Comparative Petrographic Analysis of Volcanic Rocks in Five Different Areas Using “R”

J Ramdani¹, F Septiana¹, D E Irawan²*, A Darul¹

¹Institut Teknologi dan Sains Bandung, Kompleks Industri Deltamas, Bekasi
²Institut Teknologi Bandung, Bandung, 40132
³Author to whom any correspondence should be addressed, dasaptaerwin@gmail.com
dasaptaerwin@outlook.co.id

Abstract. Petrographical analysis is a visual method to describe the mineral properties of a rock sample, related to its composition and classification. The mineral composition is mostly determined using a comparator, thus it is merely a quantitative method. This paper uses multivariable statistics as tools to support the visual classification. We took 89 thin sections from five locations, based on secondary data, gathered from five final projects from Department of Geology ITB: Mt. Lamongan-Probolinggo (LAM), Mt. Banyuresmi-Bogor (BAN), Mt. Kromong-Palimanan (KRM), Mt. Sangkur-Garut (SAN), and Mt. Wangi-Purworejo (WAN). The locations is Quaternary volcanic (LAM, KRM, SAN, and WAN) and Tertiary (BAN), consists of 85 igneous rock samples, three pyroclastics, and one sedimentary rock sample. The mineral percentage data was then analyzed using Principal Component Analysis (PCA) and Cluster Analysis (CA) with R. Our data frame size was 89 rows x 24 columns. Our preliminary results show a separation between Quaternary and Tertiary sample s on PCA and CA plots. We identified three clusters: cluster 1 BAN; cluster 2 LAM, KRM, SAN, WAN, and BAN, and cluster 3 LAM, KRM, SAN, WAN, and BAN. Based on our PCA plot, cluster 1 shows a strong influence of: chlorite, calcite, and quartz; cluster 2: pirokxene, plagioklas, olivine, glass and cluster 3: olivine, pirokxene, opaque, glass, and feldspar. Following the results we see that there are samples from BAN detached from the rest. We believe this is due to an anomalous volcanic process on that spot. On the other hand we see the similarity between samples from different locations. Based on those analyses, we perceive that multivariable statistics can support visual petrographic analysis by reading the data structure, selecting stronger variables and extracting clusters.

Key words: petrography, multivariable statistics, volcanic rock classifications

1. Background
Petrographical analysis is one of the most used method in geological mapping for rock classification. This method uses visual strength to identify the microscopic structure and texture of minerals. The interesting bit that we would like to introduce is whether a quantitative analysis, based on multivariable statistics principles, can support the visual analysis by reading mineral percentage data set, reducing the number of variables, creating a new set of variables, and extracting clusters.
We used Principal Component Analysis (PCA) and Cluster Analysis (CA) technique in this case. The methods have been widely used in the field of hydrogeology [1, 2], hot water classification [3], and ecology [4], to capture hidden patterns in the data structure [5].

2. Materials and methods
The analysis in this paper began with the preparation of thin section of rock sample[6] and petrographic analysis under a microscope[7]. Then the mineral identified in previous activity was tabulated before went to the statistical analysis. Our toy data set was secondary data set extracted from five final project reports published by The Department of Geology ITB library repository. The thin section samples were taken from five locations: Mount Lamongan-Probolinggo (LAM)[8], Banyuresmi-Bogor (BAN)[9], Gunung Kromong-Palimanan (KRM)[10], Mount Sangkur-Garut (SAN)[11], and Mount Wangi-Purworejo (WAN)[12] (see Figure 1). Based on the time scale, the five locations are Quartenary volcanic system (LAM, KRM, SAN, and WAN) and Tertiary volcanic system (BAN). The total samples were 89 consists of 85 (igneous rocks) from basalt to andesite, three samples (pyroclastics) tuffaceous sands and tuff crystal, and one sedimentary deposit. We stored curated data as separate open access data and hosted online[13].

![Java Island, Research location](image)

Figure 1. Study are five locations with active volcanism

Basically, we would like to know the classification of the volcanic rocks in those five locations based on mineral composition derived from previous petrographical analysis under statistical equations. Our particular interest is how we can draw connections between samples from different locations and how we markup anomalous samples using PCA and CA[14]. The PCA will reduce the number of variables and creates a set of new variables (or grouping of old variables), while CA will make a classification of samples based on the data structure (Figure 2).
We used R, an open source statistical software [14] and R Studio IDE[15]. The dataset was parsed as csv (comma separated value) and then analyzed using the following packages: ggplot2[16], dplyr[17], factomineR[18], factoExtra[19], cluster 20, ggcorrplot[21], and ape[22].

3. Results and discussions

Our preliminary results show a separation between Quartenary and Tertiary samples on PCA and CA plots and also volcanic rocks from sedimentary rocks. Figure 3 shows clustering of variables in four quadrants:

- Quadrant 1 (upper left), 2 (upper right), and 4 (lower left) show strong variables from volcanic rocks,
- While Quadrant 3 (lower right) shows strong variables for sedimentary rocks.

We identified three groups with various strong variables (see Figure 5 and Figure 6):

- Group 1: LAM, KRM, SAN, WAN, and BAN with a strong influence of: piroxene, plagioklas, olivine, glass, clay, iron oxide;
- Group 2: LAM, KRM, SAN, WAN, and BAN with strong influence of: piroxene, plagioklas, olivine, glass;
- Group 3: WAN and BAN with strong control of clay, quartz, and foraminifera.

The following results we see that there are samples from BAN detached from the rest. We believe this is due to an anomalous volcanic process on that spot. On the other hand we see the similarity between samples from different locations. Based on those analyses, we perceive that multivariable statistics can support visual petrographic analysis by reading the data structure, selecting stronger variables and extracting clusters.

Figure 4 shows three PCA groups:

- Group 1: LAM, KRM, SAN, WAN, and BAN with a strong influence of: piroxene, plagioklas, olivine, glass, clay, iron oxide;
- Group 2: LAM, KRM, SAN, WAN, and BAN with strong influence of: piroxene, plagioklas, olivine, glass;
- Group 3: WAN and BAN with strong control of clay, quartz, and foraminifera.

As we see in Figure 4, there are many points assembled at the center of the plot indicating a proportional influence of the variables, and some are plotted further to the side to indicate stronger influence from certain variables than the other. This plot clearly could show us the role of variables building the system spatially.
Another interesting point is the occurrence of foraminifera (as organic substances) in the cluster where most volcanic rock samples are located. This should be an indication of interaction between volcanic rock with sedimentary rock, possibly under the volcanic rock, possibly in erosion process during volcanic eruption and other unknown reasons. First impression from us, the clustering is able to capture a minimum value of foraminifera content and point its spatial location in the plot.

Consistently, Figure 5 shows three clusters:
- Cluster 1: LAM, KRM, SAN, WAN, and BAN dominated by volcanic-based minerals: olivine, pyroxene (piroksen), opaque, plagioclase, and glass,
- Cluster 2: mostly WAN samples with clay mineral, foraminifera, and a small fraction of opaque and quartz.
- Cluster 3: mostly BAN samples with plagioclase, quartz, dan calcite, as cluster marking.
Figure 4 Clusplot PCA to show the spatial distribution of samples based on the influence of variables in each quadrant.

Figure 5 Dendogram from CA

4. Remarks

According to our experiment, the cluster results are consistent with our visual observations. This could be an initial signal that this method can be used for massive petrographical sample analysis. We suggest more toy data set with various volcanic forming history should be added to test the clustering system in this paper. Another question we have is whether this method could also be applied for other rock type, limestones or sand and clay. This method benefits to a lot of petrologists to make their rock
classification easier and time effective. More detail description is needed to add more values to our model.

The method could detect strong and weak roles of the variables in the model. This could help geoscientists to evaluate and specifically classify large amount of data effectively and efficiently. But we need to watch for computer memory usage in relation to the number of data. Memory size of 8 GB is recommended for such purpose. We also believe the usage of open source software are the future in geology. R and Python are the two main programming languages in this context as they provide a straight forward programming algorithm with thousands of free and usable add on package to make coding as a pleasant activity.

In this paper, we also need to point out the importance of open access practice in research. To support this movement and setting an example, aside to only put the manuscript online, we also provide related data and R code freely available under CC-BY (Creative Commons Attribution) 4.0 license in INARxiv preprint server, a community-driven server for Indonesian academia.

REFERENCES

[1] Irawan DE, Puradimaja DJ, Notosiswoyo S, and Sumintadireja P 2009 Hydrogeochmistry of Volcanic Hydrogeology Based on Cluster Analysis of Mount Ciremai, West Java, Indonesia. Journal of Hydrology 376: 221–234, url: https://doi.org/10.1016/j.jhydrol.2009.07.033.

[2] King AC, Raiber M, and Cox ME 2014 Multivariate statistical analysis of hydrochemical data to assess alluvial aquifer–stream connectivity during drought and flood: Cressbrook Creek, Southeast Queensland, Australia, Hydrogeology Journal, 22, 481–500, doi:10.1007/s10040-013-1057-1, url: http://link.springer.com/10.1007/s10040-013-1057-1.

[3] Sumintadireja P, Irawan DE, Rezky Y, Ggio PU, and Agustin A 2016 Classifying hot water chemistry: Application of multivariate statistics, ITB International Geothermal Workshop. Zenodo. url: http://doi.org/10.5281/zenodo.45630.

[4] Wiharto M 2013 Analisis Klaster Menggunakan Bahasa Pemrograman R Untuk Kajian Ekologi, url: http://ojs.unm.ac.id/index.php/bionature/article/view/1450, accessed October 20, 2016.

[5] Hatvani IG, Magyar N, Zessner M, Kovács J, and Blaschke AP 2014 The Water Framework Directive: Can more information be extracted from groundwater data? A case study of Seewinkel, Burgenland, Eastern Austria, Hydrogeology Journal, 22, 779–794, doi:10.1007/s10040-013-1093-x, url: http://dx.doi.org/10.1007/s10040-013-1093-x.

[6] Subagyo dan Tukidjo 2003 Perparasi sayatan tipis dan poles contoh batuan dari Jumbang I, Kalan - Kalimantan Barat dan Ketapang, Madura - Jawa Timur, Kumpulan laporan hasil penelitian BATAN tahun 2003, ISBN 978-979-99141-2-5, url: http://digilib.batan.go.id/e-prosiding/File%20Prosiding/Geologi/Laporan_Pen._2004-2006_PPGN_berkas_A/artikel/subagyo_es_472.pdf.

[7] Open University 2017 An introduction to minerals and rocks under the microscope, Chapter: Minerals and the microscope, url: http://www.open.edu/openlearn/science-maths-technology/science/introduction-minerals-and-rocks-under-the-microscope/content-section-2.3.2.

[8] Jhonny 2006 Petrografi Gunung Api Lamongan Kabupaten Probolinggo-Lumajang Jawa Timur, Skripsi S1 Prodi Teknik Geologi, Institut Teknologi Bandung.

[9] Tanghamap DY 2009 Geologi Dan Petrografi Batuan Vulkanik Daerah Banyuresmi Dan Sekitarnya Kecamatan Cigudeg Kabupaten Bogor Jawa Barat, Skripsi S1 Prodi Teknik Geologi, Institut Teknologi Bandung.
[10] Hadinata J 2009 Geologi dan Alterasi Permukaan Gunung Kromong Dan Sekitarnya Provinsi Jawa Barat, Skripsi S1 Prodi Teknik Geologi, Institut Teknologi Bandung.

[11] Wisnu A 1989 Geologi Daerah Banjar Dan Sekitarnya Dan Petroografi Satuan Vulkanik Gunung Sangkur Kabupaten Ciamis, Skripsi S1 Prodi Teknik Geologi, Institut Teknologi Bandung.

[12] Wahyuningsih R 1992 Geologi Dan Petrogenesa Batuan Vulkanik Daerah Gunung Wangi dan Sekitarnya Purworejo Jawa Tengah, Skripsi S1 Prodi Teknik Geologi, Institut Teknologi Bandung.

[13] Ramdani J, Septiana F, Irawan DE, and Darul A 2017 Comparative Petrographic Analysis of Volcanic Rocks in Five Different Areas Using R. Skripsi S1 Teknik Pertambangan, ITSB, Open Science Framework. url: https://osf.io/sj5m4.

[14] Wilkinson DJ 2014 Multivariate Data Analysis using R: a course notes, Tech. rep., url: https://www.staff.ncl.ac.uk/d.j.wilkinson/teaching/mas8381/notes14.pdf.

[15] Gio PU and Irawan DE 2015 Belajar Statistika dengan R, USU Press, ISBN: 979 458 806 7, url: http://www.olahdatamedan.com/.

[16] R Core Team 2016 R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, url: https://www.R-project.org/.

[17] R Studio Team 2015 RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, url: http://www.rstudio.com/.

[18] Wickham H 2009 ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, url: http://ggplot2.org.

[19] Wickham H and Francois R 2016 dplyr: A Grammar of Data Manipulation, url: https://CRAN.R-project.org/package=dplyr, R package version 0.5.0.

[20] Lê S, Josse J, and Husson F 2008 FactoMineR: A Package for Multivariate Analysis, Journal of Statistical Soft-ware, 25, 1–18, url: http://dx.doi.org/10.18637/jss.v025.i01.

[21] Kassambara A and Mundt F 2016 factoextra: Extract and Visualize the Results of Multivariate Data Analyses, url: https://CRAN.R-project.org/package=factoextra, R package version 1.0.3, 2016.

[22] Maechler M, Rousseuw P, Struyf A, Hubert M, and Hornik K 2016 cluster: Cluster Analysis Basics and Extensions, url: https://cran.r-project.org/web/packages/cluster/index.html, R package version 2.0.4.

[23] Kassambara A 2016 ggcorrplot: Visualization of a Correlation Matrix using ’ggplot2’, url: https://CRAN.R-project.org/package=ggcorrplot, R package version 0.1.1, 2016.

[24] Paradis E, Claude J, and Strimmer K 2004 APE: analyses of phylogenetics and evolution in R language, Bioinformatics, 20, 289–290, url: http://bioinformatics.oxfordjournals.org/content/20/2/289.abstract.