

THE USE OF CLUSTER ANALYSIS IN ASSESSING THE SUSTAINABILITY OF ORGANIC FARMS. PART I. METHODOLOGICAL CONSIDERATIONS

Maciej Sporysz^{a*}, Maria Szczuka^b, Sylwester Tabor^a, Krzysztof Molenda^a, Maciej Kuboń^a

^a Department of Production Engineering, Logistics and Applied Computer Science,
University of Agriculture in Krakow, Krakow, Poland
Maciej.Sporysz@urk.edu.pl, ORCID 0000-0003-4192-7235; Sylwester.Tabor@urk.edu.pl, ORCID
0000-0003-4614-0247; Krzysztof.Molenda@urk.edu.pl, ORCID 0000-0002-1019-7519;
Maciej.Kubon@urk.edu.pl, ORCID 0000-0003-4847-8743

^b Department of Machinery Management, Ergonomics and Production Processes,
University of Agriculture in Krakow, Poland, Maria.Szczuka@urk.edu.pl, ORCID 0000-0002-
7279-2049

*Corresponding author: e-mail: maciej.sporysz@urk.edu.pl

ARTICLE INFO

Article history:

Received: November 2019
Received in the revised form:
December 2019
Accepted: December 2019

Key words:

cluster analysis,
sustainable farm,
ecological farm

ABSTRACT

The modern agriculture model is a model based on the principles of sustainable development, i.e. protecting the environment and the cultural landscape on the one hand, and on the other, ensuring adequate income for residents. It is based on three orders: ecological, social and economic. This paper attempts to use cluster analysis to assess the economic and ecological sustainability of organic farms. It also indicates the factors that statistically influenced the assessment of a farm as sustainable, or not. The first part of the work is dedicated to the characteristics of the problem and the methodology of research and analysis. The second part contains the results of the research and discusses them.

Introduction

Since 2000, a European model of multifunctional agriculture was introduced to Poland, adopting management in accordance with the principles of sustainable development to protect the environment and biodiversity of rural areas. It also included caring for the cultural landscape and maintaining an adequate income for the rural residents, to ensure good living conditions and the development of the community. The development of agriculture has been based mainly on the concept of three orders that resulted in sustainable production:

- ecological (environmental) - assuming the preservation of soil fertility and reduction of environmental degradation by the introduction of mechanized work processes and an appropriate level of material and energy expenditures;
- social - increasing the mechanization of work processes, which leads to reduced effort and arduousness, thus ensuring greater comfort of work;

- economic - ensuring income parity for an agricultural producer - the level, scale and efficiency of production (Sawa, 2008).

Environmental issues are increasingly important in economic development since globalization and urbanization affect the lifestyle and behavior of consumers. Sustainable economic development requires promotion, an environmentally friendly policy at all levels of the economy, and transformation of production and consumption patterns for human activities to always support the system of sustainable development. The assessment of the economic system in this regard requires special attention due to the principles of efficiency, integrity and respect for nature. At the same time, the growing interest in environmental issues and the behavior of entities involved in economic processes are changing in response to new trends and requirements. Therefore, in business operations and in leveraging the environmental potential, we should strive to ensure the sustainable use of natural resources, and ultimately to improve and maintain a high level and quality of life (Ziętara, 1998). Hence, for every economic entity these days, maximizing profits is and will be inextricably linked to ecological considerations, as consumers will understand the imperative of environmental protection for present and future generations (Luchs et al., 2015).

Due to continuous improvement, sustainable and ecological development, introduces new methods of production and consumption, as well as new methods of business organization, in which environmental protection and quality of life have become the main pillar of every strategy (Aceleanu, 2016). Compliance with these principles may also be possible in a properly managed farm with at least two production lines (vegetal and animal). Animal production maintains a durable and efficient farming system and ensures the closest possible circulation of nutrients, with a balanced feed and fertilizer management. Stocking density in this type of farm should not be greater than its feed productivity, with the simultaneous production of manure necessary for fertilizing its agricultural land. It is also necessary to expand and modernize the technical infrastructure in rural areas and on the farms (Kukuła and Krasowicz, 2007). Striving for the improvement of the efficiency of direct technical expenditure should start by modifying and improving the construction of machines and tools, as well as choosing the production technology and rational selection of technical means so that they correspond to local conditions and the size of the enterprise (Popczyk, 2009). Only rationally selected and applied technical means will allow the proper implementation of production processes, scope and quality-wise, which may result in an increase in the obtained production results and a reduction in expenses. Therefore, in such conditions, productivity and labor productivity can be considered as the basic measures of economic sustainability (Tabor, 2006; 2008), however they are not always directly related to the measures of ecological sustainability. However, compared to other types of farms, it is organic farms that should be especially characterized by meeting both ecological and economic sustainability criteria. Therefore, conducting research aimed at understanding the conditions that impact the meeting of these criteria should be considered important, because only then will the opportunity to develop this management system be provided.

Purpose and scope of work

An agricultural holding must meet standards of the food producer market and operate according to strategies appropriate for production enterprises. Therefore, effective use of available resources is essential to allow delivering value that exceeds the costs incurred, as

only then can the farm develop and consolidate its position on the market. Otherwise, resources decrease, which results in reduced production capacity, and in the long run, in displacing food producers from the market. Unfortunately, organic food producers must also operate in such economic conditions. However, in their case, the rational use of resources, including human labor, must be accompanied by care for the preservation of environmental values and compliance with ecological sustainability standards. Therefore, for this type of production, it is characteristic to conduct processes based on the natural circulation and reproduction of nutrients, mainly minerals (N, P, K) and the organic matter. At the same time, due to economic conditions, it is necessary to generate enough agricultural to enable adequate parity of the family income and to secure the investment needs of the farm. Therefore, in this context, it is necessary to meet the economic and social sustainability standards, in which economic labor productivity remains the basic measure. It is the ratio of the value of the generated agricultural income to the used human labor resources.

With low material inputs from outside the farm, it can be assumed that the main factor determining the level of efficiency will be agricultural technology. Properly selected, it will reduce the use of human labor resources, and at the same time will be a source of expenditure that should dominate in the structure of production costs.

The main purpose of the work was to check the usefulness of cluster analysis (clustering) in assessing the sustainability of organic farms, i.e. whether the clusters obtained in the process of modeling group farms well in terms of sustainability. Apart from presenting the work methodology, the first part of the work aims to answer the question regarding the optimal number of clusters for the purpose of analysis. Too many clusters will cause excessive fragmentation of objects and prevent conclusions, and too small may not allow indicating statistically significant features of the farms.

Next, seeking the answer to the following questions should be considered as particularly important:

1. Which factors decide about ecological sustainability, and which about economic sustainability, and to what extent?
2. Do the dominant factors allow a proper description of the phenomenon depending on the type of the production, and in the case of economic sustainability also on technical equipment characteristic for this type?
3. Is it possible to distinguish between ranges or classes of balance, appropriate to the type of the production and technical equipment characteristic for this type, due to the dominance of certain factors and the similarity of the objects?
4. How strong is the impact of factors common to ecological and economic sustainability, including that, which determine technical equipment, within specific ranges and classes of sustainability?
5. Can these factors have a decisive influence on the choice of the type of the production and selection of technical equipment characteristic for this type?

The assessments and analyzes provide the opportunity to answer the questions posed, and the conclusions may have not only cognitive, but also utilitarian significance.

Methodology

The scope of research covered events from 50 ecological farms located in four provinces of southern Poland: Małopolskie, Podkarpackie, Śląskie and Świętokrzyskie. Basic research covered events from 2 years (2011, 2012) and were conducted as part of the National Centre for Research and Development (NCBiR) development project "Innovative impact of technology and IT-based management support on production efficiency of organic farms, NR12 016510". This project was implemented at the Faculty of Production and Power Engineering in Krakow. Reports submitted by farm owners to units supervising organic farms and economic event registers kept by farmers were used to collect the data for the purposes of the research project. As reports are submitted to several supervisory bodies, including certification bodies and agencies supervised by the Minister of Agriculture and Rural Development, significant divergence in entries have been identified in many. To verifying the reasons for these discrepancies, and due to the project's sustainability, the research in selected farms were supplemented until the end of 2015. All data was then verified and the results identified as affected by random phenomena were removed.

Income per full-time employee was adopted as an indicator of economic sustainability. This variable did not participate directly in the clustering, also known as object grouping, i.e. data segmentation and clustering. It is a method that is used in data mining techniques, an advanced data analysis tool used to group variables or objects into unknown groups. It includes dividing (usually multidimensional) data sets into groups (clusters) in such a way that the elements in the same group (cluster) are similar to each other, and at the same time as different from the elements in other groups as possible (Gatnar and Walesiak, 2004). The main purposes of clustering include the classification of data suggested by their natural grouping. In the case of multivariate data, visualization often doesn't allow it (Steliga, 2011; Morzy, 1999). Groups (clusters) are sets of such elements whose distance between any two elements in a cluster is smaller than the distance between any element in a cluster and any element from another cluster (Fig. 1).

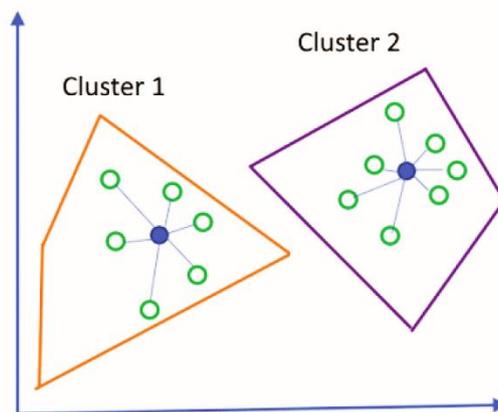


Figure 1. Graphic illustration of clusters based on two base components

In clustering, the data set is divided into groups using different criteria, which allows determining the similarities of objects in a given group (cluster). Such categorization causes maximum distinction of elements between groups, while maintaining their coherence (similarity) within the same group. Thanks to this method, maximum homogeneity in groups is achieved while maintaining the best heterogeneity between the groups (Marek, 1989).

Subject literature offers many variants of cluster analysis methods and proposals related to their classification (Sneath and Sokal, 1973). This paper uses the iterative optimization (non-hierarchical) method, which involves moving objects from one cluster to another. Its purpose is to find the best set of clusters according to a given criterion, but it requires prior specification of the number of clusters into which a given set is to be divided. The choice of this number is very important and has a great impact on the quality of the segmentation obtained. A large number of clusters will cause the determined clusters to be internally homogeneous, however, it is difficult to interpret and apply the results in practice. However, a small number of clusters yields lesser internal homogeneity of the cluster (Grabiński, 1992).

The "silhouette" factor was used to assess the quality of clustering internally. Objects with a high value of this factor are considered well allocated to clusters, objects with a low value can be outliers. This index works well with k - means clustering algorithm, and is also used to determine the optimal number of clusters.

Assuming that the set of objects was divided into k C_1, \dots, C_k clusters, and for each object i the average distance of this object from other objects belonging to the same cluster C_i was determined, according to the formula:

$$\alpha(i) = \frac{1}{\text{card}(C_i) - 1} \sum_{j=1(j \neq i)}^{\text{card}(C_i)} d(i, j)$$

where: $\text{card}(C_i)$ means the abundance of the i -th cluster, while $d(i, j)$ is the distance (according to the accepted metric) between objects i and j . The calculated value of $\alpha(i)$ can be interpreted as the quality level of the assignment of object i to the cluster - the lower the value, the better the quality).

Then, for each object i a cluster C_k mismatch level was specified with the equation:

$$\beta(i) = \frac{1}{\text{card}(C_k)} \sum_{j=1}^{\text{card}(C_k)} d(i, j)$$

determining the average distance of object i to all objects in the cluster C_k .

The level of general mismatch of object i to clusters not including this object was determined by the equation:

$$\gamma(i) = \min_{k \neq i} \beta(i)$$

High value of $\gamma(i)$ indicates a poor match of the i -th object to its neighboring clusters.

The "silhouette" value for the object i was described by the formula:

$$s(i) = \begin{cases} 0 & \text{gdy } \text{card}(C_i) = 1 \\ \frac{\gamma(i) - \alpha(i)}{\max\{\gamma(i), \alpha(i)\}} & \text{gdy } \text{card}(C_i) > 1 \end{cases}$$

This factor takes values from -1 to 1. The value close to 1 indicates a correct classification of the object into its cluster, while the value close to -1 indicates that classifying the object into the cluster adjacent to the assigned would be recommended. Zero means that the object is on the border of the cluster to which it has been assigned, and the adjacent cluster.

The value of the "silhouette" coefficient for the entire set of objects, and broken down into k clusters is defined by the formula

$$S(k) = \max_{j=1, \dots, k} \overline{s(j)}$$

where: $\overline{s(j)}$ is the arithmetic mean of the values $s(i)$ for all objects i belonging to the cluster j .

This value is a measure of the quality of the set's division into clusters and is used to determine the optimal number of clusters for a specific set of objects. Repeating the calculations of $S(k)$ for different values of k , the optimal division is the one for which $S(k)$ is the greatest.

Findings

To achieve the main purpose of the work, i.e. to check the usefulness of cluster analysis in assessing the sustainability of organic farms, it was necessary to determine the optimal number of clusters into which the collected research material will be divided. The analyzes were performed using the R analysis and visualization package.

Automatic division into k clusters was made using the silhouette factor method. For ecological sustainability (Fig. 2), the optimal number is 7. Due to the insufficient number of objects in some clusters, it was decided to reduce their number to 3 at the expense of slightly reducing the silhouette factor.

In the case of economic sustainability, the optimal number of clusters was determined as 8 and, as in the previous case, it was limited to 3 (Fig. 3). As a result, the number of objects in individual clusters was increased. Too much fragmentation would hinder generalization in the inference.

The number of clusters thus determined will be used in the second part of the work.

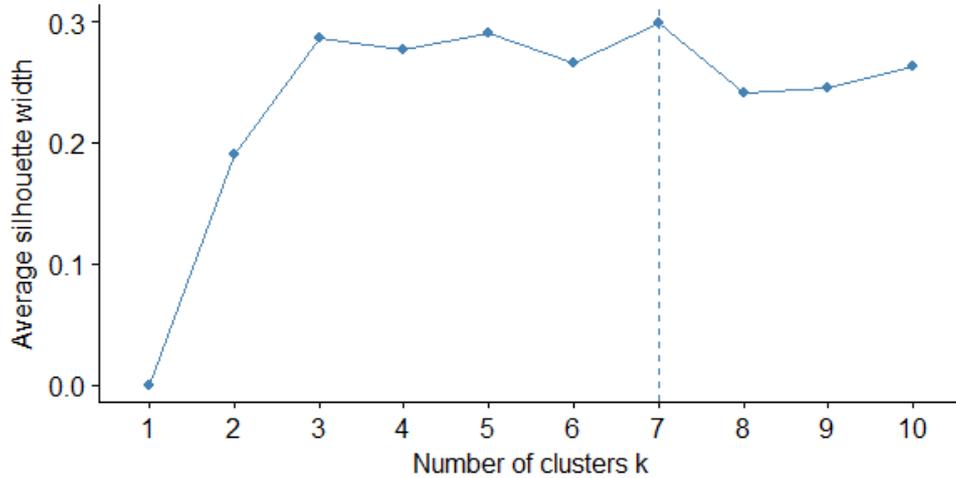


Figure 2. Optimal number of clusters for ecological sustainability analysis

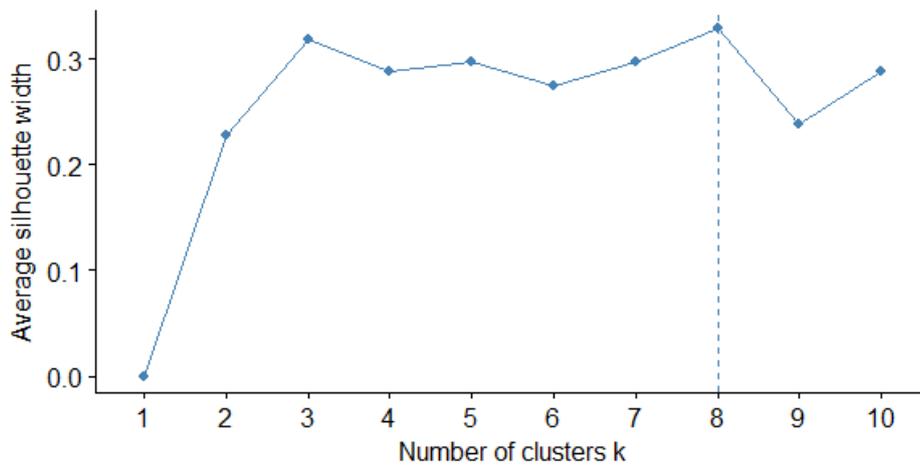


Figure 3. Optimal number of clusters for economic sustainability analysis

Conclusions

Both in the case of ecological and economic sustainability, the division into 3 clusters was considered the most useful. The analysis of silhouette values indicated the optimal division of 7 and 8, respectively, however, due to too large fragmentation in this case (some clusters would include a few farms) it was decided to divide into a smaller number of clusters. This division is characterized by a slightly lower silhouette value; instead, it significantly increases the number of objects in individual clusters.

The division into 3 clusters thus proposed will be used in the second part of the work, to assess the sustainability of organic farms.

References

- Aceleanu, M.I. (2016). Sustainability and Competitiveness of Romanian Farms through Organic Agriculture. *Sustainability* 8(3), 245.
- Gatnar, E., Walesiak, M. (2004). *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wydawnictwo Akademii Ekonomicznej, Wrocław.
- Grabiński, T. (1992). *Metody taksonometrii*, Cracow University of Economics.
- Kukuła, S., Krasowicz, S. (2007). Główne problemy i uwarunkowania zrównoważonego rozwoju rolnictwa w Polsce. *Problemy Inżynierii Rolniczej* 1, 5-15.
- Luchs, M.G., Phipps, M., Hill, T. (2015). Exploring consumer responsibility for sustainable consumption. *Journal of Marketing Management*, 31, 1449-1471.
- Marek, T. (1989). *Analiza skupień w badaniach empirycznych*. Państwowe Wydawnictwo Naukowe, Warsaw.
- Morzy, T. (1999). *Eksploracja danych problemy i rozwiązania*. 5th PLOUG Conference Zakopane.
- Popczyk, J. (2009). *Bezpieczeństwo elektroenergetyczne w społeczeństwie postprzemysłowym na przykładzie Polski*. Wydawnictwo Politechniki Śląskiej. Gliwice.
- Sawa, J. (2008). Próba oceny zrównoważenia procesów produkcji rolniczej. *Inżynieria Rolnicza* 2(100), 257-262.
- Sneath, P.H.A., Sokal, R.R. (1973). *Numerical Taxonomy*. W. H. Freeman and Co., San Francisco.
- Steliga, M. (2011). *Analiza skupień segmentów sygnału akustycznego w rozpoznawaniu mowy*. Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki Katedra Elektroniki, Krakow.
- Tabor, S. (2006). *Postęp techniczny a efektywność substytucji pracy żywej pracą uprzedmiotowioną w rolnictwie*. Rozprawa habilitacyjna. Inżynieria Rolnicza 10(85), Krakow. ISSN 1429-7264.
- Tabor, S. (2008). Produkcyjność i uzbrojenie techniczne ziemi a wydajność pracy w rolnictwie. *Postępy Nauk Rolniczych*, 4-5, 81-91.
- Ziętara, W. (1998). *Metodyczne aspekty oceny efektywności gospodarowania w rolnictwie*. Zeszyty naukowe SGGW EiOGŻ No 34.

WYKORZYSTANIE ANALIZY SKUPIEŃ W OCENIE ZRÓWNOWAŻENIA GOSPODARSTW EKOLOGICZNYCH. CZEŚĆ I. ASPEKTY METODYCZNE

Streszczenie. Nowoczesny model rolnictwa, to model oparty na zasadach zrównoważonego rozwoju, tzn. chroniący środowisko i krajobraz kulturowy z jednej strony, a jednocześnie zapewniający odpowiednie dochody mieszkańcom. Oparty jest na trzech ładach: ekologicznym, społecznym i ekonomicznym. W pracy podjęto próbę wykorzystania analizy skupień do oceny zrównoważenia ekonomicznego i ekologicznego gospodarstw ekologicznych. Wskazano czynniki, które statystycznie wpływały na decyzję o tym, które gospodarstwo należy uznać za zrównoważone, a które nie. Część pierwsza pracy poświęcona została charakterystyce problemu i metodyce badań i analiz. Część drugą to wyniki badań i ich omówienie.

Słowa kluczowe: analiza skupień, gospodarstwo zrównoważone, gospodarstwo ekologiczne