Geostatistics analysis as a tool to create the distribution model of Rawa Danau volcano deposits, Banten Province

M J W Mulyo, R Syahputra and T L Indra
Geoscience Study Program, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia
Corresponding author’s email: syahputra.reza@sci.ui.ac.id

Abstract. The eruption of an ancient volcano at the early Pleistocene had deposited volcanic material evenly, especially in the western part of Rawa Danau. Many smaller mountains emerged around Rawa Danau volcano as the Pleistocene volcano erupted. Then, at the end of the Pleistocene, the ancient volcano erupted for the second time with a very large force that destroyed most of the mountain and created a caldera. The volcanic material from the second eruption was distributed to evenly around the ancient volcano. We used geostatistical analysis with Inverse Distance Weighted (IDW) method to interpolate the spreading of volcanic deposits both in the first and second eruption. The objective of this research is to predict the volcanic deposits in the area which has not been researched yet and it is possible to have more eruptions in the future since geothermal manifestations such as hot springs are found in the research area. We used the outcrop data from geological mapping to interpolate the data. The output of this paper would be the interpolation map of facies deposits in Rawa Danau.

Keywords: Interpolation, volcanic, deposits, inverse distance weighted, geostatistics.

1. Introduction
Before the Rawa Danau Volcano are formed, there were already volcanic activities in the northwest part of West Java that had formed low volcanic hills of the late Tertiary age. This activity was proven by the oldest volcanic rocks exposed there. The rock classified as a Basaltic Batukuya-Anakanakan (QTvb) rock unit composed of basalt lava flows. Thus, the intensity of the volcanic activity in the area was increasing so that it presented a large ancient volcano at the beginning of the Pleistocene which is called the ancient Danau volcano. This mountain had two main eruption periods.

The first eruption period, namely the Ancient Danau volcano occurred at the beginning of the Pleistocene. That eruption left steep cliffs in the northern part of the lake named Mount Galenggang (655 meters above sea level) and Mount Sarengan (711 meters above sea level) in the form of Basaltic lava flow and volcanic breccia. The deposits from this first eruption were spread to the southwest and north area of the volcano in the form of Tufa Banten Bawah formation composed of tuff breccia, pumice breccia, tuff lapilli and agglomerate. Then, the second period of the Danau volcano eruption occurred in the middle to the end of the Pleistocene around one million years ago. The eruption was so strong that it destroyed most of the volcanic body of the Danau Volcano and left steep slopes and became a caldera. The sediment from the second period is evenly distributed around the Rawa Danau Volcano leaves the mountainous morphology which is now scattered in the northwest of Rawa Dano with several tops such as Mount Gede (774 meter above sea level), Mount Support (702 meter above sea level) Mount Pabeasan.
(588 meter above sea level) and several other hilltops with altitudes of 400 to 680 meter above sea level [1, 2]. The research area is shown in figure 1 [3].

Rawa Danau is now in the form of a lake because it has been filled with meteoric water for hundreds of thousands of years. In addition, volcanic material was deposited as a result of erosion from the caldera in the form of mud and sand, as well as water from landslide cliffs.

The author used interpolation method, especially inverse distance weighted (IDW), to obtain the lithology prediction of the area that has not been observed around Rawa Danau. This method is applied in data that have spatial properties, and usually be used in the geoscience field. The data came from the geological mapping that had been done by the authors’ team around the caldera of Rawa Danau.

2. Methodology

Interpolation is a mathematical method or function that predicts values in locations where data is not available. Spatial interpolation assumes that attributes are continuous in space and these attributes are spatially related [4]. Both assumptions indicate that the data attribute estimation can be done based on the surrounding locations. IDW is a model with a deterministic interpolation method. The IDW method used in this study assumes that the increase in the distance between the data point will reduce the local influence of that data point. By using this assumption, this method will give more weight in closer data points rather than further data points. Thus, certain radius with data points could be used to determine the output value for each location. The disadvantage of IDW interpolation is that it cannot estimate values above the maximum value and below the minimum value of the sample points [5].

IDW was used to interpolate the eruption data of the first and second periods so that the distribution of volcanic deposits of Rawa Danau was obtained. The formulation used in the IDW method was:

\[
Z_p = \frac{\sum_{i=1}^{n} \left( \frac{Z_i}{d_i^p} \right)}{\sum_{i=1}^{n} \left( \frac{1}{d_i^p} \right)}
\]

(1)

where \(d_i\) implies the distance between point \(i\) and the unknown point; \(p\), exponent parameter; \(Z\), the weight of the point. \(P\) will be an impactful factor for weighting, thus means by increasing the distance, exponential decrease in weighting could be expected. A power of 2 was used in this research [6].

Figure 1. The location of the research area, Serang, Banten Province.
Figure 2 shows the methods used in the research. The steps begin with compiling all the lithology data from all the researchers in Rawa Danau. The second step is giving value to each volcanic facies with the highest value in proximal facies. The third step is transforming the value to interval data so the software ArcGIS could process. The fourth step is processing the data into the software and finally, we get the result as an interpolation map.

The data came from geological mapping and it was not in the form of numbers. So, the author transformed the numerical data into ordinal data by giving rank to each facies of the volcanic deposits (Lykert scale). After obtaining the ordinal data in the form of rank, the author began to transform the data into interval data by using Microsoft Excel. The rank given was highest for central facies and getting lower to distal facies.

Table 1 shows the step to complete the transformation from the raw data into interval data. In the first line contains some abbreviations. Each column underwent one statistical calculation. The same value in the interval column means that the ordinal data was in zero frequency so it would not be included in the calculation of interpolation. This interpolation method was carried out using geostatistics tools of ArcGIS® 10.2 by ESRI, a specific application to produce a continuous space that depends on the data calculated from the sampled points.

| Row | Ordinal | Frek | Prop | Prop Kum | Zval | Z*Val | Sv | Interval |
|-----|---------|------|------|----------|------|-------|----|----------|
| 0   | 5       | 0    | 0.128205 | 0.128205 | -1.1349166 | 0.209516 | -1.63422 | 1 |
| 1   | 0       | 0    | 0.128205 | -1.1349166 | 0.209516 | 2.634223 |
| 2   | 0       | 0    | 0.128205 | -1.1349166 | 0.209516 | 2.634223 |
| 3   | 7       | 21   | 0.179487 | 0.307692 | -0.5024022 | 0.351642 | -0.79184 | 1.842378 |
| 4   | 0       | 0    | 0.307692 | -0.5024022 | 0.351642 | 2.634223 |
| 5   | 11      | 55   | 0.282051 | 0.589744 | 0.22688544 | 0.388805 | -0.13176 | 2.502462 |
| 6   | 0       | 0    | 0.589744 | 0.22688544 | 0.388805 | 2.634223 |
| 7   | 3       | 21   | 0.076923 | 0.666667 | 0.4307273 | 0.3636 | 0.327669 | 2.961892 |
| 8   | 9       | 72   | 0.230769 | 0.897436 | 1.26707575 | 0.178766 | 0.800947 | 3.43517 |
| 9   | 0       | 0    | 0.897436 | 1.26707575 | 0.178766 | 2.634223 |
| 10  | 4       | 40   | 0.102564 | 1         | 1.742966 | 4.377189 |
Abbreviation information:
- Row = Lykert scale scoring
- Ordinal = Ordinal data (using COUNTIF formula in Microsoft Excel)
- Frek = Lykert scale times ordinal data
- Prop = Proportion of the frequency
- Prop kum = Cumulative proportion
- Zval = Z value of each cumulative proportion
- Z*val = Z limit value (the value of the solid probability function in abscissa Z)
- Sv = Scale value
- Interval = Final result of the calculation (with equation $Sv + |Sv_{min}| + 1$)

3. Results and discussion

Figure 3 presents the surface of Rawa Danau Caldera and the sample points of lithology outcrop. The red area represents central facies containing igneous rock, breccia, and auto-breccia. It distributed in the East and Northwest of Rawa Danau caldera. Based on the regional geology map [1, 2], the eastern part of Rawa Danau caldera aged early Pleistocene and the red part in the Northwest came from the second period of the explosion, aged late Pleistocene. The green one represented proximal and medial facies containing agglomerate, tuff breccia, lapilli tuff, and lahar. It was distributed around the central facies especially in the North and Southeast area. The last one for the distal facies, coloured in blue. It was mainly in the Southwest area and in the centre of the caldera. The lithology includes tuff, conglomerates, sandstone, and claystone.

The lithology prediction is derived from the nearest 15 data points around it. The power value controls the smoothness of the surface area. The lower the value, it will give a smoother surface, if the value is higher, the results will be more precise [7].

Figure 4 shows the detail of the analysis. This analysis contained 431 samples with a mean value of 0.01672446 and root mean square (RMS) value: 0.8380873. The prediction will get near to the correct value if the value of Root Mean Square is less. Root Mean Square indicated the error in the analysis and represents the difference between the results value and the prediction value.

![Figure 3. The interpolation results](image-url)
The errors caused by uneven data distribution. There are some spaces that have no data and there are some spaces that have a lot of data surrounding. The lack of samples will affect the correct prediction because the samples that are far away from that point will affect the value of that point. Example, the eastern part of the caldera colored in red. It was affected by the igneous rock value in the steep slope of the caldera. The lithology there should be alluvium, sand and mud because it is in the form of the swamp. It contained the eroded sediment from the slope of the caldera. Nevertheless, the validation test demonstrated that samples from the geological mapping were good enough that can satisfactorily approximate the lithology distribution of the Rawa Danau volcano.

4. Conclusion
Based on the data processing, it can be concluded that the interpolation map can be used to represent the distribution of volcanic deposits of Rawa Danau. The map can visualize the spread of the deposits based on the geological mapping data. The central facies mainly deposited in the eastern and northwest area of Rawa Danau. The proximal and medial facies distributed around the central facies and the medial facies deposited in the caldera and in the southwest area of Rawa Danau. For further study, it is suggested to collect the data evenly in the research area. If it is not distributed evenly, means there is one location concentrated with data and other locations lack of data, the result will not be accurate. In this case, the center of the caldera has no data, thus the data around it will affect the lithology prediction in the center of the caldera.

Acknowledgments
This work was financially supported by Universitas Indonesia under research grant PITTA with grant contract number No. 2323/UN2.R3.1/HKP.05.00/2018.

References
[1] Rusmana E, Suwitodirdjo K and Suharsono 1991 Peta Geologi Lembar Serang, Jawa (Bandung: Pusat Penelitian dan Pengembangan Geologi)
[2] Santosa A 1991 *Peta Geologi Lembar Anyer, Jawa Barat* (Bandung: Pusat Vulkanologi dan Mitigasi Bencana Geologi)

[3] Serang, Banten Province 2018 available at www.maps.google.com

[4] Houlding S W 2000 *Practical Geostatistics: Modelling and Spatial Analysis* (Berlin: Springer-Verlag).

[5] Pramono G H 2008 *Forum Geografi* 22 97-110

[6] Agris 1998 *Reference Manual Version 5.3* (Georgia: Agris Corp. Rosewell)

[7] Bellotti F, Capra L, Sarocchi D and D’Antonio M 2010 *J. Volcanol. Geotherm. Res.* 191 117-28