Cluster analysis for slope geotechnical prioritization of intervention for the Estrada de Ferro Vitória-Minas

Abstract

This article proposes the geotechnical prioritization of intervention of slopes with landslide scars for the Estrada de Ferro Vitória-Minas by cluster analysis and also the proposition of a relationship between area and volume in landslide scars. Cluster definition helps the decision-making associated to containment measures, mapping and study of landslides for the Estrada de Ferro Vitória-Minas. The database is composed of the variables: slope’s height, inclination, scar area and scar volume. The distance measure used was Gower’s index, with Ward’s methods to build the clusters. Eight characteristic groups were identified. It was possible to identify stretches that need attention in relation to the propensity of landslides, such as Group 7, stretches 362+600, 093+xxxxE and 419+000. Group 7 presented high values for the scarred area and volume, such as maximum area 9.75 x 10^4 m² and minimum area 7.49 x 10^4 m², and maximum volume 9.20 x 10^5 m³ and minimum volume 4.08 x10^5 m³. Group 7 presented high ranges for slope height and inclination. The set of results about Group 7 can be interpreted as stretches with a predisposition for landslides. In relation to intervention measures, Group 7 presents the sections with priority. The relationship between area and volume of landslide scars obtained by the research was compared with the relationships established in literature.

Keywords: geotechnical prioritization of intervention of railway slopes, Estrada de Ferro Vitória-Minas, cluster analysis, multivariate statistics applied to geotechnical database, landslide scars.

1. Introduction

Predicting the area and volume of rock mass in landslides is a matter of extreme complexity. Some cases allowed equating landslides in different parts of the world. The empirical relationships allow estimating parameters of landslides as a function of characteristics that can be determined, such as slope inclination and height. The empirical relationships allow the assessment and delimitation of areas of risk, allowing conviviality with natural phenomena. In addition, they also allow the preparation of programs to prevent, alert, and implement containment construction (Polanco, 2010).

This study was developed taking as...
Cluster analysis for slope geotechnical prioritization of intervention for the Estrada de Ferro Vitória-Minas

received, through a contract signed with Brazil’s government, the concession for the complete operation of its rail transportation services. EFVM handled 119.2 million tons of ore, in addition to other cargo transported to other companies, such as coal and agricultural products. In relation to passenger transport, EFVM transported almost one million people in 2014. (Vale, 2017).

The present article proposes to identify clusters in the slopes with scars in a specific section of the Estrada de Ferro Vitória-Minas; this clustering allows to interpret the sectorization of the railroad and prioritization of intervention on slopes with landslides. The prioritization will be based on slope height, inclination, scar area and volume variables analyzed by cluster analysis. The cluster’s definition helps the decision-making associated to containment measures, mapping and study of landslides for the Estrada de Ferro Vitória-Minas. In this context this work can contribute to studies about landslides in Brazil, through the proposition of empirical relationships for landslide control estimates.

2. Materials and methods

The database used in this article was compiled from the data of Gomes’ work (2014) in conjunction with image analysis of the Google Earth Pro software and field work in the studied stretches of the railway. The variables collected on the slopes are slope’s height (HT) and inclination (IT). These variables are qualitative, classified by Gomes (2014) in bands presented in Tables 1 and 2.

The processing of database and cluster analysis, along with the final script was performed in software R, R Core Time (2016), version 1.0.136. Ward’s method presented better suitability for the database. According to Ward (1963) Ward’s method is to use the individuals themselves to construct the overall measure of group heterogeneity (W), that is, the method assumes that each individual forms a group. W represents the sum of the squared Euclidean distances between each individual, i, and the mean of their corresponding group, g, see Equation 1.

\[ W = \sum_{g} \sum_{i \in g} (X_i - \bar{X}_g)^2 \]

by Gower (1971) and is presented in Equation 2.

\[ G_{jk} = \left( \frac{\sum_{i=1}^{n} W_{ij} \cdot S_{ik}}{\sum_{i=1}^{n} W_{ij}} \right) \]

\[ S_{ik} = \text{Partial similarity of variable } i \text{ between the individuals } j \text{ and } k \]

The database is composed of quantitative and qualitative variables, mixed variables, and therefore with the use of Google Earth Pro software images and tools. Figure 1 shows the 493+600 stretch with a slide scar, field image and Google Earth Pro image.

Figure 1
Partial view of landslide’s scar in stretch 493+600 of EFVM (left). Image of the year 2016 from Google Earth Pro of stretch 493 + 600 with polygon marking the scar (right).
3. Results and discussions

In order to analyze the relationship between the variables to obtain the parametric relationship, a regression was performed between the two variables: scar’s area and volume. Figure 2 presents the result.

![Figure 2](image)

Relation between variables: scar’s area and volume.

The correlation coefficient between the scar’s area and volume variables was 0.66. Equations 3 and 4 present the relationship between the two variables, which was obtained by means of the graph presented in Figure 2.

\[
\log(V) = 0.7338 \log(A)^{1.2249} \\
V = 0.1416(A)^{1.2237}
\]

Equations 3 and 4 can be compared with the relationships proposed by Simonett (1967), Rice et al. (1969), Innes (1985), Guthrie and Evans (2004), Korup (2005), Imaizumi and Sidle (2007), Guzzetti et al. (2008) and Imaizumi et al. (2008). The main differences between the relationships may be due to the local geology, the data collection period, and the number of slopes studied. Table 3 presents the relationships cited and the relationship found in this research, and it is possible to conclude that the relationship is in agreement with literature.

<table>
<thead>
<tr>
<th>Author</th>
<th>Relationship</th>
<th>Cases</th>
<th>Areas bands</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present work</td>
<td>(V=0.1416\times A^{1.2237})</td>
<td>64</td>
<td>(5.8\times10^2 &lt; A &lt; 1.0\times10^3)</td>
<td>m²</td>
</tr>
<tr>
<td>Simonett (1967)</td>
<td>(V=0.2049\times A^{1.3680})</td>
<td>207</td>
<td>(2.5\times10^1 \leq A \leq 2.0\times10^6)</td>
<td>ft²</td>
</tr>
<tr>
<td>Rice et al. (1969)</td>
<td>(V=0.2340\times A^{1.1100})</td>
<td>29</td>
<td>(2.1\times10^2 \leq A \leq 2.0\times10^2)</td>
<td>m²</td>
</tr>
<tr>
<td>Innes (1985)</td>
<td>(V=0.0329\times A^{1.3852})</td>
<td>30</td>
<td>(3.0\times10^2 \leq A \leq 5.0\times10^2)</td>
<td>m²</td>
</tr>
<tr>
<td>Guthrie and Evans (2004)</td>
<td>(V=0.1549\times A^{1.0905})</td>
<td>124</td>
<td>(7.0\times10^2 &lt; A &lt; 1.2\times10^3)</td>
<td>m²</td>
</tr>
<tr>
<td>Korup (2005)</td>
<td>(V=0.0200\times A^{1.9500})</td>
<td>23</td>
<td>(A &gt; 1.2\times10^6)</td>
<td>m²</td>
</tr>
<tr>
<td>Imaizumi and Sidle (2007)</td>
<td>(V=0.3900\times A^{1.1300})</td>
<td>51</td>
<td>(1.0\times10^1 &lt; A &lt; 3.0\times10^3)</td>
<td>m²</td>
</tr>
<tr>
<td>Guzzetti et al. (2008)</td>
<td>(V=0.084\times A^{0.424})</td>
<td>539</td>
<td>(1.0\times10^1 &lt; A &lt; 1.0\times10^9)</td>
<td>m²</td>
</tr>
<tr>
<td>Imaizumi et al. (2008)</td>
<td>(V=0.1900\times A^{1.1000})</td>
<td>11</td>
<td>(5.0\times10^1 &lt; A &lt; 4.0\times10^3)</td>
<td>m²</td>
</tr>
</tbody>
</table>

Table 3
Empirical relationships between volume and area of landslide scars.

In order to calculate the similarities between the slopes, the Gower’s distance was used and to create the grouping, the Ward’s method was used. The cophenetic correlation in Ward’s method was 0.68, presenting an adequate value for the research. The choice of number groups was helped by the analysis of distance chart in groups and by the dendrogram analysis presented in Figure 3.
For analysis of the dendrogram, presented in Figure 4, with the characteristics of the slope’s scars, eight groups were selected for the studied stretches of the EFVM.

For the results of the cluster analysis, in relation to the quantitative variables, scar’s area and scar’s volume, the cluster technique was able to separate the area and volume bands characterizing each group. Figures 4 and 5 show the boxplot of the variables with the bands comprising each group for the area and volume variables, respectively.
When analyzing only the quantitative variables, it was observed that this technique was able to identify ranges for the scar’s area and volume. In the case of the scar’s area variable, the bands in some cases overlap; however, it does not happen in the whole set values because each track presents unique values. For the volume, it is possible to observe that bands differ as to maximum values, not causing differences in the minimum values, except for Group 7.

Analyzing the final results of clustering in the whole database, qualitative and quantitative variables, it was possible to build Table 4, which presents the final characteristics of each group.

Among the grouped stretches, Group 7 stands out as the group that needs the most care in relation to the landslide development. In group 7, stretches 362 + 600, 093 + xxxE and 419 + 000 present high values of scar area and volume with high ranges of slope height and inclination, which can be interpreted as areas with predisposition for landslide occurrence. In cases of intervention measures, these would be the stretches that should be given priority.

Other stretches that need attention are in Group 3. Group 3 has seven slopes with high values for slope height and inclination, as well as ranges of values for the scar’s area and volume. The stretches in Group 6 present average values for slope height with high slope inclinations, and when compared with the groups, stands out presenting high values for the scar area and volume variables.

The geographical aspect was analyzed, and allowed to sectorize the stretches in EFVM by cluster analysis. The groups formed are identified by characteristics of stretches in the group. Table 5 presents the stretches of EFVM with the results of cluster analysis.
4. Conclusions

The application of cluster analysis allowed the identification of eight groups with intrinsic characteristics. In the classified groups, there can be identified stretches of group 7, such as 362 + 600, 093 + xxxE and 419 + 000 with a tendency for landslides, due to the high values for slope height and inclination, along with considerable dimensions of scar area and volume. Other groups can be cited, groups 3 and 6, with slope height and inclination values and considerable scar area and volume values.

Acknowledgements

The authors are grateful to Universidade Federal de Ouro Preto (UFOP), Fundação de Desenvolvimento da Pesquisa de Minas Gerais (FAPEMIG) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financial support and to Núcleo de Geotecnia da Escola de Minas of the Universidade Federal de Ouro Preto.

References


INNES, J.N. Magnitude–frequency relationships of debris flows in northwest Europe.

Table 5
Sectorization of stretches in EFVM.

<table>
<thead>
<tr>
<th>Stretches</th>
<th>Group</th>
<th>Stretches</th>
<th>Group</th>
<th>Stretches</th>
<th>Group</th>
<th>Stretches</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>011+000 D</td>
<td>1</td>
<td>094+330D</td>
<td>8</td>
<td>458+600</td>
<td>3</td>
<td>524+108</td>
<td>4</td>
</tr>
<tr>
<td>011+000 E</td>
<td>2</td>
<td>094+330E</td>
<td>1</td>
<td>465+042A</td>
<td>4</td>
<td>535+419</td>
<td>4</td>
</tr>
<tr>
<td>018+500</td>
<td>3</td>
<td>100+800</td>
<td>4</td>
<td>465+042B</td>
<td>4</td>
<td>535+420</td>
<td>4</td>
</tr>
<tr>
<td>019+500</td>
<td>2</td>
<td>103+100D</td>
<td>2</td>
<td>472+228</td>
<td>3</td>
<td>535+421</td>
<td>4</td>
</tr>
<tr>
<td>021+700</td>
<td>2</td>
<td>103+100E</td>
<td>1</td>
<td>472+307</td>
<td>2</td>
<td>093+xxxD</td>
<td>6</td>
</tr>
<tr>
<td>034+000</td>
<td>4</td>
<td>104+100D</td>
<td>1</td>
<td>477+225</td>
<td>3</td>
<td>093+xxxE</td>
<td>7</td>
</tr>
<tr>
<td>056+500</td>
<td>4</td>
<td>104+100E</td>
<td>2</td>
<td>478+500</td>
<td>4</td>
<td>419+000</td>
<td>7</td>
</tr>
<tr>
<td>056+xxx</td>
<td>4</td>
<td>108+150</td>
<td>2</td>
<td>482+299</td>
<td>4</td>
<td>442+400</td>
<td>8</td>
</tr>
<tr>
<td>062+xxxD</td>
<td>5</td>
<td>359+675A</td>
<td>6</td>
<td>482+930</td>
<td>4</td>
<td>521+000E</td>
<td>4</td>
</tr>
<tr>
<td>062+xxxE</td>
<td>4</td>
<td>359+675B</td>
<td>2</td>
<td>493+600</td>
<td>3</td>
<td>521+522</td>
<td>3</td>
</tr>
<tr>
<td>063+800D</td>
<td>6</td>
<td>359+675C</td>
<td>2</td>
<td>511+077</td>
<td>4</td>
<td>079+410D</td>
<td>4</td>
</tr>
<tr>
<td>063+800E</td>
<td>4</td>
<td>361+900</td>
<td>2</td>
<td>511+670</td>
<td>4</td>
<td>079+410E</td>
<td>6</td>
</tr>
<tr>
<td>064+xxx</td>
<td>4</td>
<td>362+600</td>
<td>7</td>
<td>511+899</td>
<td>4</td>
<td>404+000</td>
<td>2</td>
</tr>
<tr>
<td>072+500</td>
<td>6</td>
<td>369+200D</td>
<td>4</td>
<td>513+200</td>
<td>3</td>
<td>409+000</td>
<td>5</td>
</tr>
<tr>
<td>073+500</td>
<td>5</td>
<td>369+200E</td>
<td>2</td>
<td>515+300</td>
<td>4</td>
<td>519+900</td>
<td>4</td>
</tr>
<tr>
<td>078+900</td>
<td>5</td>
<td>395+850</td>
<td>2</td>
<td>518+790</td>
<td>4</td>
<td>521+000D</td>
<td>2</td>
</tr>
</tbody>
</table>


R CORE TEAM. R: A language and environment for statistical computing


Received: 5 December 2017 - Accepted: 19 February 2018.