Fuzzy cluster analysis of indicators for assessing the potential of recreational forest use

Evstakhii Kryzhanivs'kyi[0000-0001-6315-1277], Liliana Horal[0000-0001-6066-5619], Iryna Perevozova[0000-0002-3878-802X], Vira Shiyko[0000-0002-2822-0641], Nataliia Mykytiuk[0000-0001-3194-3891] and Maria Berlous[0000-0003-2856-9832]

Ivano-Frankivsk National Technical University of Oil and Gas,
15 Karpatska Str., Ivano-Frankivsk, 76019, Ukraine
rector@nung.edu.ua, liliana.goral@gmail.com, perevozova@ukr.net,
viraSh@i.ua, nataliamykytiukmmm@gmail.com, masher@i.ua

Abstract. Cluster analysis of the efficiency of the recreational forest use of the region by separate components of the recreational forest use potential is provided in the article. The main stages of the cluster analysis of the recreational forest use level based on the predetermined components were determined. Among the agglomerative methods of cluster analysis, intended for grouping and combining the objects of study, it is common to distinguish the three most common types: the hierarchical method or the method of tree clustering; the K-means Clustering Method and the two-step aggregation method. For the correct selection of clusters, a comparative analysis of several methods was performed: arithmetic mean ranks, hierarchical methods followed by dendrogram construction, K-means method, which refers to reference methods, in which the number of groups is specified by the user. The cluster analysis of forestries by twenty analytical grounds was not proved by analysis of variance, so the re-clustering of certain objects was carried out according to the nine most significant analytical features. As a result, the forestry was clustered into four clusters. The conducted cluster analysis with the use of different methods allows us to state that their combination helps to select reasonable groupings, clearly illustrate the clustering procedure and rank the obtained forestry clusters.

Keywords: cluster analysis, k-means clustering method, forestry, recreation.

1 Introduction

The intensive development of recreation in the world creates motivation to use significant reserves of recreational resources. To expand the use of forest recreational resources, it is necessary to use for this purpose not only nature reserves, but also to involve more and more forests of state forestry farms in this use. The reserves of recreational forest use on the territory of Ukraine are significant. Therefore, there is a need to assess their development on the basis of the classification of forestry areas on many analytical grounds. Taking into account the fact that such classification is a rather time-consuming task, it is proposed to carry out forests clustering with the help of software.

The use of cluster analysis methods is dictated primarily by the fact that they help to build scientifically based classifications, identify internal links between the observed population units. In addition, cluster analysis methods can be used to compress information, which is an important factor in the conditions of constant increase and complication of statistical data flows. That is why this type of statistical analysis is of great importance when analyzing the development of recreational facilities. It should be noted that recently cluster analysis has received considerable attention from domestic and foreign experts in various scientific fields. One of the reasons is that modern science is increasingly relying on classification for its development. Moreover, this process deepens as knowledge specialization grows, which in its turn is based largely on objective classification. Another reason is related to the accompanying deepening of specialized knowledge, the increase in the number of variables, taken into account in the analysis of certain objects.

Clustering of the studied forests will allow the effective management of recreational areas, taking into account the reserves for improving the development of areas for selected components and also to develop at the state level the Strategy of recreational forest use development in Ukraine for the maintenance of the National recreational product competitive in the domestic and world markets. Taking into consideration the fact that each region of Ukraine is characterized by its natural and climatic conditions, ethnic traditions and historical and cultural recreational features, there is a problem of qualitative analysis and assessment of the level of recreational facilities development.

2 Background

The foreign scientists, who studied the issue of recreational forest management, are Simon Bell [2], William M. Murphy [11], Lloyd C. Irland, Darius Adams, Ralph Alig, Carter J. Betz, Chi-Chung Chen, Mark Hutchins, Bruce A. McCarl, Ken Skog and Brent L. Sohngen [6], Nerida Anderson, Rebecca M. Ford, Lauren T. Bennett, Craig Nitschke and Kathryn J. H. Williams [1], Artti Juutinen, Anna-Kaisa Kosenius and Ville Ovaskainen [7], Markus A. Meyer, Joachim Rathmann and Christoph Schulz [10], Tina Gerstenberg, Christoph F. Baumeister, Ulrich Schraml and Tobias Plieninger [5], Kee-Cheo Lee and Kee-Rae Kang [9], Hyun-Kyu Shin and Hong-Chul Shin [12], Yevstakhii Kryzhanivskyi, Liliana Horal, Vira Shyiko, Oleksii Holubchak and Nataliia Mykytiuk [8].

Markus A. Meyer, Joachim Rathmann and Christoph Schulz proved in [10] that visitors cluster along major paths or regions in urban and rural forest, recreation of the local population is highly driven by relaxation, forest structures and demographic factors play a minor role for forest benefits, forest benefits do not strongly vary within the area of the forests, forest management should focus on avoiding nuisances to support forest benefits. They found a weak connection between recreational behavior and demand for specific forest characteristics. For local recreation, we recommend to provide a basic level of highly rated FB and to avoid nuisances rather than designing forests for a desired appearance.

Tina Gerstenberg, Christoph F. Baumeister, Ulrich Schraml and Tobias Plieninger in [5] identified frequencies of activities in urban forests, visualized activity-specific hot routes, and unveiled the contributions of landscape features to recreational use intensity. The hot route maps represent an advancement of existing forest function maps, as they were based on more reliable spatially explicit data on where people move in forests. They used a public participation mapping procedure as a basis for visualizing recreational use intensity. These maps may aid forest managers to tailor management according to residents' forest uses and preferences, prioritize objectives, and prevent conflicts between re-creational user groups, conservationists and representatives of the timber industry. They conclude that urban forest managers may promote outdoor recreation by maintaining large proportions of broadleaved dominated stands. Finally, accessibility to water bodies as well as unique structural compositions – as represented by protected habitats – may enhance recreational use [5].

The purpose of Kee-Cheo Lee and Kee-Rae Kang [9] is to classify the forests by considering the supplier's perspective as well as the user's perspective in order to provide fundamental materials for the operation of the natural recreation forests. A factor analysis was conducted to identify the common characteristics of the selected twelve variables by pre-selection and survey of experts. K-means cluster analysis was conducted among those factors to classify the natural recreation forests in Korea. Four factors were drawn after the factor analysis and the factors were named according to the variables and sizes as 'The use performance and visiting condition factor', 'Education and settlement factor', 'Internal activation factor' and 'Potential factor'. In addition, the cluster analysis of the matrix was conducted for the points of the drawn factors and the final classification consists of five groups. The results of this study may contribute to providing fundamental materials for the operation and management of natural recreation forests. Also, it may act as a reference when investigating the natural recreation forests of Korea. Proposing the classification natural recreation forests could be helpful in selecting the proper recreation forest in the future. Based on the established model, fundamental materials could be provided to improve the profitability of the natural recreation forests by effectively expanding the number of tourists, creating new natural recreation forests and proper maintenance and management [9].

Hyun-Kyu Shin and Hong-Chul Shin in [12] segmented recreational forest's visitors for marketing based on purpose of visit. Using the factor analysis, cluster analysis, cross tab, and t-test to find out different behavioral intention in each cluster, the result elicited some implications. First, 2 clusters were founded and has difference in behavioral intentions. Cluster 1 (married, 200~300 hundred won income) has higher satisfaction, revisit intention, recommendation intention. The result shows that market researcher in recreational forest should approach different marketing strategy and has various facilities, active program. This research needs to survey broad region to generalized result [12].

Thus, having considered the scientific works of both foreign and domestic researchers of the recreational forest management problems and without diminishing their scientific value to improve development of recreational forest management, it is possible to consider and necessarily classify the recreational region for a component that is its own manufacturer [8].

3 Methodology

As it is known, for complex evaluation of every economic process or its components, the methods of integrated indicators calculation are conventionally applied using different economic and mathematical methods and approaches. The complex evaluation is required to define the potential of recreational forest management, considering the development of all its components. Therefore, in [8] we propose to evaluate the potential of recreational forest use by performing the following steps: to identify the recreational forest use potential components; to develop and form a system of quantitative and qualitative indicators (indices) in order to evaluate the efficiency of recreational forest use potential by its component composition; to evaluate the efficiency of recreational forest use of the regional territories by individual components of the recreational forest use potential using certain indicators; to comprehensively evaluate the efficiency of each recreational forest use potential component; to conduct an integrated evaluation of the efficiency of recreational forest use by means of using taxonomic analysis methods and fuzzy set theory; to determine the level of the recreational forest use potential by comparing the integrated indicator value with its standard (critical) values [8]. Based on the previous studies of recreational forest management, the following structural components of recreational forest management potential can be formed: a resource component, social component, economic component, innovation and investment component. Each component of recreational forest use is characterized by a system of performance indicators. According to the above characteristics of each component, the following system of indicators can be proposed, considering the attributes of recreational activity, which are listed in table 1 [8].

Economic and mathematical modelling of evaluation of the recreational forest management potential determined the efficiency of recreational forest use of regional territories by individual components of recreational forest management potential using indicators specified in table 1. A taxonomic method based on determination of taxonomic indicators of each component [8] was used for this stage.

To approve the methodology of assessing the recreational potential of forest use, a typical forestry of the Western region of Ukraine was selected, including 8 forestries. It is worth mentioning that as a result of the underdeveloped information and statistical infrastructure of forestries, it was not possible to calculate a required system of indicators, shown in table 1. However, the taxonomic indicators were calculated based on the actual statistical base on the resource and social components of each forestry. The calculation results of forestry activity were summarized in table 2.

Therefore, based on obtained calculations we can conclude that recreational forest management in Ukraine is low, confirmed by the level of recreational forest management potential (table 2). Of 8 analysed forests only in Forestry 1 the potential level is average, in two forestries the integrated indicator of recreational forest management potential level has been set at a level below average, and the remaining 5 forests have a low level of recreational forest management. Graphically obtained results are shown in figure 1 [8].

Table 1. Evaluation indicators of the recreational forest management potential components.

Component	Indicator	Substantiation
Resource component	Area of recreational territories, km^2	Total area of forestry intended for recreational forest use
	Number of recreational places, quantity	Number of recreational places located on the forestry territory intended for recreational forest management
	The level of attractiveness of natural and recreational resources	The indicator can be evaluated according to the following criteria: exoticism, uniqueness, aesthetics, comfort, etc.
	Quality factor of forest vegetation	It describes the level of recreation applicability
	Exoticism degree (contrast) of recreational territory	It is determined as a contrast ratio degree of the resting place relative to a recreant's permanent residence
Economic component	Proportion of total forestry costs on maintenance of recreational places, %	It shows the proportion of the total costs on maintenance of recreational territories
	Efficiency factor of recreational forest management	It shows attractiveness of recreational forest management
	Wear coefficient of recreational fixed assets (FA)	It characterizes wear level of recreational fixed assets
	Volume of marginal costs for growing 1 ha of recreational forest	They reflect the effect, achieved by improving the forest as a means of labor in recreation sphere
	Capacity of a single recreational load	It shows the maximum permissible number of persons on recreational territory
Social component	Proportion of recreant employees	It shows a proportion of recreant employees in the total number of staff involved in recreational activities
	Recreational capacity	The capacity of recreation centres (resorts, tourist, health, recreational complexes) is a simultaneous number of recreants that can be located in this centre, without disturbing ecological balance within this centre and surrounding territories
	Recreational load per 1 ha of forest	It determines attendance intensity for any segment of the day, during weekends, weekdays
	The average stay of vacationers on the recreational territory, h	It shows an average length of stay of visitors on the recreational territory of forest area
Innovation and invest-	Cost amount on marketing activities of recreational territories	It characterizes the development level of marketing activities

Component	Indicator	Substantiation
ment compo-nent	Efficiency of innovation implementation of re-creational forest mana-gement	It characterizes the innovation level and efficiency of recreational innovation use
	Amount of investments in recreational activity	It shows the amount of investment resources aimed at recreational activities
	Proportion of foreign in-vestments in recreatio-nal activities financing	It shows amount of recreational activity financing at the expense of foreign financial sources
	Quantity of the won grants (programs) to fi-nance recreational acti-vities	It characterizes relevance of the recreational sphere development

Table 2. Taxonomic analysis results of recreational forest management of a typical forestry.

Indicator	Forestry 1	Forestry 2	Forestry 3	Forestry 4	Forestry 5	Forestry 6	Forestry 7	Forestry 8
Taxonomic indicator of resource component	1.00	0.51	0.33	0.36	0.32	0.33	0.31	0.33
Taxonomic indicator of social component	1.00	0.56	0.56	0.30	0.30	0.21	0.39	0.39
Taxonomic indicator of economic component	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Taxonomic indicator of innovation and investment component	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Integrated indicator of recreational forest management potential level	0.50	0.27	0.22	0.16	0.16	0.14	0.17	0.18

Thus, according to the results of economic and mathematical modelling of the integrated indicator of recreational forest management potential level, it can be concluded that the recreational forest management potential in Ukraine is low (figure 1), so measures should be taken to improve recreational activity results and develop this industry. As the calculations indicate, first of all, it is urgent to develop economic and innovation investment components of the recreational forest management potential in Ukraine.

Thus, having obtained the results of calculating the integrated indicator of the recreational forest use level in the studied forests, we consider it necessary to conduct a fuzzy cluster analysis of forestry based on the analysis of forest use potential individual indicators for the studied objects. The main stages of cluster analysis of the recreational forest use level by predetermined components are shown in the figure 2.

To implement the clustering process, it is necessary to develop a matrix of observations x_{ij}. In this case, the original set consists of *m* elements described by *n* parameters, and each of its lines can be interpreted as a point or vector placed in *i*-dimensional space with coordinates equal to the value of *n* features for a particular forestry. Thus, in the observation matrix x_{ij} is the value of feature *i* for *j* forestry; *j* – a number of classification objects (forestry); *i* – a number of features of the objects.

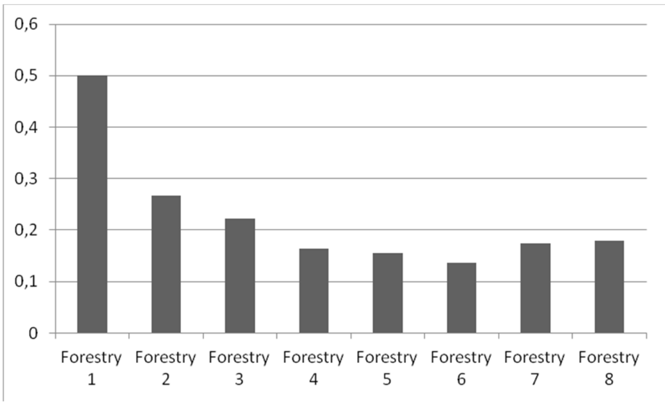

Fig. 1. Integrated indicator of recreational forest management level.

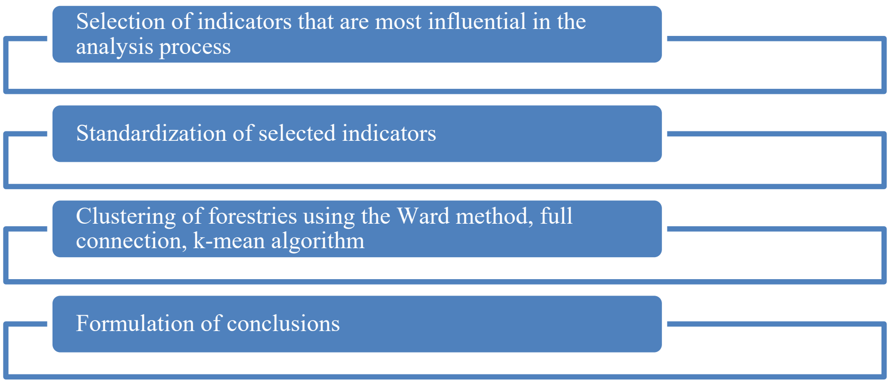

Fig. 2. The main stages of cluster analysis of the recreational forest

Using element multiplicity *w*, described by *n*-signs, each unit can be interpreted as a point of *n*-dimensional space with coordinates equal to the value of n attributes for the analysed unit. Let us represent the matrix as follows:

$$X = \begin{bmatrix} x_{11}x_{12}\ldots x_{1k}\ldots x_{1n} \\ x_{21}x_{22}\ldots x_{2k}\ldots x_{2n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x_{i1}x_{i2}\ldots x_{ik}\ldots x_{in} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x_{w1}x_{w2}\ldots x_{wk}\ldots x_{wn} \end{bmatrix} \quad (1)$$

where: w is the number of study periods, n is the number of indicators of each recreational forest management potential, x_{ik} – indicator value k of each specific component for a year ($k = 1 \div n$, $i = 1 \div w$).

As indicators of recreational forest use management level assessment are reflected in various measures, they need to be standardized. One of the most common means of statistical generalization for inhomogeneous populations is the standardization of indicators by the ratio of deviation (x_i) to the unit of standardization. In our case, σ_i is chosen as the standardization unit. These features should be normalized using the following formula:

$$z_{ij} = \frac{x_{ij} - \bar{x_i}}{\sigma_i} \quad (2)$$

when

$$\bar{x_i} = \frac{1}{w} \sum_{i=1}^{w} x_{ij} \quad (3)$$

$$s_k = \left[\frac{1}{w} \sum_{i=1}^{w} (x_{ik} - \bar{x}_k)^2 \right]^{\frac{1}{2}} \quad (4)$$

where: z_{ij} – standardized value of indicator j for the i-th study period; x_{ij} – standardized value of indicator j for the i-th study period; x_j – arithmetic mean of kj indicator; σ_j – standard deviation of k indicator; w – a number of periods.

The main feature of clusters is that objects belonging to one of them are more similar to each other than objects from different clusters. Such a classification with the help of software and computer system STATISTICA, can be performed simultaneously on a fairly large number of analytical features. In our case, clusters will be called geographically concentrated and interconnected by the level of recreational potential of forestry.

Among the agglomerative methods of cluster analysis, which are intended for grouping and combining objects of study, it is common to distinguish three most common types: hierarchical method (I) or the method of tree clustering; K-means Clustering Method (II) and two-step aggregation method (III).

I. Hierarchical clustering is used in the formation of clusters by determining the distances between objects and allows you to graphically visualize the results of the study in the form of a dendrogram. These distances can be determined in one-dimensional or multidimensional space. However, an important step in conducting a cluster analysis is to select the correct method for calculating the distances between the studied objects. The main ways to determine distances are: Euclidean distance, square of Euclidean distances, distance of city squares (Manhattan), Chebyshev distance, power distance.

II. K-means Clustering Method is the most common among non-hierarchical methods of cluster analysis. Unlike hierarchical methods, which did not require prior assumptions about the number of clusters, to be able to use this method it is necessary to have a hypothesis about the most probable number of clusters. K-means Clustering Method builds k clusters located at as large distances from each other as possible. Note that the K-means Clustering Method assumes that the number of clusters includes observations with the closest average value. The method is based on minimizing the sum of the distances squares between each observation and the center of its cluster, i.e. the function. In this case, the choice of the number of clusters is based on the research hypothesis. If it is not present, it is recommended to create 2 clusters, further 3, 4, 5, comparing the received results. The input will be $X_u = \{x_{1u}, x_{2u}, \ldots, x_{mu}\}$ – a set of unmarked data; $X_{kl} = \{x_{1l}, x_{2l}, \ldots, x_{pl}\}$ is a set of marked data in the class k, X_l=K_k=1X_{kl}. At the output, we want to obtain separated K sets $\{C_k\}$ K_k = 1 of X_u, which minimizes the objective function in *k*-means. Set parameters:

1. $t = 0$.
2. Initialization of cluster centers:

$$\mu_k = \frac{1}{|x_k^l|} \sum_{x \in x_k^l} x \tag{5}$$

3. Repeat until convergence:

provide cluster data:

For marked data: $x \in x_k^l$ provide x to the cluster C_k^{t+1}.

For unlabeled data: for $x_{iu} \in x_u$ provide to C_k^{t+1} a cluster obtained from $k = arg\ min_k||x_i^u - \mu_k^t||^2$.

4. Update centers:

$$\mu_k^{t+1} = \frac{1}{|C_k^l|} \sum_{x \in c_k^l} x \tag{6}$$

$t \leftarrow t+1$.

Another component of the algorithm is based on the discrepancy *KL*, which is a measure of the mismatch between the two probability distributions. Taking into account the *K*-dimensional probability vector of assignment of clusters *p* and *q* corresponding to points respectively x_p and x_q, the discrepancy *KL* between *p* and *q* is given by the formula:

$$KL\ (p\|q) = \sum_{i=1}^{K} p_i \log \frac{p_i}{q_i}, \tag{7}$$

where *K* is the number of clusters. In this approach, we use a symmetric variant of the discrepancy *KL*, because we are dealing only with the optimization of the loss function for *p* and *q* simultaneously:

$$L_{p,q} = KL(p\|q) + KL(q\|p) \tag{8}$$

Losses are obtained by first fixing p and calculating the discrepancy *q* with *p* and vice versa.

The described method makes it possible to automate the process of cluster data analysis, especially if the number of clusters is unknown from the beginning. For this purpose, the model of the neural network-based cluster data analysis system was described on the basis of *k*-means and *KL* discrepancy methods.

III. The two-way aggregation method is used in cases when you want to perform simultaneous clustering of objects (columns) and observations (rows) [11].

The key to the adequacy of the economic objects cluster analysis results is a reasonable choice of factors by which the grouping is carried out. Regarding the factor characteristics, we used a four-component system of indicators, which are shown in table 1.

The main purpose of cluster analysis is to break down the set of studied objects and features into homogeneous in the appropriate sense groups or clusters. This means that the task of classifying data and identifying the appropriate structure in it is solved. Methods of cluster analysis can be used in different cases, even when it comes to a simple grouping, and which all comes down to creating groups by the number of similarities.

The need for an objective division of different economic objects into groups exists constantly, because this classification allows you to find methods for effective management of these objects. Methods of cluster analysis allow to solve the following tasks: classification of objects taking into account the features that reflect the essence, nature of objects; verification of the assumptions about the presence of some structure in the studied set of objects, i.e. search for the existing structure; building new classifications for phenomena that have been little studied when it is necessary to establish the existence of relationships within the population and try to introduce a structure into it.

Cluster analysis has certain shortcomings and limitations. In particular, the composition and number of clusters depends on the selected breakdown criteria. When reducing the original data set to a more compact form, certain distortions may occur, and individual features of individual objects may be lost by replacing their characteristics with generalized values of cluster parameters.

When classifying objects, the possibility of the absence of any cluster values in the considered set is often ignored. In the cluster analysis it is considered that: 1) the chosen characteristics allow, in principle, a desirable division into clusters; 2) the units of measurement (scale) are chosen correctly.

The quality criterion of clustering to some extent reflects the following informal requirements: 1) within groups, objects must be closely related; 2) objects of different groups must be far from each other; 3) other things being equal, the distribution of objects by groups must be uniform. The key point in cluster analysis is the choice of metrics (or measures of proximity of objects), which crucially depends on the final version of the objects division into groups with a given algorithm of division.

The task of cluster analysis is to, based on the data of the set X, divide the set of objects G into m (m is an integer) of clusters (subsets) $G_1, G_2, \ldots, G_m$, so that each object G_j belongs to one and only one subset of the breakdown and that objects belonging to the same cluster are similar, while objects belonging to different clusters are heterogeneous. The solution to the problem of cluster analysis is the breakdowns that satisfy some criterion of optimality. This criterion may be some functionality that expresses the levels of different breakdowns desirability and groups, called the objective function. For further research, it was possible to use the methods of theories of complex systems and equipment made by tools used to examine the necessary systems of complexity, which were used in conventional [4; 3; 14; 13].

Let's perform cluster analysis according to the *K*-means Clustering method described above for each of the selected components (table 3).

Table 3. Substantiation of component's indicator.

<table>
<tr><th>Component</th><th>Indicator</th><th>Substantiation</th></tr>
<tr><td rowspan="5">Resource component</td><td>Area of recreational territories, km^2</td><td>var2</td></tr>
<tr><td>Number of recreational sites, quantity</td><td>var3</td></tr>
<tr><td>The level of attractiveness of natural and recreational resources</td><td>var4</td></tr>
<tr><td>Quality factor of forest vegetation</td><td>var5</td></tr>
<tr><td>Exoticism degree (contrast) of recreational territory</td><td>var6</td></tr>
<tr><td rowspan="5">Economic component</td><td>A proportion of total forestry costs on maintenance of recreational sites, %</td><td>var7</td></tr>
<tr><td>Efficiency factor of recreational forest management</td><td>var8</td></tr>
<tr><td>Wear coefficient of recreational fixed assets (FA)</td><td>var9</td></tr>
<tr><td>Volume of marginal costs for growing 1 ha of recreational forest</td><td>var10</td></tr>
<tr><td>Capacity of a single recreational load</td><td>var11</td></tr>
<tr><td rowspan="4">Social component</td><td>Proportion of recreant employees</td><td>var12</td></tr>
<tr><td>Recreational capacity</td><td>var13</td></tr>
<tr><td>Recreational load per 1 ha of forest</td><td>var14</td></tr>
<tr><td>The average stay of vacationers on the recreational territory, h</td><td>var15</td></tr>
<tr><td rowspan="5">Innovation and investment component</td><td>Cost amount on marketing activities of recreational territories</td><td>var16</td></tr>
<tr><td>Efficiency of innovation implementation of recreational forest management</td><td>var17</td></tr>
<tr><td>Amount of investments in recreational activity</td><td>var18</td></tr>
<tr><td>Proportion of foreign investments in recreational activities financing</td><td>var19</td></tr>
<tr><td>Quantity of grants (programs) won to finance recreational activities</td><td>var20</td></tr>
</table>

To begin with, we will standardize certain input data and summarize the results in table 4.

Table 4. The results of the standardization of the features of the recreational forest use assessment features.

	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13	Var14	Var15	Var16	Var17	Var18	Var19	Var20
1	-0,5669	-0,8472	-1,0818	-1,3149	-0,6656	-1,1348	-1,5557	-0,4606	-1,52518	-0,72738	-0,02347	0,20112	-0,5031	-0,25273	-2,4015	-0,15213	-1,24083	-1,68157	-0,60911
2	-0,4002	-0,8472	-0,3442	-0,5164	0,10241	-0,5857	0,92388	-0,8844	-0,82125	-0,42176	0,328617	1,039122	0,440211	0,288836	-1,01602	-0,96713	-0,05909	-0,62619	-1,08498
3	-0,2334	-0,4621	-1,0818	0,28214	0,87045	-0,5857	0,37286	-0,4606	-0,11732	-0,1467	-0,37556	1,877124	1,383519	1,371973	0,369462	1,477862	0,531783	0,429187	0,342624
4	0,43351	-0,4621	0,39338	1,08063	1,63849	0,23794	0,64837	-0,0367	0,586608	-0,17726	-0,02347	-1,47488	-0,5031	1,913541	1,062203	-0,15213	1,122653	0,956876	-0,60911
5	0,60024	-0,4621	1,13097	1,87913	-0,6656	0,51249	0,56572	0,3871	1,290537	-0,11614	0,328617	-0,63688	0,440211	-0,25273	1,754944	1,477862	1,713523	1,484565	0,342624
6	2,10083	-0,8472	1,13097	1,47988	-0,6656	-0,3112	0,51061	0,81093	0,938573	0,189486	1,736974	-0,97208	-0,5031	-0,7943	0,369462	-0,15213	-0,64996	0,851338	1,294358
7	-0,0667	-0,4621	-0,3442	0,28214	0,10241	-1,1348	0,81368	0,6414	-0,82125	0,800733	-0,72765	0,368721	1,383519	-1,33587	-0,32328	1,477862	-1,24083	0,956876	2,246092
8	0,43351	-0,077	0,39338	0,68139	-0,6656	0,23794	0,89633	0,81093	0,234643	1,411979	-0,37556	0,536321	0,440211	-0,7943	-0,18473	-0,15213	-0,35452	0,7458	1,294358
9	-1,2338	-0,077	-0,3442	-1,3149	-1,4337	-0,5857	-1,5557	0,25995	-1,17322	-1,64425	1,736974	1,039122	-0,5031	-1,33587	-1,01602	-0,15213	-0,64996	-1,68157	-0,60911
10	-0,5669	0,69319	-1,0818	-1,1552	-0,6656	-0,8603	0,92388	1,23476	1,290537	-1,33863	-0,72765	1,039122	0,440211	-0,7943	0,923655	-1,29313	1,713523	0,7458	0,342624
11	1,76737	1,07829	0,39338	-0,5164	0,10241	-0,5857	-1,0047	-1,7321	0,586608	-1,03301	-0,02347	0,03352	1,383519	-0,25273	-0,32328	-0,96713	0,531783	-1,25942	-1,08498
12	0,43351	3,00381	-1,0818	-0,6761	0,87045	0,23794	-1,2802	-2,1559	-0,53968	-0,42176	-0,37556	-1,47488	-1,44641	0,288836	-0,04618	1,477862	-1,24083	-0,41512	-0,89463
13	-0,4002	-0,4621	1,13097	0,60154	1,63849	0,51249	-1,2802	-0,0367	1,008966	0,800733	-1,07974	-0,63688	-1,44641	0,830405	-0,18473	-0,15213	-0,64996	-0,30958	-0,70428
14	-1,5673	0,30808	1,86857	-0,2768	-1,4337	1,61068	0,64837	0,3871	0,586608	1,717603	-1,78392	-0,13408	0,440211	1,371973	0,369462	-0,96713	-0,05909	-0,20404	-0,60911
15	-0,7336	-0,077	-1,0818	-0,5164	0,87045	2,43433	0,37286	1,23476	-1,52518	1,106356	1,384885	-0,80448	-1,44641	-0,25273	0,646558	-0,80413	0,531783	0,007036	0,342624

In the first stage of the cluster analysis, we find out whether the selected objects of study (Forestris) form “natural clusters”. To do this, use the method of hierarchical classification, in which we select the following characteristics: Amalgamation (joining) rule: Complete Linkage, Single Linkage and Ward’s method; Distance metric is: Euclidean distances (non-standardized). The obtained clustering results are shown in figures 3-6.

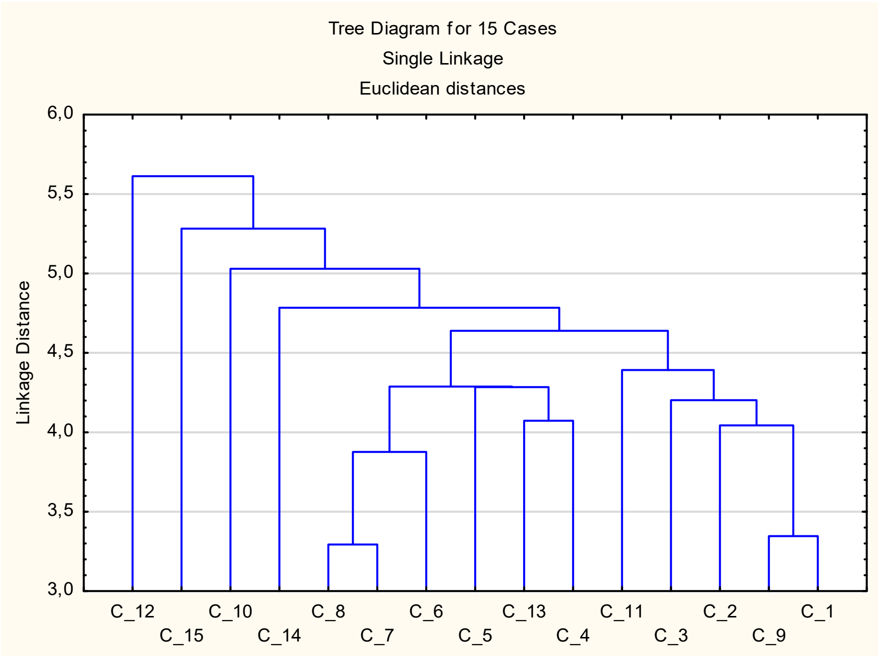

Fig. 3. Tree diagram for 15 forestries (Single Linkage).

Complete Linkage defines a relationship between clusters as the longest distance between two objects in different clusters (“the farthest neighbor”). Distance metric is Euclidean distances is a geometric distance in n-dimensional space and is calculated by the formula:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{9}$$

From the obtained calculations and the constructed dendrogram it is possible to draw conclusions that the investigated forestries form 5 natural clusters. Let’s test the above hypothesis by dividing the original data of K-means clustering into 5 clusters and check the significance of the difference between the obtained groups.

The best results in terms of meaningful interpretation were obtained by using an iterative method of cluster analysis, in particular the K-means clustering algorithm with division into three clusters. After the procedures performed by using the previously mentioned computer program, the results of clustering were obtained, which are shown in figure 6.

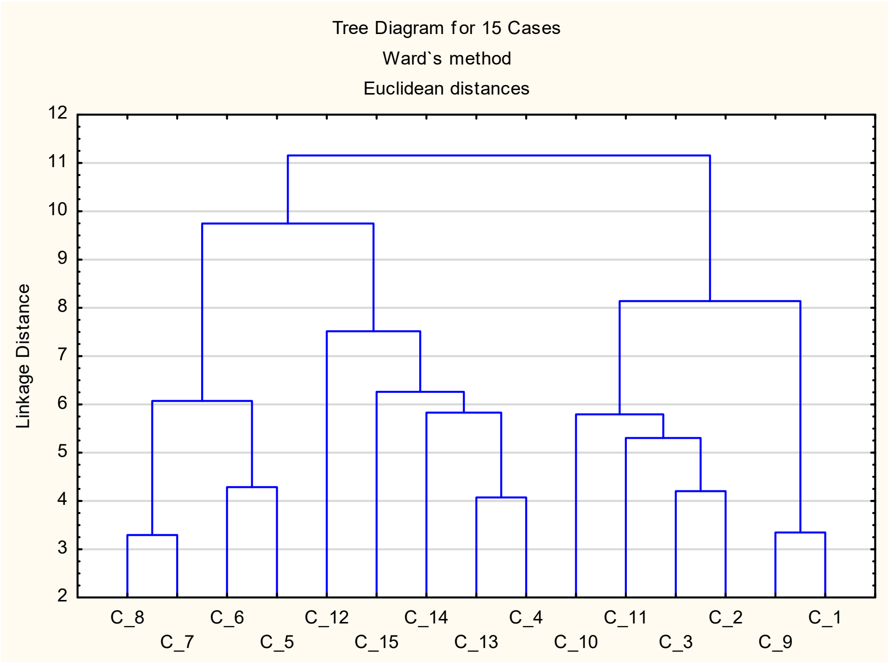

Fig. 4. Tree diagram for 15 forestries (Ward's method).

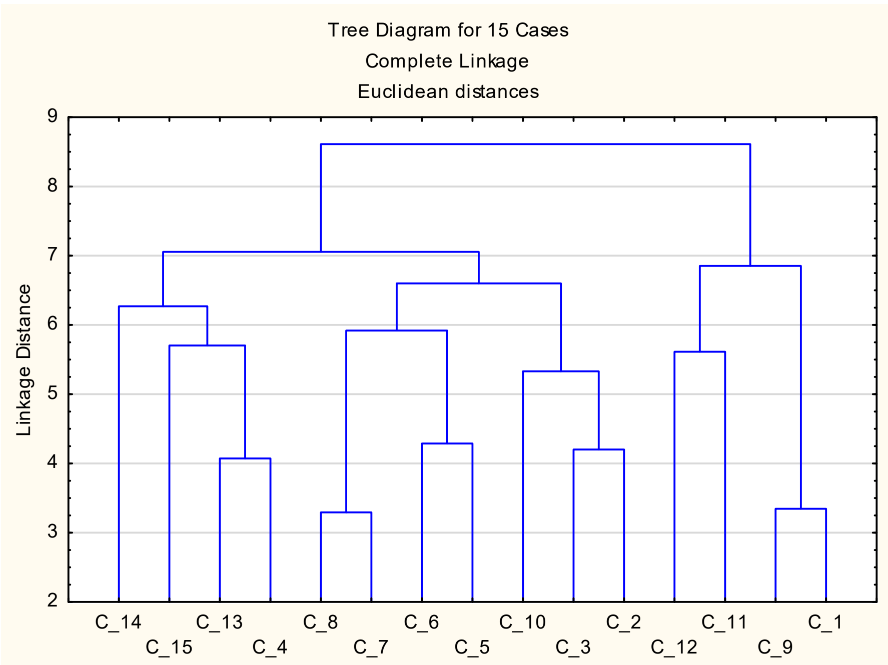

Fig. 5. Tree diagram for 15 forestries (Complete Linkage).

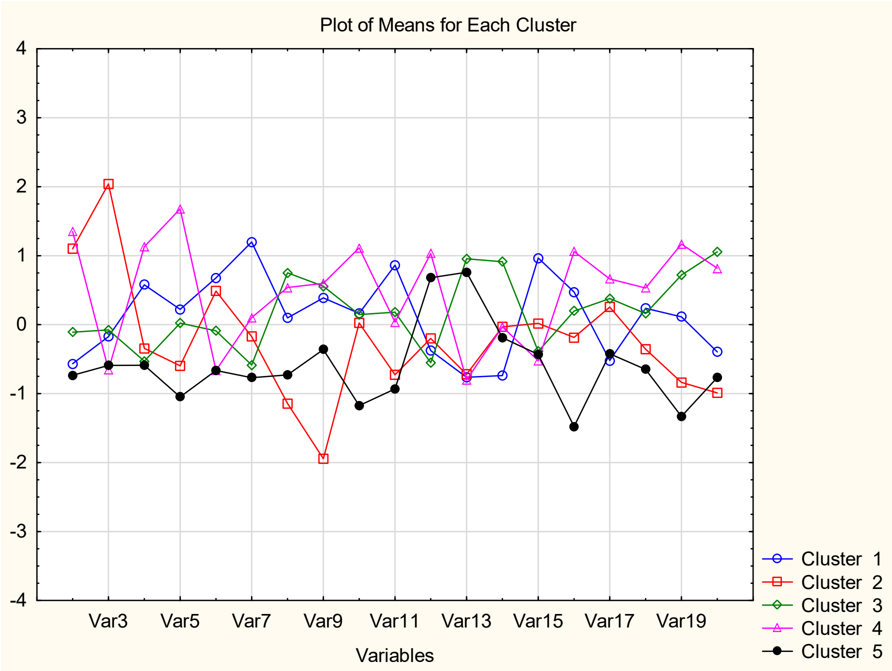

Fig. 6. Average level of normed values of indicators for the selected clusters.

To check the quality of the clustering, a variance analysis was performed, the results of which (table 5) indicate the relative quality of the clustering procedure: intergroup values of variances (Between SS) do not significantly exceed intragroup values (Within SS), except for 9 factors and the level of p- significance reaches the optimal value only for 9 characteristics.

Next, for qualitative clustering in the cluster analysis, we include the 9 most significant features of the previously performed analysis of variance. To implement clustering, we use the method of hierarchical classification, in which we select the following characteristics: Amalgamation (joining) rule: Complete Linkage, Single Linkage and Ward's method; Distance metric is Euclidean distances (non-standardized). The obtained clustering results are shown in the figures 7-9.

From the obtained calculations and the constructed dendrogram we can conclude that the studied forests form 4 natural clusters. Let's test the above hypothesis by dividing the original data of K-means clustering into 4 clusters and check the significance of the difference between the obtained groups.

The best results in terms of meaningful interpretation were obtained by using an iterative method of cluster analysis, in particular the K-means clustering algorithm with division into four clusters. After the procedures performed by using the previously mentioned computer program, the results of clustering are obtained, which are shown in figure 10.

Table 5. Analysis of variance.

Variable	Analysis of Variance (Апробація) Between SS	df	Within SS	df	F	signif. p
Var2	9,01688	4	4,98312	10	4,52371	0,024111
Var3	10,37888	4	3,62112	10	7,16553	0,005449
Var4	6,29275	4	7,70725	10	2,04118	0,164208
Var5	9,85135	4	4,14865	10	5,93648	0,010325
Var6	4,56180	4	9,43820	10	1,20833	0,366127
Var7	8,97487	4	5,02513	10	4,46500	0,025055
Var8	7,08283	4	6,91717	10	2,55987	0,103927
Var9	10,50677	4	3,49323	10	7,51937	0,004596
Var10	6,80881	4	7,19119	10	2,36707	0,122708
Var11	6,76535	4	7,23465	10	2,33783	0,125890
Var12	5,38430	4	8,61570	10	1,56235	0,258010
Var13	10,04167	4	3,95833	10	6,34211	0,008287
Var14	5,62076	4	8,37924	10	1,67699	0,230981
Var15	5,44553	4	8,55447	10	1,59143	0,250834
Var16	9,92733	4	4,07267	10	6,09387	0,009470
Var17	3,19534	4	10,80466	10	0,73934	0,586234
Var18	2,41334	4	11,58666	10	0,52072	0,722951
Var19	11,55609	4	2,44391	10	11,82130	0,000831
Var20	10,15565	4	3,84435	10	6,60426	0,007224

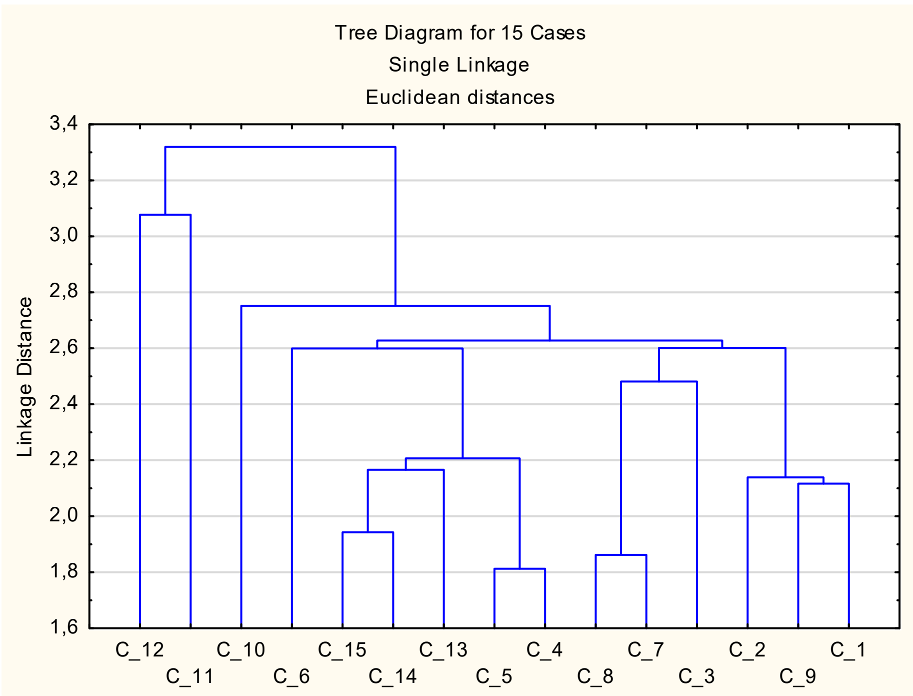

Fig. 7. Tree diagram for 15 forestries (Single Linkage).

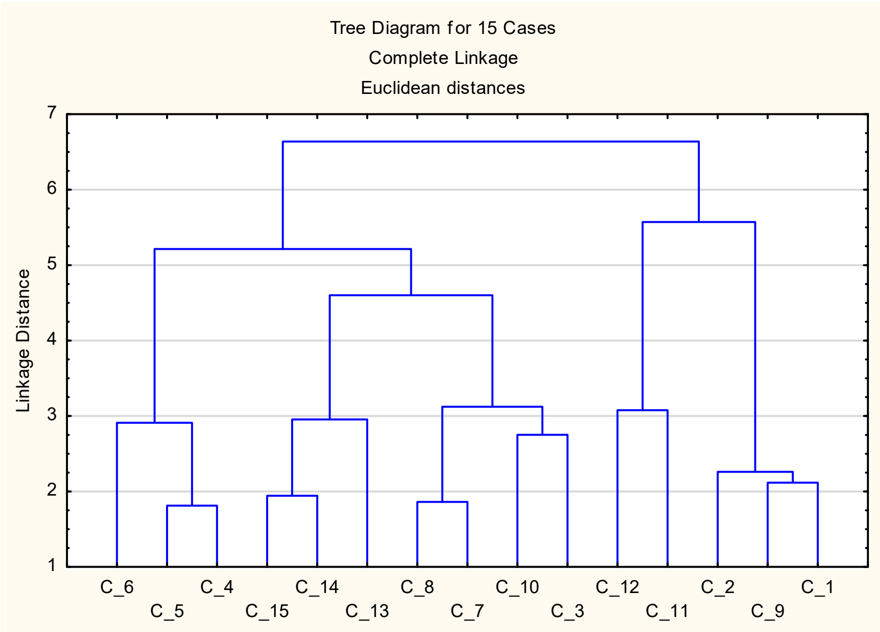

Fig. 8. Tree diagram for 15 forestries (Complete Linkage).

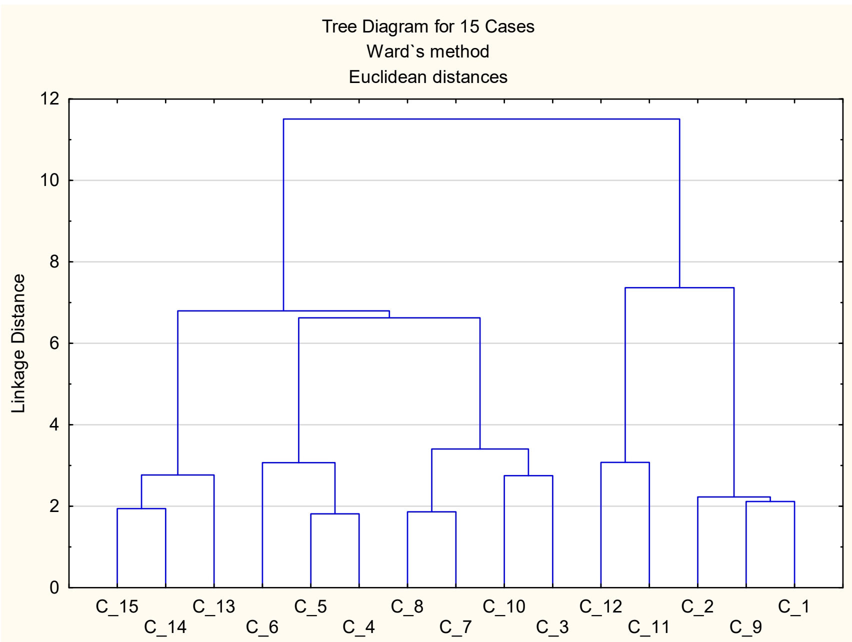

Fig. 9. Tree diagram for 15 forestries (Complete Linkage).

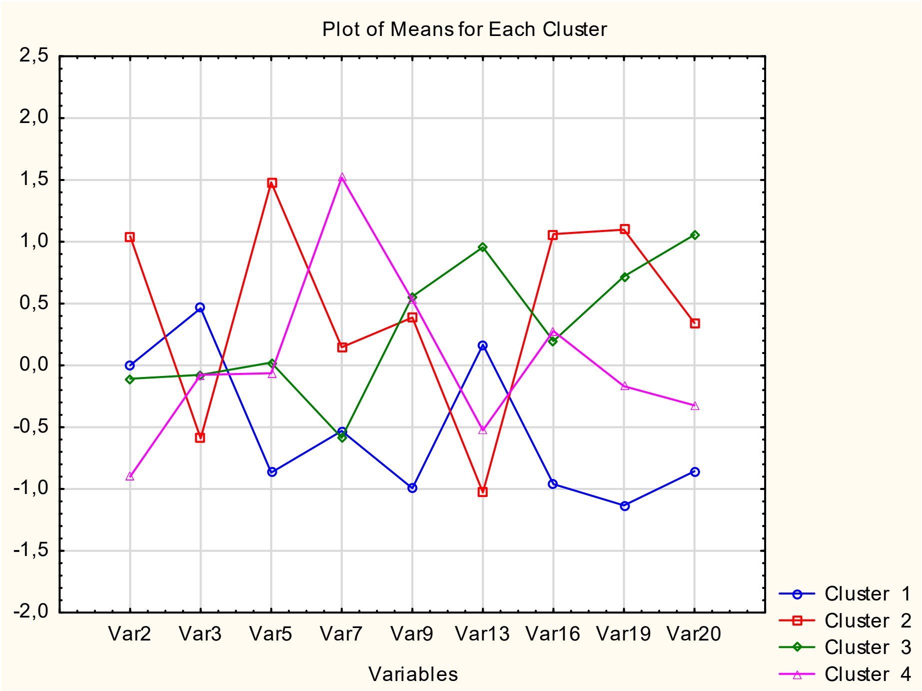

Fig. 10. Average level of normed values of indicators for the selected clusters.

The distance between the clusters, which are selected by K-means Clustering Method, was calculated by a simple Euclidean distance and are presented in table 6.

Table 6. Euclidean distances between clusters.

Cluster Number	Euclidean Distances between Clusters (Апробація) Distances below diagonal Squared distances above diagonal			
	No. 1	No. 2	No. 3	No. 4
No. 1	0,000000	2,445802	1,394819	1,277132
No. 2	1,563906	0,000000	1,068632	1,250228
No. 3	1,181025	1,033747	0,000000	1,106422
No. 4	1,130103	1,118136	1,051866	0,000000

To check the quality of the clustering, a dispersion analysis was performed, the results of which (table 7) indicate the high quality of the clustering procedure: intergroup values of variances (Between SS) significantly exceed intragroup values (Within SS), and the level of p-significance is much better than the normative (0.05).

Also, the contribution to the division of objects into groups is characterized by the values of Fisher’s criterion (F-criterion) and its significance level (p): the higher the values of the first and the smaller the values of the second, the better the clustering. For all parameters, without exception, the significance level approaches 0, which indicates the high statistical significance of the F-criterion. Depending on the levels of these indicators, forestry was grouped into four clusters (table 8).

Table 7. Euclidean distances between clusters.

Variable	Analysis of Variance (Апробація)					
	Between SS	df	Within SS	df	F	signif. p
Var2	9,79064	4	4,209359	10	5,81481	0,011049
Var3	10,46539	4	3,534605	10	7,40210	0,004860
Var5	10,59468	4	3,405316	10	7,77805	0,004073
Var7	10,19347	4	3,806533	10	6,69472	0,006896
Var9	10,47683	4	3,523172	10	7,43423	0,004786
Var13	10,41854	4	3,581461	10	7,27255	0,005172
Var16	10,38962	4	3,610375	10	7,19428	0,005373
Var19	12,47685	4	1,523151	10	20,47867	0,000083
Var20	8,85809	4	5,141908	10	4,30681	0,027826

Table 8. Forestry clusters.

Forestry group	Forestry
1 cluster	1, 2, 9, 11, 12
2 cluster	4, 5, 6
3 cluster	3, 7, 8, 10
4 cluster	13, 14,15

4 Results and conclusion

For the correct selection of clusters, a comparative analysis of several methods was performed: the arithmetic mean, hierarchical methods followed by dendrogram construction, K-means Clustering Method, which refers to reference methods in which the number of groups is specified by the user. The cluster analysis using different methods allows us to state that their combination helps to select reasonable groupings, visually illustrate the clustering procedure and rank the obtained clusters.

Thus, the results of the cluster analysis on 9 analytical grounds confirmed the hypothesis of separation of 4 clusters from 15 forestries. The first cluster is formed by five forestries 1, 2, 9, 11, 12, which are characterized by an average area of recreational territories, biggest number of recreational sites and recreational capacity, lowest quality factor of forest vegetation, proportion of total forestry costs on maintenance of recreational sites, wear coefficient of recreational fixed assets, cost amount on marketing activities of recreational territories, proportion of foreign investments in recreational activities financing, quantity of grants (programs) won to finance recreational activities. The second cluster is formed by three forestries 4, 5, 6. This cluster is characterized by the highest level of recreational territories, quality factor of forest vegetation, cost amount on marketing activities of recreational territories, proportion of foreign investments in recreational activities financing, an average level of recreational capacity and number of recreational sites, lowest level of proportion of total forestry costs on maintenance of recreational sites, wear coefficient of recreational fixed assets, quantity of grants (programs) won to finance recreational activities. The third cluster includes four forestries 3, 7, 8, 10, which have the following characteristics: the highest level of wear coefficient of recreational fixed assets, recreational capacity and quantity of grants (programs) won to finance recreational activities, average area of recreational territories, number of recreational sites and recreational capacity, quality factor of forest vegetation, cost amount on marketing activities of recreational territories and quantity of grants (programs) won to finance recreational activities, lowest proportion of total forestry costs on maintenance of recreational sites. The fourth cluster includes 3 forestries 13, 14, 15 and is characterized by the highest level of the proportion of total forestry costs on maintenance of recreational sites and wear coefficient of recreational fixed assets, lowest number of recreational sites and recreational capacity, quality factor of forest vegetation, recreational capacity, cost amount on marketing activities of recreational territories and quantity of grants (programs) won to finance recreational activities and quantity of grants (programs) won to finance recreational activities, the lowest of recreational sites.

For the proper selection of the clusters, a comparative analysis of several methods was performed: arithmetic mean, hierarchical methods followed by dendrogram construction, K-means method, which refers to the reference methods in which the number of groups is specified by the user. The cluster analysis, using different methods, allows us to state that their combination allows to select reasoned groupings, visually illustrate the clustering procedure and rank the obtained clusters.

The obtained results of clustering will help to develop separate development strategies for each isolated cluster, which will increase the efficiency of recreational areas management in the future. In addition, the results can be used to form an effective model for the development of recreational clusters.

References

1. Anderson, N., Ford, R.M., Bennett, L.T., Nitschke, C., Williams, K.J.H.: Core values underpin the attributes of forests that matter to people. Forestry: An International Journal of Forest Research **91**(5), 629–640 (2018). doi:10.1093/forestry/cpy022
2. Bell, S.: Forest recreation: New opportunities and challenges for forest managers. Rad. Sumar. inst. Izvanredni broj **10**, 155–160 (2005)
3. Bielinskyi, A., Soloviev, V., Semerikov, S., Solovieva, V.: Detecting stock crashes using Levy distribution. CEUR Workshop Proceedings **2422**, 420–433 (2019)
4. Derbentsev, V., Semerikov, S., Serdyuk, O., Solovieva, V., Soloviev, V.: Recurrence based entropies for sustainability indices. E3S Web of Conferences **166**, 13031 (2020). doi:10.1051/e3sconf/202016613031
5. Gerstenberg, T., Baumeister, C.F., Schraml, U., Plieninger, T.: Hot routes in urban forests: The impact of multiple landscape features on recreational use intensity. Landscape and Urban Planning **203**, 103888 (2020). doi:10.1016/j.landurbplan.2020.103888
6. Irland, L.C., Adams, D, Alig, R., Betz, C.J., Chen, C.-C., Hutchins, M., McCarl, B.A., Skog, K., Brent, L.: Assessing Socioeconomic Impacts of Climate Change on US Forests, Wood-Product Markets, and Forest Recreation: The effects of climate change on forests will

trigger market adaptations in forest management and in wood-products industries and may well have significant effects on forest-based outdoor recreation. BioScience **51**(9), 753–764 (2001). doi:10.1641/0006-3568(2001)051[0753:ASIOCC]2.0.CO;2
7. Juutinen, A., Kosenius, A., Ovaskainen, A.: Estimating the benefits of recreation-oriented management in state-owned commercial forests in Finland: A choice experiment. Journal of Forest Economics **20**(4), 396–412 (2014). doi:10.1016/j.jfe.2014.10.003
8. Kryzhanivskyi, Ye., Horal, L., Shyiko, V., Holubchak, O., Mykytiuk, N.: Economic and Mathematical Modelling for Evaluation of Potential Recreational Forest Utilization. Advances in Economics, Business and Management Research **99**, 173–178 (2019)
9. Lee, K.-C., Kang, K.-R.: Classification of Recreation Forests through Cluster Analysis. Journal of the Korean Institute of Landscape Architecture **37**(1), 9–17 (2009)
10. Meyer, M.A., Rathmann, J., Schulz, C.: Spatially-explicit mapping of forest benefits and analysis of motivations for everyday-life's visitors on forest pathways in urban and rural contexts. Landscape and Urban Planning **185**, 83–95 (2019). doi:10.1016/j.landurbplan.2019.01.007
11. Murphy, W.: Forest Recreation in a Commercial Environment. In: Small-scale forestry and rural development: The intersection of ecosystems, economics and society. Proceedings of IUFRO 3.08 Conference, Hosted By Galway-Mayo Institute of Technology, Galway, Ireland, 18-23 June 2006, pp. 347–356
12. Shin, H.-K., Shin, H.-C.: Market Segmentation on Recreational Forest Visitors by Cluster Analysis. The Journal of the Korea Contents Association **10**(3), 364–372 (2010). doi:10.5392/JKCA.2010.10.3.364
13. Soloviev, V., Bielinskyi, A., Solovieva, V.: Entropy Analysis of Crisis Phenomena for DJIA Index. CEUR Workshop Proceedings **2393**, 434–449 (2019)
14. Soloviev, V.N., Belinskiy, A.: Complex Systems Theory and Crashes of Cryptocurrency Market. Communications in Computer and Information Science **1007**, 276–297 (2019)