, India*.

Parra-González SD and Rodriguez-Valenzuela J 2017. Determination of the soil quality index by principal component analysis in cocoa agroforestry system in the Orinoco Region, Colombia. *Journal of Agriculture and Ecology Research International* **65**: 1-8.

Piccini C, Marchetti A, Rivieccio R and Napoli R 2019. Multinomial logistic regression with soil diagnostic features and land surface parameters for soil mapping of Latium (Central Italy). *Geoderma* **352**: 385-394.

Pinto da Costa JF and Cabral M 2022. Statistical methods with applications in data mining: A review of the most recent works. *Mathematics* **10**(6): 993.

Polisgowdar BS, Roa AR and Bachina HB 2019. Land use land cover change detection using remote sensing and geographic information system in canal command area, Karnataka. *Indian Journal of Ecology* **46**(4): 683-690.

Samuel-Rosa A, Heuvelink GBM, Vasques GM and Anjos LCH 2015. Do more detailed environmental covariates deliver more accurate soil maps? *Geoderma* **243**: 214-227.

Sapra SK 2010. Robust vs. classical principal component analysis in the presence of outliers. *Applied Economics Letters* **17**(6): 519-523.

Semy K and Singh MR 2021. Comparative assessment on physico-chemical properties of coal mining affected and non-affected forest soil at Changki, Nagaland. *Indian Journal of Ecology* **48**(1): 36-42.

Shafeeva E, Miftakhov I, Ishbulatov M and Lykasov O 2022. Creating a soil map using digital soil mapping on the example of the Diurtiulinsky Municipal District. *Asian Journal of Water, Environment and Pollution* **19**(3): 89-95.

Shankar SV, Radha M and Kumaraperumal R and Gowsar SRN 2019. Statistical evaluation of physico-chemical properties of Soils of Coimbatore district using Dimensionality Reduction Technique. *International Archive of Applied Sciences and Technology* **10**(2): 84-89.

Wani OA, Sharma KR, Sharma V, Kumar S, Mahdi SS, Hussain N and Singh G 2022. Digital mapping of soil physicochemical properties of Ramban district of Jammu and Kashmir using geographic information system. *Indian Journal of Ecology* **499**(5): 1654-1660.

Zahariah S and Midi H 2022. Minimum regularized covariance determinant and principal component analysis-based method for the identification of high leverage points in high dimensional sparse data. *Journal of Applied Statistics* **12**: 1-19.

Zhang G, Feng L and Song X 2017. Recent progress and future prospect of digital soil mapping: A review. *Journal of Integrative Agriculture* **16**(12) :2871-2885.

Received 23 May, 2023; Accepted 25 October, 2023