



## DATA CLASSIFICATION TO IMPROVE THE ALGORITHM FOR ASSESSING AN EXTERNAL PATCHWORK OF LAND OWNERSHIP

Grzegorz Oleniacz, Karol Noga, Adrian Moskal

### Summary

Research conducted for many years in Poland and around the world has demonstrated that defects in the spatial structure of agricultural land resulting from the common phenomenon of land fragmentation constitute one of important factors that contribute to the lack of rational land management. Reconstruction of the spatial structure of rural areas is essential for sustainable development of these areas. The process of land consolidation and exchange is a tool that can arrange space and lead to the desired structural changes, however, it has to be performed systematically and primarily in those areas where it is most urgently needed. With limited budget resources, it is reasonable to select objects for land consolidation in such a way as to obtain the best possible effect. In case of a selected group of neighbouring villages, a joint land consolidation procedure will enable better consolidation effects, inter alia by eliminating the external patchwork of land ownership. This article describes the modification and improvement in the methodology of designating areas with concentration of the external patchwork of land ownership. It is based on Czekanowski's diagram and was originally presented in 2017. The applied classification of numerical data enables the elimination of undesirable numerical effects of calculations and simplifies the interpretation of the final results.

### Keywords

rural areas • spatial analysis • land fragmentation • land consolidation • data classification • plot patchwork

### 1. Introduction

The existence of land fragmentation and a patchwork of land ownership in any area is economically harmful and unfavourable to landowners as it leads to lower labour productivity due to time lost for accessing plots dispersed over large areas and for travelling from plots to plots. In checkerboarded areas, the costs and time of transportation and thus the overall costs of agricultural production are higher than in more consolidated areas. In addition, land fragmentation makes it difficult for farmers to apply crop rotation and manage land in a rational way [Noga 2001, Stręk and Noga 2019]. With

regard to the administrative division, there is a distinction between an internal patchwork of land ownership (i.e., within the limits of the village) and an external one. The latter can occur between villages and also between communes, counties, voivodeships, and even between countries [Rabczuk 1968, Dudzińska 2012]. Research on the causes and effects of this phenomenon conducted for many years have led to a general conclusion that a patchwork pattern of land ownership results from the enforcement of inheritance law [King and Burton 1982, Tan et al. 2004, Niroula and Thapa 2005, Hung et al. 2007]. The problem associated with excessive fragmentation of land and patchwork of land ownership exists in many countries in Europe [Gonzalez et al. 2003, Di Falco et al. 2009, Demetriou 2013] and all over the world [Bentley 1987, Manjunatha et al. 2012, Guo et al. 2015]. Reconstruction of the spatial structure of areas affected by this phenomenon is necessary for the permanent and sustainable development of these areas. The process of land consolidation and exchange has been extensively used as a powerful tool for managing the spatial structure of rural areas in numerous countries. Land consolidation is a complex project involving agricultural landscape and natural resources, restoring the spatial structure of villages, and transferring land as well as the improvement in the road network and farmland irrigation systems [Chen et al. 2018, Zeng et al. 2018]. It has also an impact on proper determination of land ownership, the transformation of agricultural production methods and the transfer of rural labour [Shi et al. 2018, Zeng et al. 2018]. Land consolidation and exchange tasks require significant financial expenses, which with limited budget resources, can be conducted in few areas during a given funding period. Hence, in order to spend funds rationally, it seems essential to develop a methodology that leads to the selection of such areas where performing this type of work is urgent.

Conducting land consolidation and exchange tasks in one single village improves the spatial structure of this village, but it does not eliminate the phenomenon of an external patchwork of land ownership. Therefore, it has been proposed to apply an algorithm in search for areas with the external patchwork of land ownership and group them into clusters according to the method of cluster analysis and with the use of Czekanowski's diagrams [Leń et al. 2017, Oleniacz 2018, Oleniacz and Leń 2018]. In general, this method can be classified as non-hierarchical, but it has such an advantage that the number of clusters does not need to be predefined. In addition, it provides a clearer graphical presentation of results than dendrograms in hierarchical methods. The clusters determined in this way can be treated as whole objects subjected to land consolidation and exchange, for which the process will bring the greatest benefits in terms of the elimination of the external patchwork of land ownership.

The above-mentioned algorithm is based on four matrices (arrays) of mutual dependencies between villages. Three of the arrays contain statistical data referring to the features that are the most significant when identifying the existence of patchwork of land ownership:

- $A = [a_{ij}]$  – the percentage of non-resident owners in the total number of plot private owners,
- $B = [b_{ij}]$  – the percentage of non-resident owners' plots in the total number of private landowners' plots within a cadastral unit (village),

- $C = [c_{ij}]$  – the percentage of the area of non-resident owners' plots in the total area of a cadastral unit (village).

The application of these percentages is a way to create a uniform comparative scale for the whole studied area. The fourth matrix ( $D = [d_{ij}]$ ) is a first-order spatial weight matrix. Individual elements of this matrix are 1, if the villages have a common border, or 0, if they do not border each other. In order to include all the features in the analysis, all the matrices are combined into one ratio matrix  $W = [w_{ij}]$  according to the formula:

$$w_{ij} = (a_{ij}) \cdot (b_{ij}) \cdot (c_{ij}) \cdot (d_{ij}) \quad (1)$$

The  $W$  matrix is then transformed into a symmetrical dependency matrix  $Z$  as shown below:

$$Z = W + W^T \quad (2)$$

The  $Z$  matrix represents the mutual dependencies between individual pairs of villages with regard to the examined features and it replaces the classic form of a distance matrix in the proposed algorithm. According to the assumption of the Czekanowski's diagram method [Czekanowski 1909, Czekanowski 1910, Sołtysiak and Jaskulski 1998], the  $Z$  matrix is ordered in line with the criterion proposed by Sołtysiak [Sołtysiak 1997] while the graphical visualization of the results allows for the identification of clusters.

The above-described algorithm is an enhanced and multivariate version of an external patchwork of land ownership cluster analysis first shown by prof. Noga [Noga 1977].

## 2. Research and method – modification of the algorithm

Originally, the values included in the matrices from  $A$  to  $D$  were raw data obtained from a digital database of the land and building register or values calculated directly from these data. However, the first attempts to apply the presented algorithm in its original form revealed some numerical problems, which resulted in favouring some pairs of villages and contradicted the intentions of the authors and the rational interest of conducting this type of analysis. For example, in specific cases, the distance between a pair of villages calculated on the basis of centroids was smaller for villages without a common border than for the same villages and the villages adjacent to them. Hence, in the current version of the algorithm, the  $D$  matrix is a representation of the mutual position of villages in relation to their immediate neighbourhood (according to the common border criterion), and not the neighbourhood matrix according to the reciprocal distance criterion [Oleniacz 2018].

Another noticed problem is favouring pairs of villages which demonstrate an asymmetrical influence of the considered factors on each other. A pair of villages where non-residents own e.g., 9% of plots in the second village while at the same time the opposite impact is significantly lower, will be weighed higher with respect to the cluster formation than a pair of villages with a more symmetrical mutual influence (e.g., 4% in both directions). There were attempts to solve this problem by selecting additional

weights for the data contained in the dependency matrices [Leń 2018] or replacing the raw data with other representative values [Oleniacz and Leń 2018]. These proposals, to some extent, eliminate the above-described adverse effect of favouring asymmetric mutual relations in pairs of the analysed villages. They also simplify data in the tables, which has a positive effect on the efficiency of the table  $Z$  ordering algorithm. However, the new, arbitrarily adopted values make a change in the scale of mutual impact that individual pairs of villages have on each other.

In a classic task of statistical cluster analysis, the input data are used to create a matrix of mutual distances between objects according to the adopted distance meter. The smaller the distance, the more similar the objects are to each other and vice versa. Therefore, this matrix is sometimes called the matrix of similarity. In this case, due to the nature of the problem under consideration, the original values obtained from the land and building registry databases refer directly to the relationship between the studied objects and express their mutual dependence on each other in terms of the analysed feature. When conducting statistical analyses on large sets of data, very often the first step is to simplify the representation of the complete set by grouping the data into class intervals [Holcomb 1997, Sobczyk 2010].

In order to reflect the experience gained so far, it is proposed to divide the raw data in individual matrices into class intervals and replace all values in a given interval with one representative value in the form of a median from a given interval. The median, as the middle value from the interval, will allow for a more reliable representation of the scale of mutual dependencies of individual village pairs from neighbouring class intervals. The median has also the property of being resistant to outliers in the dataset and to skewness of the distribution [Bissel 1994, Bakker and Gravemeijer 2006], which should mitigate the impact of extreme values on the final result. The number of class intervals will be determined according to the Sturges' formula [Sturges 1926]:

$$K = 1 + \log_2(N) \quad (3)$$

where:

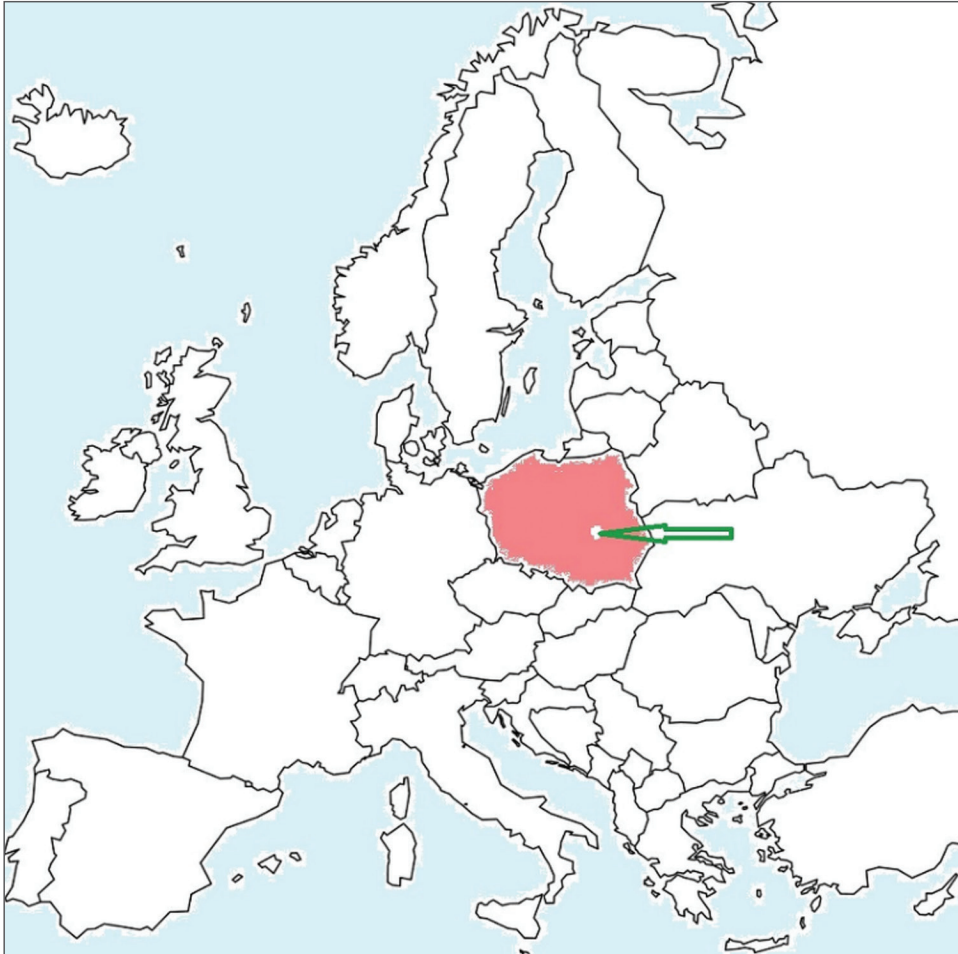
- $K$  – a number of class intervals to determine,
- $N$  – the size of the dataset.

This variant of specifying the number of intervals is widely used around the world and it is the default choice in many statistical analysis software packages [Scott 1992].

### 3. Results and discussion

The test of the improved algorithm was conducted on the data concerning the area of Sławno commune located in central Poland (Fig. 1).

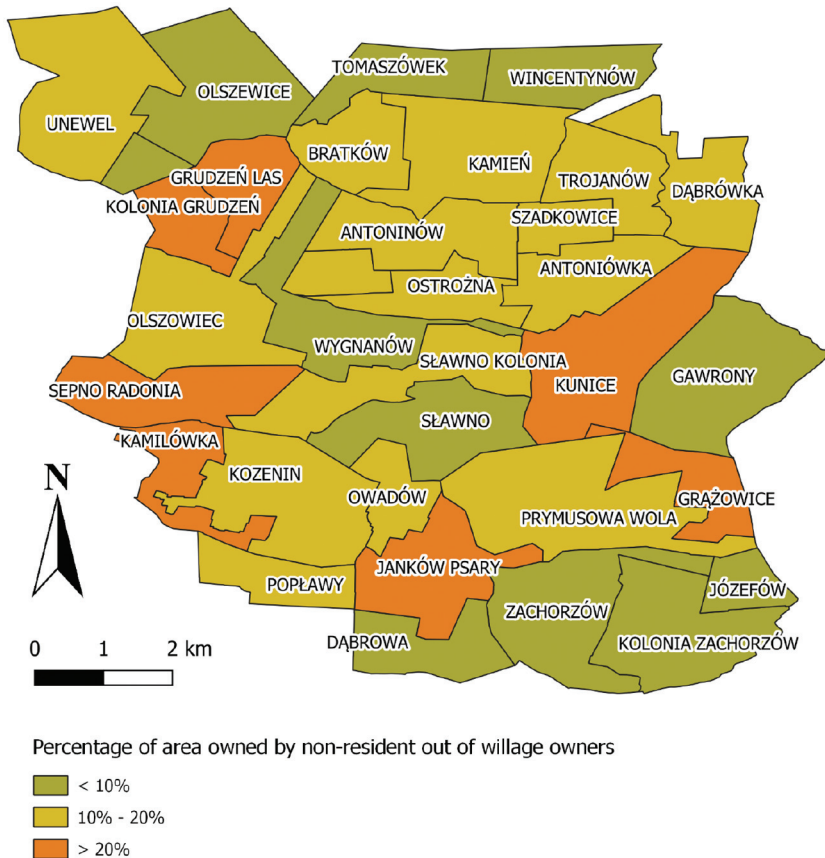
It is the area of central, eastern, and south-eastern Poland that is characterized by the most unfavourable spatial structure of rural areas [Leń 2017, Mika and Leń 2017, Mika et al. 2017]. It also concerns the analysed test area, where even more than 30% of the total number of plots belong to out-of-village non-resident owners (Fig. 2).



Source: Authors' own study

**Fig. 1.** Approximate location of the test area on the map of Europe

The problem of strong unilateral impact in individual pairs of villages is evident in many cases. For instance, the area of land that residents of Kozenin own in the village of Kamilówka is over 28 ha, which corresponds to more than 10.2% of the total area of the village, whereas residents of Kamilówka own only 4 ha in Kozenin (about 0.7% of the total area of the village). Figure 3 shows those pairs of villages where the described phenomenon is the most visible and, at the same time, where the percentage of the area owned by out-of-village non-resident owners is the highest. Thus, these pairs of villages will have the greatest impact on the formation of clusters.

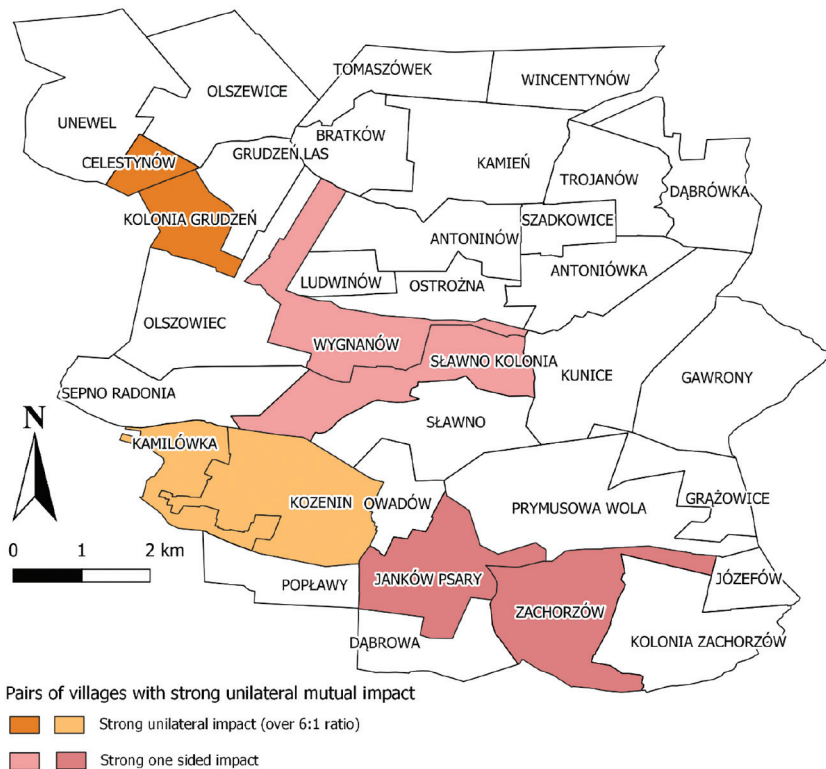


Source: Authors' own study

**Fig. 2.** Percentage distribution of the area of plots owned by out-of-village non-resident owners in individual villages of Sławno commune (it concerns only non-resident owners living in the same commune)

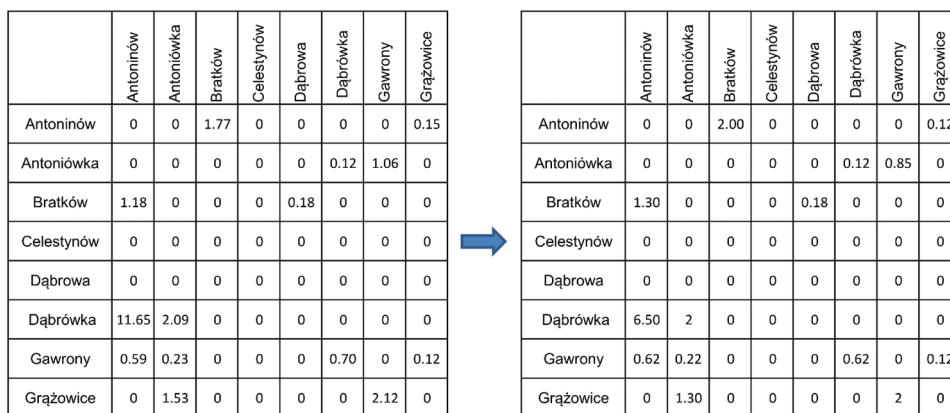
In accordance with the adopted assumption, all data contained in the matrices from *A* to *C* were classified into appropriate class intervals. In the case of Sławno commune, 9 intervals with an equal amount of data were determined for individual matrices. The median value was calculated for each interval and replaced with all the data from that interval (Fig. 4). These matrices were then transformed according to the formula (1) and (2) into a symmetric *Z* matrix which was ordered afterwards.

The *Z* matrix is ordered by rearranging its rows and columns in such a way that the pairs of villages with the greatest mutual influence are as close to the diagonal as possible forming clusters. In Czekanowski's method, such an ordered matrix is visually presented in the form of a diagram, in which the numerical values correspond to symbols. In this case, it is also divided into intervals and the symbol is assigned to a given interval.



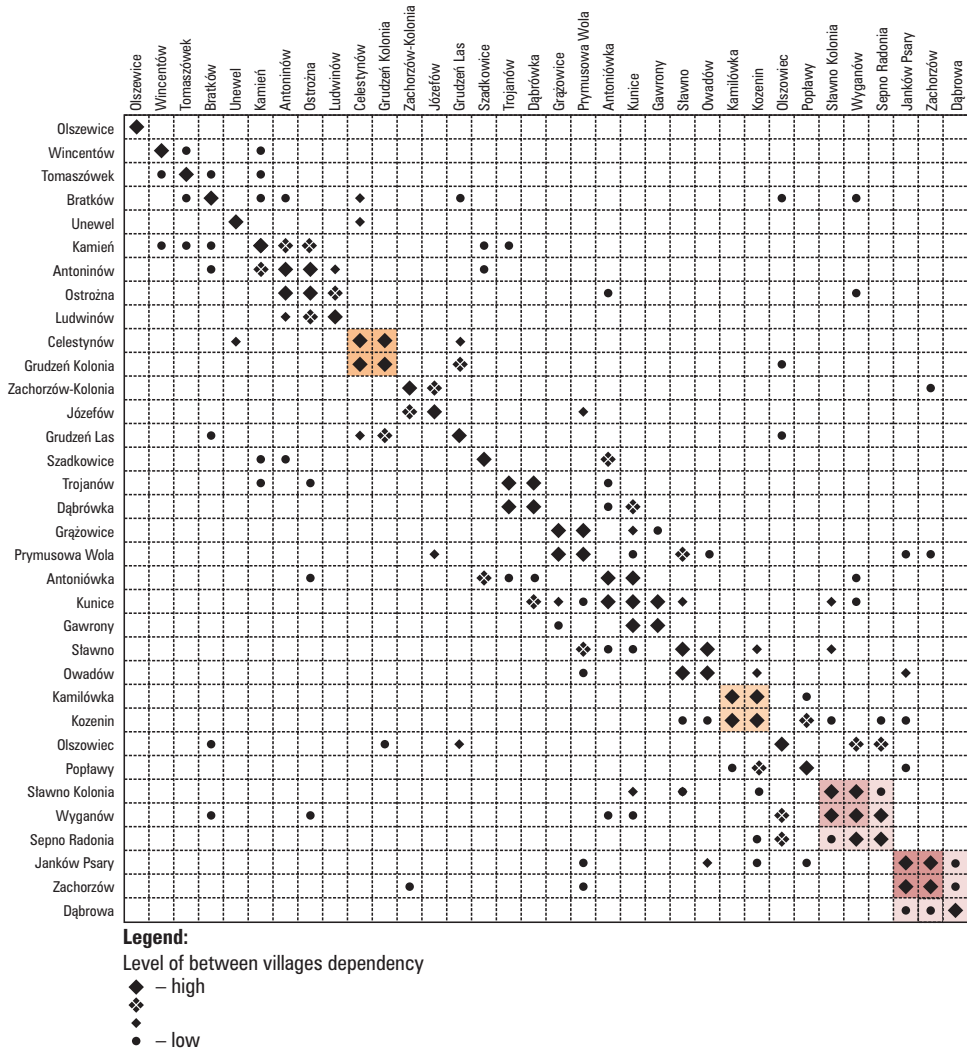
Source: Authors' own study

Fig. 3. Pairs of villages with strong unilateral mutual impact regarding the percentage of out-of-village non-resident owners' areas in the total area of private farms



Source: Authors' own study

Fig. 4. An example of replacing cell values for a fragment of the A matrix



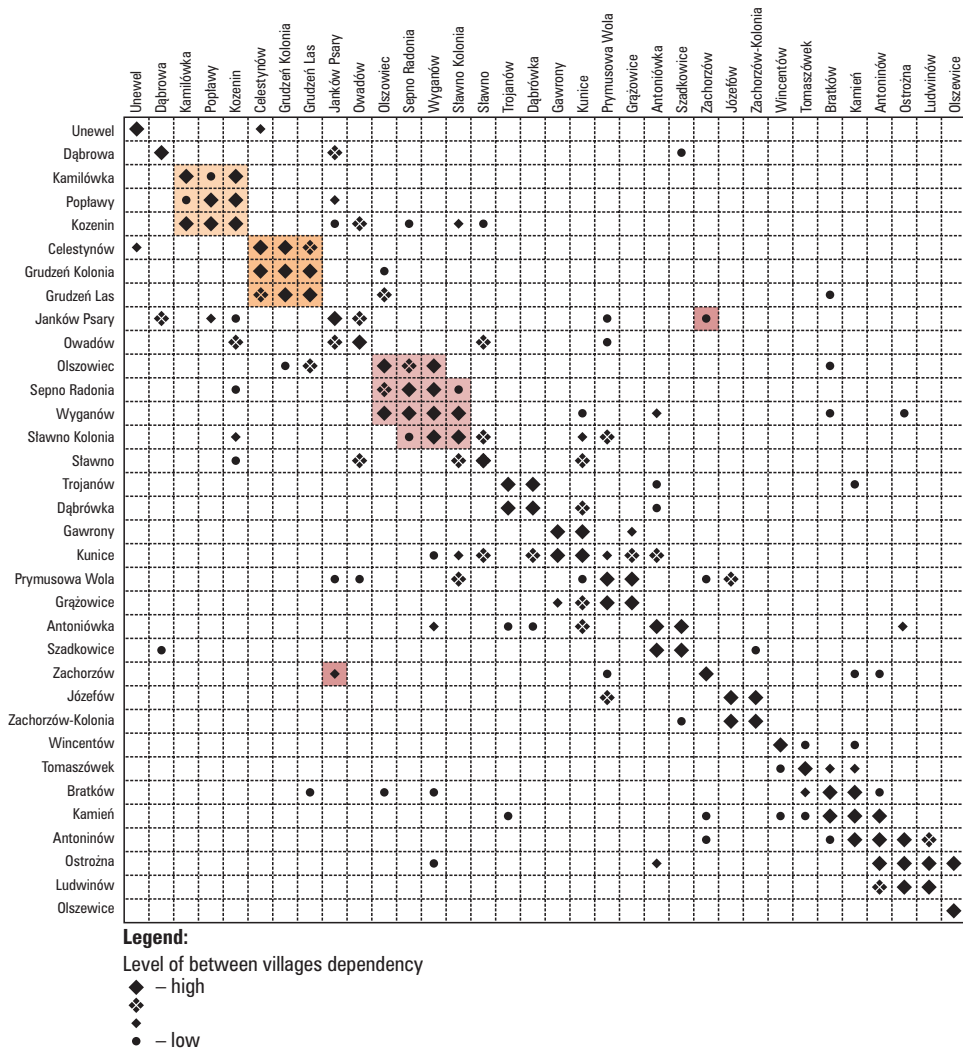
Source: Authors' own study

Fig. 5. Czekanowski's diagram for the original version of the Z matrix

For the purposes of comparison, Figure 5 presents the result of ordering the Z matrix which was created from the original versions of the A, B and C matrices. Significant discrepancies between the values of individual matrix cells cause a large dispersion in the values of the mutual dependence indicators of particular village pairs included in the ordered matrix. This results in the formation of only clusters composed of one pair of villages with the greatest mutual dependence. Moreover, the pairs of villages marked in orange in Figure 3 created such clusters. For the pairs of 'red' villages in Figure 3, slightly enlarged clusters can be distinguished. The whole diagram clearly



shows a cluster of closely related villages of Antoninów – Ostorożna – Ludwinów and Kamień as well. There is also a concentration of villages in the group of Prymusowa Wola, Grążowice, Antoniówka, Kunice and Gawrony.



Source: Authors' own study

Fig. 6. Czekanowski's diagram for the Z matrix created from the modified A, B and C matrices

The ordering of the Z matrix created from the modified A, B and C matrices provided different effects (Fig. 6). This was due to a smaller dispersion of the values of the mutual dependence indicators for particular village pairs in the Z matrix. The 'orange' clusters and a 'red' one expanded. However, the second 'red' cluster was not formed. Following in-depth analysis, it turns out that few of the inhabitants of Zachorzów village own as much as 19%

of the area of private plots in the village of Janków Psary. It is the highest percentage among all villages (three times higher than the median for this interval) and it is this value from the *C* matrix that eventually generates a high index for this pair of villages in the *Z* matrix. After the matrix modification, the final index for a pair of villages of Janków Psary and Zachorzów has a much smaller influence on the formation of clusters. The remaining villages formed the same or slightly regrouped clusters.

#### 4. Conclusions

Modifications of cell values in the *A*, *B* and *C* matrices resulted in a smaller dispersion of the mutual dependence indicators of villages in the *Z* matrix, which in turn allowed for a clearer cluster formation. However, the effects will depend on the distribution of data for the analysed objects. If there is a large number of outliers, the changes caused by the modification will be less significant.

The described algorithm is simple to use, and the graphical final form is easy for the result interpretation. It can be easily extended with additional dependency matrices created according to the selected features.

In practice, the presented methodology of calculations can be applied wherever there is access to reliable information from cadastral databases that are used to obtain a complete picture of the spatial and ownership structure of the analysed areas. If the data are digital, the algorithm can be fully automated.

Searching for areas with an excessive concentration of external patchwork of land ownership will allow for better consolidation effects through its liquidation and thus for more effective spending of funds allocated to consolidation.

#### References

- Bakker A., Gravemeijer K.P.E. 2006. An Historical Phenomenology of Mean and Median. *Educ. Stud. Math.*, 62, 149–168. <https://doi.org/10.1007/s10649-006-7099-8>
- Bentley J.W. 1987. Economic and ecological approaches to land fragmentation: In defence of a much-maligned phenomenon. *Annual Review of Anthropology*, 16, 31–67. <https://doi.org/10.1146/annurev.an.16.100187.000335>
- Bissell D. 1994. *Statistical Methods for Spc and Tqm*. 1<sup>st</sup> ed. Chapman and Hall/CRC.
- Chen F., Yu M., Zhu F., Shen C., Zhang S., Yang Y. 2018. Rethinking Rural Transformation Caused by Comprehensive Land Consolidation: Insight from Program of Whole Village Restructuring in Jiangsu Province, China. *Sustainability*, 10, 2029. <https://doi.org/10.3390/su10062029>
- Czekanowski J. 1909. Zur Differentialdiagnose der Neandertalgruppe, *Korespondentblatt der Deutschen Gesellschaft für Anthropologie. Ethnologie und Urgeschichte*, XL, 6/7, 44–47.
- Czekanowski J. 1910. Verwandtschaftsbeziehungen der Zentralafrikanischen Pygmäen, *Korespondentblatt der Deutschen Gesellschaft für Anthropologie. Ethnologie und Urgeschichte*, XLI, 9/12.
- Demetriou D. 2013. *The Development of an Integrated Planning and Decision Support System (IPDSS) for Land Consolidation*. Springer, Cham. <https://doi.org/10.1007/978-3-319-02347-2>
- Di Falco S., Penov I., Aleksiev A., van Rensburg T. 2009. Agrobiodiversity, farm profits and land fragmentation: Evidence from Bulgaria. *Land Use Policy*, 27, 763–771. <https://doi.org/10.1016/j.landusepol.2009.10.007>

- Dudzińska M. 2012. Szachownica gruntów rolnych jako czynnik kształtujący przestrzeń wiejską [Patchwork of fields as a factor which affects rural space]. *Infrastruktura i Ekologia Terenów Wiejskich*, 2(3), 45–56.
- Gonzalez X.P., Alvarez C.J., Crecente R. 2003. Evaluation of land distributions with joint regard to plot size and shape. *Agricultural Systems*, 82, 31–43. <https://doi.org/10.1016/j.agry.2003.10.009>
- Guo B., Jin X., Yang X., Guan X., Lin Y., Zhou Y. 2015. Determining the effects of land consolidation on the multifunctionality of the cropland production system in China using a SPA-fuzzy assessment model. *Eur. J. Agron.*, 63, 12–26.
- Holcomb Z.C. 1997. *Fundamentals of Descriptive Statistics*. Routledge, Chapman Hall.
- Hung P.V., MacAulay G., Marsh S.P. 2007. The economics of land fragmentation in the north of Vietnam. *The Australian Journal of Agricultural and Resource Economics*, 51, 195–211. <https://doi.org/10.1111/j.1467-8489.2007.00378.x>
- King R., Burton S. 1982. Land fragmentation: Notes on a fundamental rural spatial problem. *Progress in Human Geography*, 6(4), 475–494. <https://doi.org/10.1177/030913258200600401>
- Leń P. 2017. Patchwork of land ownership as barrier to development of rural areas. *Geomatics and Environmental Engineering*, 11, 3, 85–96. <https://dx.doi.org/10.7494/geom.2017.11.3.85>
- Leń P., Oleniacz G., Skrzypczak I., Mika M. 2017. Methodology for assessing the size and liquidation of the outer patchwork of land. *World Multidisciplinary Earth Symposium – WMESS, Prague, IOP Conf. Series: Earth and Environmental Science*, 95, 032020. [dx.doi.org/10.1088/1755-1315/95/3/032020](https://doi.org/10.1088/1755-1315/95/3/032020)
- Leń P. 2018. Methodology for the evaluation and elimination of the external patchwork of agricultural land for the purpose of increasing the effectiveness of consolidation and exchange process. *Acta Sci. Pol., Administratio Locorum*, 17(1), 41–50. <https://doi.org/10.31648/as-pal.471>
- Manjunatha A.V., Anik A.R., Speelman S., Nuppenau E.A. 2012. Impact of land fragmentation, farm size, land ownership and crop diversity on profit and efficiency of irrigated farms in India. *Land Use Policy*, 31, 397–405. <https://doi.org/10.1016/j.landusepol.2012.08.005>
- Mika M., Janus J., Tazsakowski J., Leń P. 2017. The role of the comprehensive land consolidation and exchange works in the process of modernization of real estate cadastre in Poland. *Infrastructure and Ecology of Rural Areas*, 1/2, 243–255. [dx.doi.org/10.14597/infraeco.2017.1.2.018](https://doi.org/10.14597/infraeco.2017.1.2.018)
- Mika M., Leń P. 2017. The size of land by non-residents and the possibility of liquidation on the example of the village municipality Sławno, Opoczno county, state Łódź. *Infrastructure and Ecology of Rural Areas*, 1/1, 171–182. [dx.doi.org/10.14597/infraeco.2017.1.1.013](https://doi.org/10.14597/infraeco.2017.1.1.013)
- Niroula G.S., Thapa G.B. 2005. Impacts and causes of land fragmentation, and lesson learned from land consolidation in South Asia. *Land Use Policy*, 22(4), 358–372. <https://doi.org/10.1016/j.landusepol.2004.10.001>
- Noga K. 1977. Analysis of the Intervillage Plot Patchwork on the Example of Villages Situated in the Upstream Catchment of the River Soła. *Scien. Sess.*, 7, 7. *Scien. Booklets*, 133. University of Agriculture in Krakow, Poland, 154–170.
- Noga K. 2001. *Metodyka programowania i realizacji prac scalenia i wymiany gruntów w ujęciu kompleksowym [Methodology of programming and implementation of land consolidation and exchange activities: A comprehensive perspective]*. Szkoła wiedzy o terenie. Kraków, Poland, 88.
- Oleniacz G. 2018. Spatial weights matrix in clustering of the external patchwork of agricultural land. *Acta Scientiarum, Polonorum Administratio Locorum*, 17(4), 371–383. <https://doi.org/10.31648/as-pal.2617>

- Oleniacz G., Len P. 2018. Modification of the algorithm of assessing an external patchwork of the land for the purpose of improving the efficiency of land consolidation works. Proceedings of 18<sup>th</sup> International Multidisciplinary Scientific GeoConference, SGEM, Book 2.2, 309–314. <https://doi.org/10.5593/sgem2018/2.2/S08.039>
- Rabczuk I. 1968. The problem of non-resident owners in the powiat of Proszowice, Kraków voivodeship. Current Issues in Rural Management Surveying, SGP, Warsaw, Poland, 65–70.
- Scott D.W. 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc.
- Shi Y., Cao X., Fu D., Wang Y. 2018. Comprehensive Value Discovery of Land Consolidation Projects: An Empirical Analysis of Shanghai, China. Sustainability, 10, 2039. <https://doi.org/10.3390/su10062039>
- Sołtysiak M. 2010. Statystyka opisowa [Descriptive statistics]. C.H. Beck.
- Sołtysiak A., Jaskulski P. 1998. Czekanowski's Diagram. A Method of Multidimensional Clustering. In: New Techniques for Old Times. CAA 98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, ed. J.A. Barceló, I. Briz, A. Vila. BAR International Series, 757, Oxford, 175–184.
- Sołtysiak A. 1997. UMCzek 1.00. Ewolucyjny algorytm porządkowania diagramu Czekanowskiego (UMCzek 1.00. Evolutionary algorithm for organizing Czekanowski's diagram). Biuletyn Antropologiczny, 1, 21–24.
- Stręk Ż., Noga K. 2019. Method of Delimiting the Spatial Structure of Villages for the Purposes of Land Consolidation and Exchange. Remote Sens., 11, 1268. <https://doi.org/10.3390/rs11111268>
- Sturges H. 1926. The choice of a class-interval. J. Amer. Statist. Assoc., 21, 153, 65–66.
- Tan S., Heerink N., Qu F. 2004. Land fragmentation and its driving forces in China. Land Use Policy, 23(3), 272–285. <https://doi.org/10.1016/j.landusepol.2004.12.001>
- Zeng S., Zhu F., Chen F., Yu M., Zhang S., Yang Y. 2018. Assessing the Impacts of Land Consolidation on Agricultural Technical Efficiency of Producers: A Survey from Jiangsu Province, China. Sustainability, 10, 2490. <https://doi.org/10.3390/su10072490>

---

Dr inż. Grzegorz Oleniacz  
Rzeszow University of Technology, Poland  
Department of Geodesy and Geotechnics  
e-mail: oleniacz@prz.edu.pl  
ORCID: 0000-0003-3318-0928

Prof. dr hab. inż. Karol Noga  
Rzeszow University of Technology, Poland  
Department of Geodesy and Geotechnics  
e-mail: k.noga@prz.edu.pl  
ORCID: 0000-0002-4890-9614

Mgr inż. Adrian Moskal  
Rzeszow University of Technology, Poland  
Department of Geodesy and Geotechnics  
e-mail: a.moskal@prz.edu.pl  
ORCID: 0000-0002-4888-7961