Information and computational technologies for research of natural recovery of vegetation cover

Y A Ozaryan¹*, T V Kozhevnikova², and I S Manzhula²

¹ Mining Institute, Far East Branch of the Russian Academy of Sciences, 51, Turgeneva Street, Khabarovsk, 680000, Russia
² Computing Center, Far East Branch of the Russian Academy of Sciences, 65, Kim Yu Chen Street, Khabarovsk, 680000, Russia

*Corresponding author’ e-mail: ozaryanigd@gmail.com

Abstract. The processes of natural recovery of biota in the area disturbed by mining enterprises of the Khabarovsk region have been investigated. The specifics of the formation of secondary phytocenoses on man-made tumors have been revealed. Cluster analysis methods are used as an instrument to identify the key parameters that affect the state of vegetation during natural recovery.

1. Introduction
Modern approaches to the problems of reclamation are based on the restoration or creation of economic functions of the areas disturbed by mining operations. This is absolutely true for the territories occupied by the so-called "second nature". But for the territories that belong to the zone of distribution of the natural biota located in geographically remote or inaccessible regions of Siberia and the Far East, where the economic use of lands disturbed by mining is ineffective or simply impossible, the methodology for reclamation of these lands should be based on the processes of natural recovery of biota and include two main stages. The first stage is mining, which goal is restoring the abiotic component of the disturbed area of the ecosystem and, above all, its landscape basis. The second stage is biological, aimed at ensuring the conditions for the development of natural demutational processes in the technogenic damage zones of biota [1].

The research of self-recovery of biota and its spatio-temporal dynamics on the surface of technogenic neoplasms is an urgent geoeological task, the solution of which will optimize the technology of reclamation of disturbed lands, reduce material [2], time and labor costs. Domestic and foreign researchers pay more and more attention to the study of the processes of natural recovery of vegetation in the zone of influence of mining enterprises [3–7].

Undoubtedly, the nature and intensity of self-restoration processes completely depend on the degree and characteristics of technogenic changes in the biota in each specific case and is determined through a quantitative assessment, which can be obtained through field observations of the components of the biota and, above all, of its phytocenosis. This is due to the fact that as a result of the complex anthropogenic impact, a completely different (in comparison with the primary) plant complex is formed in the territory.
2. Methodology. Study area and dataset

The methodological basis of the study was a classic systemic approach, according to which all the objects of the study are interconnected parts of the natural-technical system. Its functioning and development depend on the state and dynamics of the parameters of the external environment [8]. The work uses a set of methods: the content analysis of literature data; downloading, interpreting and analyzing medium- and high-resolution satellite images from open sources, the method of assessing the quality of the environment by the level of asymmetry of the morphological structures of vegetation. The calculation and analysis of the NDVI (Normalized Difference Vegetation Index) was carried out with using the QGIS and the Vega-Science professional information service for monitoring renewable biological resources based on the satellite data analysis. This program is provided by Space Research Institute of the Russian Academy of Sciences. The NDVI is calculated by the formula:

\[
NDVI = \frac{NIR - RED}{NIR + RED},
\]

where, NIR is the reflection in the near-infrared parts of the spectrum; RED is the reflection in the red parts of the spectrum.

The study object was the dumps of dissected rocks. The research area was the deposit of granodyorites, which have been under development for more than 100 years (Figure 1). These dumps are located on the perimeter of the quarry. They have been discontinued at various times because of natural recovery. In the area of the Korfovsky Building Stone (Granodiorite) Deposit, the surface of the dumps of dissected rocks at the age of 5, 10, 15, 20, 30, 45, and 55 years was studied. It will improve the quality of environmental measures, including reclamation and reduce the labor and material costs of the restoration of land disturbed by mining.

![Figure 1. Situational pictures of dimension stone quarry and overburden dumps](image)

To analyze the data, the hierarchical cluster analysis was used. Hierarchical clustering (also called the hierarchical cluster analysis or HCA) is a method of cluster analysis which seeks to build a hierarchy of clusters. Generally, strategies for hierarchical clustering fall into two types:

Agglomerative: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

Divisive: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.
In general, the merges and splits are determined in a greedy manner. The results of hierarchical clustering are usually presented in a dendrogram.

Hierarchical clustering algorithms assume that the analyzed set of objects is characterized by a certain degree of connectivity. By the number of signs, sometimes monothetic and polythetic methods of classification are identified. Like most visual ways of representing dependencies, graphs quickly lose clarity as the number of clusters increases. There are a number of specialized programs for building dendrograms [9, 10].

The object of the study was the secondary plant communities developing on the man-made dumps of open-cut deposits. The subject of the study is the processes of natural recovery of phytocenoses.

The primary challenge in the research is the relationship between the research objects and the signs of recovery.

The purpose of the study is to establish the patterns, criteria and conditions for self-restoration of phytocenoses on the surface of technogenic neoplasms using clustering.

The primary challenge in the research is the relationship between research objects and signs of recovery.

The purpose of the study is to establish patterns, criteria and conditions for self-restoration of phytocenoses on the surface of technogenic neoplasms using clustering.

The scientific hypothesis is formulated as follows: the quality and speed of natural vegetation recovery is directly proportional to the time factor and the distance to natural plant communities.

For more detailed research of the relationships between the researched objects we used the hierarchical clustering methods.

2.1. Methodology. Study area and dataset

At the beginning of the process, each element is in a cluster of its own. The clusters are then sequentially combined into larger clusters until all elements end up being in the same cluster. The result of the clustering can be visualized as a dendrogram, which shows the sequence of cluster fusion and the distance at which each fusion took place.

The distance between two clusters is assumed to be equal to the maximum distance between two elements from different clusters.

$$\max \{d(a,b): a \in A, b \in B\},$$  \hspace{1cm} (2)

where $d(a,b)$ is the distance between elements $a$ and $b$, owned by clusters $A$ and $B$.

2.2. Unweighted pair-group method using arithmetic averages

The following linkage criterion is used

$$\frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a,b),$$ \hspace{1cm} (3)

where $d(a,b)$ is the distance between elements $a$ and $b$, owned by clusters $A$ and $B$, $|A|$ and $|B|$ are cluster capacities.

2.3. Single-linkage clustering

Single-linkage clustering, the method is also known as nearest neighbour clustering. The distance between two clusters is assumed to be equal to the minimum distance between two elements from different clusters:

$$\min \{d(a,b): a \in A, b \in B\},$$ \hspace{1cm} (4)

where $d(a,b)$ is the distance between elements $a$ and $b$, owned by clusters $A$ and $B$. 

3
2.4. Complete-linkage clustering using distance based on the correlation
At the beginning of the process, each element is in a cluster of its own. The clusters are then sequentially combined into larger clusters until all elements end up being in the same cluster. On each level of the algorithm two clusters are combined. The distance between these clusters is minimum. The formalization of the concept of “minimum distance” may depend on modifications of the algorithm; in the complete-linkage clustering the minimum distance is defined as the maximum of the set of distances between an element of the first cluster and an element of the second cluster. That is, the distance $D(A, B)$ between clusters $A$ and $B$ is calculated by the formula:

$$D(A, B) = \max_{a \in A, b \in B} d(a, b),$$

where $d(a, b)$ is the distance between elements $a \in A$ and $b \in B$, $A$ and $B$ are different clusters.

2.5. Feature scaling hierarchical clustering
Feature scaling hierarchical clustering is the method used to normalize the range of the explanatory variables or characteristics of the data. In our case, to calculate the distance between clusters, we first calculate the mean and standard deviation of the entire vector, then “scale” each element by these values, subtracting the mean and dividing by the standard deviation.

![Figure 2. Dynamics of the NDVI secondary phytocenosis of dumps of the Korfovsky stone quarry.](image)

3. Results
It has been established that even under the conditions of a two-hundred-year or more technogenic impact of mining enterprises, ecosystems still retain the ability to function [11]. The process of self-recovery of natural biota after the end of mining operations and the closure of the enterprise and the implementation of recultivation measures will be determined by the regenerative resources of the biota outside the zone of its technogenic damage, as well as by the size of this zone and the climatic features of the region.

Natural recovery of the biota was recorded at all the objects studied, but its dynamics is significantly different.

To return the ecosystem to its original state, a cycle of successively changing stages of the state
and composition of the studied environment is required, which is called the succession.

The revealed patterns of changes in the density of undergrowth and its species composition in the territories of secondary forest communities located near the preserved areas of primary phytocenoses indicate that the distance of seed transfer of the main forest-forming species can be used as one of the criteria for regulating the size of individual land allotments in the formation of the infrastructure of the surface complex of a mining enterprise.

The total area occupied by dumps at the Korfovsky stone quarry is at least 20.89 hectares. The Korfovsky stone quarry is located at the border of the Bolshekhekhtsirsky State Nature Reserve, in this regard, the flora and fauna of the adjacent territories is very diverse. Various types of deciduous, coniferous-deciduous and dark coniferous forests are widely represented here.

The study of demutational processes was carried out on rock dumps when examining their vegetation cover (Table 1). The pioneer vegetation on the waste dumps is represented by grasses of the families: legumes (Trifolium hybridum, Melilotus officinalis), cereals (Alopecurus pratensis, Poa pratensis), asters (Artemisia vulgaris, Tanacetum vulgare, Taraxacum officinale, Crépis tectorum, etc.). Among woody species, for example, Salix schwerinii, Salix udensis, Populus maximowiczii, Betula alba, and Alnus are involved in demutational processes [12].

Table 1. Characteristics of natural resumption of vegetation at rock dumps

| Stage of formation | Number of species, pc./% |
|--------------------|-------------------------|
|                    | Total          | Woody plant | Shrub   | Herbaceous plants |
| 5–10 years         | 8/100          | 2/25        | 1/12,5  | 5/62,5            |
| 15–25 years        | 13/100         | 2/18        | 2/18    | 10/76,92          |
| 40–50 years        | 23/100         | 4/17,39     | 3/13,04 | 16/69,56          |
| Background value   | 46/100         | 16/34,78    | 7/15,22 | 23/49,99          |

The first stage of demutational processes is the basis of the succession series [13-14]; in the Korfovsky stone quarry, there are the herb-legume and cereal-herb communities. It was revealed that the herb community with a predominance of legumes was replaced by the shrub community with a predominance of downy alder, and in the case of the presence of cereals, the superiority of Salix schwerinii was noted. In addition, there is a replacement of shrubs in the first case with a young birch forest and a mixed tree-shrub community in the second one.

Obviously, there is no clear sequence in the change of stages of the succession series. The appearance of a particular species in a phytocenosis depends on the nature of the transfer of seeds (or spores) that takes place at a certain time of the year, or other specific environmental conditions.

In the first five years, the regeneration of vegetation is small: singly, Artemisia vulgaris, Alopecurus pratensis, Gramineae, Trifolium hybridum L., and Taraxacum officinale, these plants are up to 5 cm height, and their number does not exceed 18 specimens per 1 m². On dumps that have existed for 10 or more years and which are located near the forest, there are single tree species: Betula alba, Salix schwerinii, and Populus maximowiczii [8, 15].

In the next ten years, the species composition of the herbaceous cover is characterized by significant diversity and is represented by the following species: Trifolium hybridum, Oenothera biennis, Trifolium lupinaster, Crépis tectorum, Artemisia vulgaris, Artemisia scopária, Taraxacum officinale, Plantago maior, and Pulsatilla. Some species fade out, whereas others appear.

On the site within the boundaries of the background territory, 46 species were recorded, which is exactly twice the number of species on the 50-year-old dump, 16 species of woody plants, 7 species of shrub and 23 species of herbaceous plants were identified. It should be noted that the soil and vegetation cover is more diverse here, most species are not found in the zone of influence of the enterprise.

To identify the reasons for the development of natural recovery and identify the factors that
determine the spatio-temporal dynamics of the process, an analysis of the dependence of the parameters of the sites studied was performed. In Table 2, the data are presented as follows: the rows of the table indicate the objects under study, which are considered overburden dumps of different ages, the columns of the table indicate the factors that characterize each of the objects under consideration.

| NDVI | t, years | S, ha | L, m | t, years | Herbaceous plants (species) | Woody plant (species) | Shrub (species) | Total numbers of species |
|------|----------|-------|------|----------|-----------------------------|----------------------|-----------------|------------------------|
| 5    | 10       | 2.64  | 100  | 5        | 3                           | 0                    | 0               | 3                      |
| 10   | 15       | 1.12  | 100  | 15       | 4                           | 0                    | 0               | 4                      |
| 20   | 30       | 3.29  | 250  | 20       | 10                          | 2                    | 1               | 13                     |
| 30   | 45       | 2.77  | 750  | 30       | 8                           | 3                    | 3               | 14                     |
| 55   | 45       | 3.98  | 1350 | 55       | 16                          | 4                    | 3               | 23                     |
| Background value | 6.8 | 0 | >100 | 23 | 16 | 7 | 46 | |

Where L, m is the distance to natural plant communities; t, years is the vegetation recovery time; Herbaceous plants (species) denote the number of herbaceous species; Woody plant (species) denote the number of woody species; Shrub (species) show the number of shrub species; Total numbers of species means the total number of plants in a point.

In order to identify the tightness of relationships between the factors, we have built the correlation matrix in which the correlation coefficient between the corresponding parameters is found at the intersection of the corresponding row and column (Table 2).

| S, ha | L, m | t, years | Herbaceous plants (species) | Woody plant (species) | Shrub (species) | Total numbers of species |
|-------|------|----------|-----------------------------|----------------------|-----------------|------------------------|
| 1     | -0.04152 | 0.823832 | 0.757478                   | 0.78924              | 0.73868         | 0.785979              |
| -0.04152 | 1 | -0.25294 | 0.262018                   | -0.08335             | 0.133436        | 0.119298              |
| 0.823832 | -0.25294 | 1 | 0.823645                   | 0.976025             | 0.882043        | 0.911428              |
| 0.757478 | 0.262018 | 0.823645 | 1 | 0.907137 | 0.916624 | 0.977873 |
| 0.78924 | -0.08335 | 0.976025 | 0.907137 | 1 | 0.953637 | 0.973017 |
| 0.73868 | 0.133435 | 0.882043 | 0.916624 | 0.953637 | 1 | 0.968063 |
| 0.785979 | 0.119298 | 0.911428 | 0.977873 | 0.973017 | 0.968063 | 1 |

Interpreting the results of the correlation analysis presented in Table 3, we can conclude that there is a strong linear relationship ($R^2 > 0.80$) between the following factors:

- vegetation recovery time and number of herbaceous species, number of woody plant species, number of shrub species, total number of plants in a point;
- number of herbaceous species and number of shrub species, total number of plants in a point;
- number of woody plant species and number of shrub species, total number of plants in a point.
3.1. Complete-linkage clustering
We obtained a dendrogram by division into four clusters (Figure 3), in which the background belonged to the first cluster, the dumps with the oldest age (45 and 55 year-old) were combined into the second cluster, the “youngest” dumps were included in the third cluster (5-year-old and 10-year-old), and the fourth cluster included the dumps of middle age (15, 20 and 30 year-old). We obtained a dendrogram by division into five clusters, in which the background again fell into the first cluster, but the 45-year-old and 55-year-old dumps formed separate clusters, the fourth cluster included the “youngest” dumps (5-year-old and 10–year-old), and the fifth cluster included the dumps of middle age (15, 20 and 30 years old).

Figure 3. Cluster analysis dendrograms by complete-linkage clustering: a – 4 clusters; b – 5 clusters.

3.2. Unweighted pair-group method using arithmetic averages
The results of cluster analysis by division into four clusters are presented on the dendrogram (Figure 4).

We obtained a dendrogram by division into four clusters, in which the background again fell into the first cluster, the 55-year-old dumps were included in the second cluster, the third cluster included the 45-year-old dumps, and the fourth cluster included the dumps of the “youngest” and middle age (5, 10, 20 and 30 years old).

Figure 4. Cluster analysis dendrograms by unweighted pair-group method using arithmetic averages: a – 4 clusters; b – 5 clusters.

3.3. Single-linkage clustering
The results of cluster analysis by division into four clusters are presented on the dendrogram (Figure 5).

The background belonged to the first cluster, the 55-year-old dump was included in the second cluster, the 45-year-old dump was included in the third cluster, and the middle and “young” age dumps
were combined into the fourth cluster (5, 10, 15, 20 and 30-year-old). We can conclude that this method is similar to the unweighted pair-group method using arithmetic averages.

The dendrogram which was divided into five clusters has the following characteristics: the first cluster is the background, the 45-year-old and 55-year-old dumps formed separate clusters, the fourth cluster was formed of the middle-aged dumps (20 and 30 year-old), and the fifth cluster united the "young" age dumps (5, 10, and 15 year-old).

Using clustering methods based on the minimum differences of features, regardless of the quantity into which the features are divided, we obtain the following partitions:

- the background is always allocated in a separate class;
- the youngest and oldest dumps form separate clusters;
- middle-aged dumps are formed either into one class, or each dump forms its own cluster, depending on the clustering method.

3.4. Complete-linkage clustering using distance based on the correlation

The dendrogram which was separated into four clusters has the following characteristics (Figure 6): the first cluster includes the 45-year-old and 55-year-old dumps, the 15-year-old and 20-year-old dumps are combined into the second cluster, in the third cluster there are 5 year-old and 10-year dumps, and the background and 30-year-old dumps are combined into the fourth cluster.

3.5. Feature scaling hierarchical clustering

The dendrogram which was separated into four clusters has the following characteristics (Figure 7): the background belongs to the first cluster, the 45-year-old and 55-year-old dumps are included in the
second cluster, the 5-year-old and 10-year-old dumps are included in the third cluster, and the middle age dumps are combined into the fourth cluster (15, 20 and 30 years old). When applying methods that use data scaling, clustering is similar to clustering based on minimal feature differences. However, it should be noted that when using the complete-linkage clustering with distance based on correlation, the background does not stand out in a separate cluster, but falls into a cluster with a 30-year dump.

When applying methods that use data scaling, clustering is similar to clustering based on minimal feature differences. However, it should be noted that when using the complete-linkage clustering with distance based on correlation, the background does not stand out in a separate cluster, but falls into a cluster with a 30-year dump.

Figure 7. Cluster analysis dendrograms feature scaling hierarchical clustering: a – 4 clusters; b – 5 clusters.

Analysis of the results obtained using five methods of cluster analysis and the correlation matrix allowed for the following main conclusions:

- The most similar dumps of 5, 10, 15, 20, and 30 year-old age are combined by various combinations into one or two clusters, but are stably close in the parameters for all research methods.
- In three cases out of five, the backgrounds of 45, 55 year-old dumps are allocated to separate clusters.

Taking into account the qualitative and quantitative parameters of vegetation restoration and the results of cluster analysis, it can be concluded that the scientific hypothesis is correct.

Acknowledgments
Experimental studies were carried out using the resources of the Center for Shared Use of Scientific Equipment «Mineral Research Center» KhFRC FEB RAS.

The studies were carried out using the resources of the Center for Shared Use of Scientific Equipment "Center for Processing and Storage of Scientific Data of the Far Eastern Branch of the Russian Academy of Sciences", funded by the Russian Federation represented by the Ministry of Science and Higher Education of the Russian Federation under project No. 075-15-2021-663.

ORCID iDs
Y A Ozaryan, [ID] https://orcid.org/0000-0002-8892-4595
T V Kozhevnikova, [ID] https://orcid.org/0000-0002-2660-3104
I S Manzhula, [ID] https://orcid.org/0000-0002-1298-1470

References
[1] Trubetskoy K N 2001 Subsoil and ensuring the ecological safety of their development Subsoil development and environmental problems – look into the XXI century (Moscow)
[2] Arkhipova Yu A 2016 Economic evaluation of activities of the mining entities of non-ferrous metallurgy of the far east of the russian federation Mining Inf. & Analyt. Bull. S21 pp 374–88
[3] Strilesky S L, Humphreys E R and Carey S K 2017 Forest water use in the initial stages of reclamation in the Athabasca Oil Sands Region Hydrological Processes 31 pp 2781–92
[4] Wood M E, Macrae M L, Strack M, Price J S, Osko T J and Petrone R M 2016 Spatial variation
in nutrient dynamics among five different peatland types in the Alberta oil sands region *Ecology and Hydrology* 9(4) pp 688–99

[5] Nwaishi F, Petrone R M, Price J S and Andersen R 2015 Towards Developing a Functional-Based Approach for Constructed Peatlands Evaluation in the Alberta Oil Sands Region, Canada *Wetlands* 35(2) pp 211–25

[6] Petrone R M, Chasmer L, Hopkinson C, Silins U, Landhäusser S M et al. 2015 Effects of harvesting and drought on CO₂ and H₂O fluxes in an aspen-dominated western boreal plain forest: early chronosequence recovery *Can. J. Forest Res.* 45(1) pp 87–100

[7] Zenkov I V, Yuronen Yu P, Nefedov B N and Baradulin I M 2016 Remote sensing in estimation of forest ecosystem generation at crushed stone quarries in Siberia *Eurasian Mining* 1 pp 50–4

[8] Ozaryan Yu. A 2018 Assessment of natural biota rehabilitation in the influence zone of mining in the Khabarovsk territory by satellite monitoring data *Mining J.* 10 pp 84–8

[9] Berezikov V S and Lbov G S 2008 Modern trends in cluster analysis *All-Russian competitive selection of review and analytical articles in the priority direction ‘Information and telecommunication systems’* p 26

[10] Mahruse Nasma 2019 Modern Trends in Data Mining Methods: Clustering Method *Moscow Econ. J.* 6 pp 359–77

[11] Ivanova N S, Shipilova E V and Makarov A B 2018 Recultivation of mining landscapes: new view *Mining J.* 8 pp 102–107

[12] Ozaryan Yu A 2014 Features of the biota recovery in the zone of the industrial impact of the far eastern mining enterprises *Mining Inf. & Analys. Bull.* 2 pp 372–79

[13] Manakov Ju A and Kupriyanov A N 2009 Criteria for diagnostics of primary succcession stages on the rockdumps of kuzbass *Mining Inf. & Analys. Bull.* 7 pp 186–93

[14] Klaas G J N, van Lagen B and Buurman P 2001 Composition. of plant tissues and soil organic matter in the first stages of a vegetation succession *Geoderma* 100 pp 1–24

[15] Galchenko Ju and Ozaryan Ju 2018 Method of quantitative assessment of the regularities of natural restoration of biota in zones of technogenic disturbance by extractive enterprises VII *International Scientific Conference “Problems of Complex Development of Georesources”* (PCDG 2018) 56 doi: https://doi.org/10.1051/e3sconf/20185604006