Multivariate Analysis of Limestone Petrography Data on Kalipucang Formation Using R

A Darul1*, D E Irawan2, J Ramdani1, F Septiana1 and S S Sholihat3

1Fakultas Teknik dan Desain, Institut Teknologi Sains Bandung, Jalan Ganesha Boulevard LOT A-1 CBD Deltamas, Bekasi, Indonesia
2Fakultas Ilmu dan Teknologi Kebumian, Institut Teknologi Bandung, Jalan Ganesha no. 10 Bandung, Indonesia
3Fakultas Ilmu Pendidikan, Universitas Pendidikan Indonesia, Jalan Dr. Setiabudhi no. 229, Bandung, Indonesia

*darul@itsb.ac.id

Abstract. Limestone may look physically and visually have the same characteristics but basically can be distinguished one of them with thin section rock analysis, this research to support petrography analysis of thin section rock data with a more quantitative approach using the method Principal component analysis and cluster analysis statistics to parse physical and visual characteristics. The study used 57 samples of thin section rock from three locations in a single rock formation, i.e. the location of Pancatengah-Tasikmalaya (PCT); Cijulang-Ciamis (CJL) and Sindangsari-Ciamis (SDS), analysed using open source software with replication R programming languages using packages: ggplot2; Dplyr; FactomineR; FactoExtra; Cluster and Ggcorrplot. The results of the study showed consistently the existence of three significant rock sampling classifications, i.e. one group showing the samples were in the area near the deposition with the main compositon of foraminifera, algae, mud carbonate, coral fragments, Group 2 showed mixing with igneous rock with plagioclase composition, opaque, Glass, pyroxene and, Group 3 shows the rocks have been transported so that they are mixed with other sedimentary rocks having quartz compositions, iron oxides, rock fragments.

1. Introduction
Limestones may look similar physically and visually. However, the mineral composition varies due to the formation processes. A visual petrographical thin section is usually being used to identify mineral compositions using refraction microscope. The classification then is the final result referring to the final mineral description. The classification, however is largely based on very broad ranges. Therefore, we see the needs to classify rock samples (in this case limestones) using a more quantitative analysis.

Here we classify 57-samples of thin section of limestone from a formation using multivariate statistical method. This paper describes a machine learning technique based on multivariate analysis using R open-source statistical software. Machine learning has been used in various fields to make classification and prediction, including earth sciences [1][2][3]

Region of interest is in Pancatengah-Tasikmalaya,108°16'00"E - 7°39'6"S to 108°21'00"E - 7°44'30"S; Cijulang-Ciamis 108°24'00"E - 7°24'00"S to 108°28'23"E - 7°34'34"S and, Sindangsari-Ciamis 108°19,4'00"E - 7°37,4'00"S to 108°24'00"E - 7°44'00"S, West Java Province. The following
map refers to the systematic geological map of Indonesia, scale 1:100000 Karangnunggal quadrangle 1308-1 and Tasikmalaya quadrangle 1308-4 [4]

![Geological Map of Karangnunggal Quadrangle](image)

**Figure 1.** Research location: geological map of Karangnunggal [4]

2. **Materials and methods**

We use R, a command line application, because it could be easily replicated by copy and pasting the code and dataset. In the multivariate analysis, principal component analysis (PCA) and cluster analysis (CA), we will reduce the data, create new sets of data, then classify the data points using PCA and CA. This method was carried out in order to support petrographical classification. we, with a more quantitative approach to parsing the characters that are not derived from the physical and visual observation.

We analysed 57 samples of petrographical analysis taken from the following locations: (1) Pancatengah-Tasikmalaya/PCT (19 sedimentary rocks and 1 igneous rock sample), (2) Cijulang-Ciamis/CJL (20 samples), and (3) Sindangsari-Ciamis/SDS (17 samples). It describes mineral composition and then formatted as a csv (comma separated value) tables. The columns are the minerals from the petrographical observation, and the rows are the samples.

We used PCA to reduce variables and create new sets of variable, which will create a classification based on the data structure. Aside to only making a new classification, we also evaluated the interaction between samples in the same formation. Several add-on packages were use method using packages FactoExtra [5], Ggplot2 [6], Dplyr [7], Cluster [8] Ggcorrplot [9] and Ape [10], which packages for the R open-source statistical software and R Studio IDE [11], the equation of principal component analysis was written by Lê [12] as follows.
\[ F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_k x_{ki} m_k G_s(k), \]  
(1)

\[ G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_i x_{ki} p_i F_s(i), \]  
(2)

where \( F_s(i) \) denotes the coordinate of the individual \( i \) on the axis \( s \), \( G_s(k) \) the coordinate of the variable \( k \) on the axis \( s \), \( \lambda_s \) the eigenvalue associated with the axis \( s \), \( m_k \) the weight associated to the individual \( i \), \( x_{ki} \) the general term of the data table (row \( i \), column \( k \)). The final results are PCA plot and PCA classification using cross-plot. We stored the codes to load and prepare the data, calculate, and visualise it in our repository [13].

Table 1. A part of dataset from a total of 57 [13]

| Composition         | SDS-16 | CIL-22 | CIL-26 | PCT-42 | PCT-43 |
|---------------------|--------|--------|--------|--------|--------|
| Plagioclase         | ...    | 0      | 0      | 3      | 4      |
| Opaque              | ...    | 0      | 5      | 0      | 0      |
| Pyroxene            | ...    | 0      | 0      | 0      | 0      |
| Olite               | ...    | 0      | 0      | 0      | 0      |
| Calcite             | ...    | 0      | 0      | 0      | 0      |
| Olivine             | ...    | 0      | 0      | 0      | 0      |
| Glass               | ...    | 0      | 0      | 5      | 7      |
| Quartz              | ...    | 0      | 10     | 0      | 8      |
| Clay                | ...    | 0      | 0      | 0      | 0      |
| K. Feldspar         | ...    | 0      | 0      | 0      | 6      |
| Ganggang.merah      | ...    | 3      | 0      | 0      | 3      |
| Ganggang.hijau      | ...    | 0      | 0      | 10     | 15     |
| Fragmen.koral       | ...    | 5      | 0      | 25     | 15     |
| Fragmen.pelecipoda  | ...    | 10     | 0      | 0      | 0      |
| Fragmen.moluska     | ...    | 0      | 0      | 0      | 3      |
| Foraminifera.Besar  | ...    | 25     | 0      | 25     | 15     |
| Foraminifera.kecil  | ...    | 5      | 0      | 10     | 15     |
| Masa.spar           | ...    | 5      | 0      | 15     | 0      |
| Mikrit              | ...    | 10     | 0      | 15     | 3      |
| Fragmen.Batuan      | ...    | 35     | 0      | 15     | 15     |
| Oksida Besi         | ...    | 0      | 0      | 3      | 0      |
| Gastropod           | ...    | 0      | 0      | 0      | 3      |
| Ostracod            | ...    | 5      | 0      | 0      | 0      |
| Foraminifera.plankton| ...    | 0      | 0      | 0      | 0      |
| Detritus            | ...    | 0      | 0      | 0      | 0      |
| Glaucnione          | ...    | 0      | 0      | 0      | 0      |
| Sparry.calcite      | ...    | 0      | 0      | 0      | 0      |
| Lumper.Karbonat     | ...    | 0      | 0      | 0      | 0      |
| Echinoid            | ...    | 2      | 0      | 0      | 0      |
| Bryzoa              | ...    | 2      | 0      | 0      | 0      |

3. Results and Discussion
The results showed separation into three groups, namely the pure limestone, limestone group that has been transported and igneous rocks. Figures three and four explained 57 samples from the same formation divided into four quadrants (Figure 2):

- Quadrant 1 (top left), shows that the area is heavily influenced by the minerals in igneous rocks
- Quadrant 2 (top right) and 3 (bottom right) shows that the area is heavily influenced major minerals in the limestone
- Quadrant 4 (bottom left) shows that the area is influenced by other sedimentary rocks minerals

Figure 3 shows the amount of the mineral composition (%) in each sample. The analysis results explained there are three groups with different character difference. See figures 2 and 3.
- Cluster 1 (bottom left) PCT, CJL, and SDS with the main composition such as algae, mud carbonate, foraminifera, plankton, and coral fragments
- Cluster 2 (bottom right) SDS with main composition of pyroxene, plagioclase, and glass
- Cluster 3 (top left) PCT and SDS with the main composition of iron oxide, rock fragments, ka-feldspar and quartz.

![Figure 2. Biplot of variables and individuals](image)

There are gradations of samples in the plot. Limestone characteristics lies in the area following x-axis, while y-axis contains igneous/volcanic rocks characteristics (Figure 2). Therefore, in Figure 3, we could see a gradation from limestone samples around x-axis, to igneous/volcanic rocks in upper and lower part of y-axis.

The central part of the plot would be limestone of core-reef setting. Then samples located along x axis are limestone with a more mixing processes with volcanic or sandstone. The mixing was probably occurring in the erosion and sedimentation processes, which probably resembles fore-reef or back-reef setting. Sample no 38 is possibly a volcanic rock sample rather than it’s a mixed between limestone and volcanic rock, while sample no 8, 9 represents a mixed between the two lithological units. This method can also separate samples with various foraminifera species (on right side y-axis), with samples that has a less variation of foraminifera species (on left side x-axis).
Conclusion

Based on the results of the research we obtained PCT, CJL, and SDS are limestone that have the same character, but there are some samples that show the mixing that may occur due to the erosion, transport and sedimentation process, both with other sedimentary rocks and igneous rocks. Also we can separate one igneous/volcanic rock (basalt) sample. Based on the research, multivariate analysis may help us to classify the variables and cases, based on the similarities of the mineral composition values. We believe that multivariate statistics can support visual petrographical analysis.

Acknowledgments

Our gratitude to the librarian and laboratory of Geology program-Institut Teknologi Bandung (ITB). Dean of Faculty of Engineering and Design Institut Teknologi Sains Bandung (ITSB), as well as chairman of the mining exploration study program Ir. Mulyono Hadiprayitno, M.Si.

References

[1] E Mjolsness and D DeCoste 2001 Machine learning for science: state of the art and future prospects Science 293 5537-5551
[2] DE Irawan and CN Rachmi 2016 Multivariate statistics for hydrogeology: moving forward from “the present is the key to the past” PAAI conference proceeding 2016 available on INArxiv, URL: https://osf.io/2xv/
[3] S Sellars, P Nguyen, W Chu, X Gao, KL Hsu and S Sorooshian 2013 Computational Earth
science: Big data transformed into insight *Eos, Transactions American Geophysical Union* 94 32 277-278 doi: 10.1002/2013EO320001

[4] S Supriatna, T Apandi, S Kusumadinata, D Sudana, S Santosa, A Koswara and M Morioarfa 1978 *Peta Geologi Regional Lembar Karangnungsul* (Bandung: Pusat Penelitian dan Pengembangan Geologi)

[5] A Kassambara and F Mundt 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.

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

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

[8] M Maechler, P Rousseeuw, A Struyf, M Hubert and K Hornik 2016 *cluster: Cluster Analysis Basics and Extensions*, r package version 2.0.4

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

[10] E Paradis, J Claude and K Strimmer 2004 *APE: analyses of phylogenetics and evolution in R language* Bioinformatics 20 289–290

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

[12] S Lê, J Josse and F Husson 2008 *FactoMineR: an R package for multivariate analysis* Journal of statistical software 25 1 118

[13] J Ramdani, A Darul and DE Irawan 2017 *Analisis Statistik Multivariat Pada Batuan Beku, Sedimen, Dan Metamorf Dengan Metode Principal Component Analysis, INA-Rxiv*, url: osf.io/preprints/inarxiv/jw9h5