Land Consolidation based on Cluster Analysis

János Katona[1], Kornél Czimber[2], Andrea Pődör[1]

[1]Alba Regia Technical Faculty, Óbuda University, Budai út 45, H-8000 Székesfehérvár, Hungary, {katona.janos, podor.andrea}@amk.uni-obuda.hu

[2]Faculty of Forestry, University of Sopron, Bajcsy-Zsilinszky u. 4, H-9400 Sopron, Hungary, czimber.kornel@uni-sopron.hu

Abstract: The optimisation of land use structure is crucial to have a competitive agricultural production. In Hungary land consolidation lacks some important conditions such as a reasonable decision making support system based on Geoinformatics. In the paper, we optimize the land structure based on landownership. The present paper analyses the DigiTerra software that operates on the basis of Cluster Analysis and it provides a solution for the development. The improved software is introduced on a sample area. The efficiency of the allocation is proven by the internationally accepted fragmentation (Simmons, Januszewski, Igozurike) indices for parcels.

Keywords: land consolidation; cluster analysis; land valuation

1 Introduction

One of the main challenges of our time is to provide sustenance and clean water for people. The primary aim of the agriculture is to produce raw food materials to provide that sustenance. On the other hand, it is well known that monocultural production of thousands of hectares is not sustainable due to the stresses on the ecosystem and lowering of employment. It follows that adequate land politics can be both globally and locally provided that can support a land structure serving both personal and social interests.

The availability of agricultural assets in Hungary is favourable. However, the present fragmented land structure – resulting from the compensation and distribution of shareholdings in the 1990s – is not appropriate for competitive agricultural production. Therefore, it is necessary to consolidate the land.

According to the 2010 land use registry, on the average [1]:

- the number of land parcels used by a private person is 4.44 pieces
- the size of land parcels used by a private person is 9.54 hectares
- the number of land parcels used by professional farmers is 39.42 pieces
- the size of land parcels used by professional farmers is 296.38 hectares

The above-mentioned land sizes are not available for the maintenance of viable farms. The lease of land can somewhat help this situation, although it cannot provide a real solution because it removes capital from the production process. Since 2012 the policy to abolish the undivided joint ownership has also increased the extent of fragmentation. Consequently, the fragmented land structure cannot be maintained anymore. According to spatial data, the basis of land consolidation can be Geoinformatics.

2 The Background of Land Consolidation

2.1 The Measurement of Land Fragmentation

The efficiency of land consolidation methods can be measured by analysing land fragmentation, although land fragmentation has no objective measure. The number of parameters taken into the measurement is high and the scalability of parameters is not trivial. In this section, we introduce the index calculation accepted by international literature.

The index number by Simmons [2] gives a relevant value for a farm. This value is made up from the number of parcels *(n)*, the size of parcels (*a)* and the size of the whole farm (*A)*.

$$\mathrm{FI} = \frac{\sum_{i=1}^{n} a_i^2}{A^2} \tag{1}$$

Dorving [3] additionally, uses the distance that is taken by the farmer to reach one of his parcels. According to some critics [4], it would be more realistic to use both the back and forth distances and the annual frequency in the calculation.

A similar process was carried out by Januszewski [5], as well. He combined the number of land parcels belonging to a certain farm and their size distribution into a *K* factor. The value of this factor can change between 0 and 1, the nearer the value of K is to 0, the bigger the fragmentation is.

$$\mathrm{K} = \frac{\sqrt{\sum_{i=1}^{n} a_i}}{\sum_{i=1}^{n} \sqrt{a_i}} \tag{2}$$

The following consequences can be determined:

- the extent of fragmentation proportionally grows with the number of parcels,
- the fragmentation grows if the parcels are small,
- the fragmentation decreases if the size of big parcels grows and the area of small parcels falls off at the same time.

According to Igozurike [6], the average size of land parcels and the back and forth distances between them should also be taken into account.

$$P_i = \frac{1}{S_i / 100} Dt \tag{3}$$

, where P_i = fragmentation of the farm; S_i = the size of parcels; Dt = the whole back and forth distance.

The above-mentioned indices have three significant disadvantages:

- they neglect some spatial factors such as the index of land parcel per owner, the shape of parcels and some nonspatial factors like the type of the ownership or the accessibility of parcels,
- they are not flexible since the complex mathematical equation does not allow the separate handling of each and every member,
- they are not problem-oriented since they take the factors equally into account.

According to the literature [7, 8], each and every criterion (factor) and its entirety should fulfill some preconditions. Each and every criterion should be comprehensive to be measured objectively and reach its aim. The set of criterion should be complete, important aspects cannot be neglected. The criterion system should be flexible so that the problem can be divided into small parts such as economic, environmental, societal, etc. The final criterion system should be determined in a way to avoid the duplication of consequences of the decision. Duplication can be avoided in case of an additive result. If the correlation coefficiency of a criterion couple is near to 0, the two criteria are independent and not redundant. Therefore, six variables can be used [4]:

- the spatial location of land parcels,
- the size of land parcels,
- the shape of land parcels,
- the accessibility of land parcels,
- the type of the ownership,
- diversification of the ownership.

The so-called LandFragmentS Model [4] provides an index per land parcels by using the above-mentioned aspects and factors. The factors (f_{ij}) are determined with different weights (w_j). The weighted factors are contracted land parcel by land parcel (LFIi - land fragmentation index). The index referring to the whole area (GLFI - global land fragmentation index) comes from the summation of the land parcel index (LFI_i) / land parcel number (n) quotients.

$$LFI_i = \sum_{j=1}^{m} f_{ij} \cdot w_j \tag{4}$$

$$GLFI = \sum_{i=1}^{n} LFI_i / n \tag{5}$$

This complex evaluation method can provide an objective index for land proportion before and after the modification.

In the course of the DigiTerra software development, an evaluation method has been worked out on the basis of spatial data [9]. Besides land quality – that can be found in the real estate registry – , the method takes into account the shape, the size, the location, the accessibility, the relief and the slope conditions, the drainage, the irrigational conditions and environmental protection. Land quality found in the real estate registry can be modified with the above-mentioned factors, so we can get a modified Golden Crown value (mGC). In the present paper, we deal with the method of the allocation on the basis of previous studies.

2.2 Mathematical and Informatical Methods for Land Consolidation

The spread of information technology in the 1960s provided a new prospective for land consolidation, as well. In Germany, graphical data processing (e.g. David) and database management (e.g. Oracle) programs became the means of computer-aided planning. In 1984 the first CONEF (COmputerunterstützte NEuverteilung in der Flurbereinigung) land consolidation program package was developed. The program automatically processed alphanumerical data; however, the graphical data were processed manually. In 1990 at the Technical University of Munich the Chair of Land Management, the further development of CONEF, the CARE (Computer Aided Reallotment) was improved, which could manage both the regional data and the numerical demands of the owners.

Computer-aided land consolidation has antecedents in Hungary, as well. The task has already been approached by mathematical programming [10], combinatorial modelling [11] and Cluster Analysis [12]. According to the present operative law, the land consolidation module of the DigiTerra Map software is the most appropriate, which is based on voluntary land exchange institution. However, the original land structure is unchanged; the software re-allocates the ownership rights in a way that they are distributed to the nearest centroid of the previously owned land parcels. To implement the approach, it is essential to weigh land parcels in advance. “During scalability, three parameters are analysed: the owners’ previous land property, the distance of the nearest centroid, the rate of the distance of the second nearest and the furthest centroid. Therefore, the program provides a classification order, on the basis of which it rates parcels to the nearest centroid.” [12]

3 The DigiTerra Map Land Consolidation Module

3.1 The Work of the Module

The module carries out the planning based on three parameters (own land area, the distance of the first district, the distance of the second district) that can be weighted one by one. The weighting of the parameters can be fulfilled in various variables and the operation can be iteral. The coherence among the weighting of parameters, the number of iterations, the number of the parcels and owners of the planning should be defined numerically. The following analysis has been performed with the use of a generated test area containing 100 land parcels.

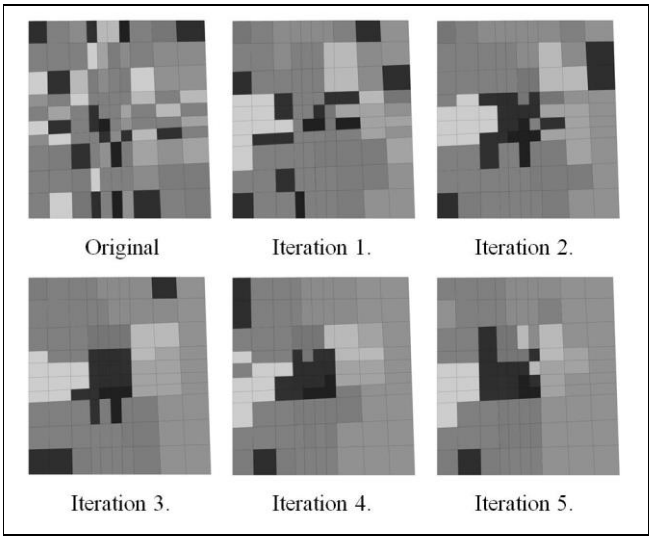

Figure 1

Distribution in 5 iterations with the help of the DigiTerra land consolidation module

Simmon’s land parcel index, c.f. eq. (1) can show the efficiency of the planning. Fig. 1 displays the analysis of the sample with 20 owners in 5 iterations. The weighting of parameters: 1.0, 10.0 and 1.0, respectively for the own land area, the distance of the first district, the distance of the second district, and the district distance is 10000 m. During the planning, Simmons’s land parcel index was determined by the combination of land parcels distributed next to each other. If an owner has not been given an area, the index cannot be counted, so its value is 1. The analysis shows that the highest degree can be weighted during the very first allocation. Some improvement can be observed, but not consequently, since its value is below some of the previous values (Fig. 2).

Figure 2
The coherence between the iterations and the fragmentations

There is no unambiguous connection between the shape factor of the land parcels and the number of the iterations. It is proven by Student's t-test, which says that the connection between the two data sets is $t = 4.209 > t_{0.05} = 4$, so they are found to be different from each other with 95% probability.

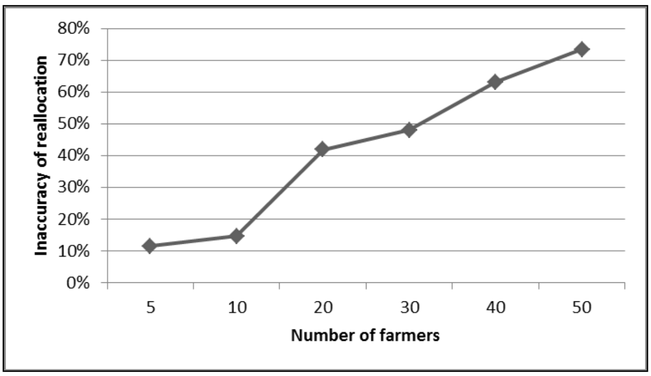

Figure 3
The inaccuracy of the allocation according to the number of the owners

As for the new land structure, an important aspect is value correspondence, in other words, the accurate allocation of the determined values. This accuracy is influenced by the number of the owners (Fig. 3) among other factors. The percentage accuracy of the distribution has been analysed in an area containing 100 parcels according to 6 different owners.

The positive correlation can be observed between the inaccuracy of the allocation and the number of the owners, the correlation coefficiency is r=0.984. A similar coherency can be observed between the number of the iterations and the inaccuracy of allocations (Fig. 4).

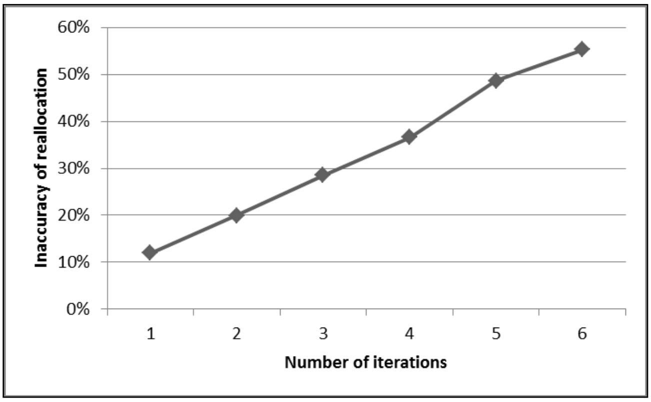

Figure 4

The accuracy of the allocation according to the iterations

The weighting of parameters can significantly influence the result (Fig. 5). To prove this, three different weighting variables have been analysed: 1) parameter: own land area, 2) parameter: the distance of the first district, 3) parameter: the distance of the second district). Noticeably, among the analysed variables, the bigger-than-average weighting of the own land area is the worst, while the most favourable result was given by the above-average weighting of the second district.

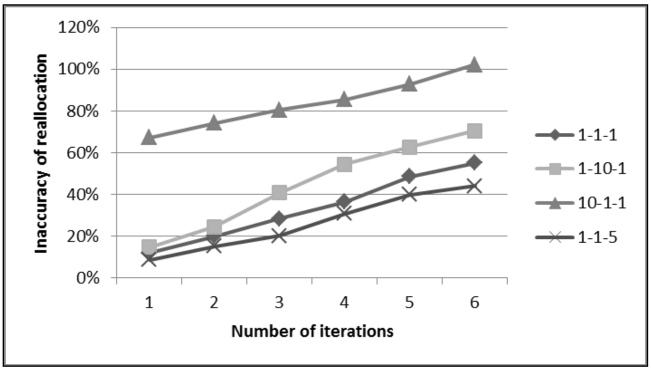

Figure 5

The inaccuracy of the allocation according to the weight of parameters

To plan optimally, it is not practical to carry out a huge number of the iterations since it can reduce the accuracy of allocation. However, it does not raise the fragmentation and the shape factor value of the parcels considerably. Since the sample data is not articulated by linear infrastructure, the results are true in general, as well.

3.2 The Development of the Module

Applicable conditions of the development:

- to keep the exact rate of ownership in the course of allocation

  The software does not modify the land structure but it assigns new owners to the existing land parcels. The accuracy of reallocation can have a maximum value that is correlated to the exchange value of the land parcel with the smallest value. Therefore, owners with few land parcels can be given areas whose exchange value is significantly different from the original. The number of iterations worsens the value protection as well. The present work of the module allows the repetition of the allocation. It is possible to boost land concentration by growing the number of iterations, on the contrary, the difference between the initial and the final value in exchange will be extensive. The accuracy of re-allocation can be grown by parcelling the area for units. This operation can be carried out in land parcels or tables. The latter can be supported only in the case of institutionalised land consolidation since new land structure and land parcel boarders come into existence.

- to give the demands of the owner

  Based on the present work of the module, the coordinates belonging to the owners are determined so that the module counts the centroid of the areas that belong to the same owner. This method is objective; however, it does not necessarily give an optimal solution. For instance, if earlier an owner lived further from the place of production, this problem is still present in the new land structure. To solve this problem, instead of centroid dispersion, the farming premises and the place of living may be defined as a preliminary condition. This method would be in line with the operative land law. According to this law, residents would be preferred for the purchase of land.

As a summary, the following hypothesis can be formulated: based on the owners' aspects, giving the farming premises and lamellating the area would be more accessible and favourable.

4 Planning on the Sample Area

4.1 Sample Data

We would like to carry out the analysis both in mountainous and plain areas, but the Ministry of Agriculture has merely provided the data of Mesterszállás for the analysis. This village is on the Great Hungarian Plain. It is 42.92 km2 and its population is 702 people. 508 land parcels have been involved in the planning. Out of these 508 land parcels 182 owners own 335 land parcels with 1/1 ownership proportion; 245 owners own 173 land parcels with undivided joint ownership. On the sample area 424 natural people and 3 legal people are registered into Land Registry. According to the modified Golden Crown (mGC) [9], land size distribution is displayed in Table 1. The size of the area is 67578.05 mGC, the average size of the area per person is 158.26 mGC.

The relevant authority has given the digital and certified cadastral maps of the sample area. The Department of the Land and Geoinformatics of the Ministry of Agriculture and Regional Development has provided the data of the owners and land users free of charge with the condition that the personal data can be accessed only in encoded format. The encoding of owners and users was carried out with a 6-digit-tag. Since the encoding was fulfilled independently, they cannot correlate with each other. The owners' data are formed into an .xls format with 11columns and 2716 rows. The attributes of the statement: location, profile number, the owner's tag, legal status, counters of the ownership interest and denominator of the ownership interest. The land user data table contains the following information: location, profile number, land usage, quality class, used area and the user's tag.

According to the licence of the Ministry, the Institute of Geodesy, Cartography and Remote Sensing has provided the following mapping data:

- the settlement boundary of the municipality
- parcel boundary in the municipality
- profile numbers of the parcels
- the boundary of land usage
- building boundary
- labels of the cadastral map
- 1:10 000 topographical map
- aerial photograph without deformation
- relief model
- 1:50 000 land cover
- 1:100 000 land cover
- MePAR 2012 block map
- MePAR 2012 thematic layers

4.2 Planning

In the course of planning, more alternatives have been carried out. The number of parcels and the owners are almost equal, so planning based on parcels and ownership proportion cannot provide an appropriate solution. According to the present land policy, it is legitimate to establish a claim to a planning based on land usage. Land lease – along the farmers' demand – basically means use-based land concentration in the sample area as well. The number of land users in the sample area (83 pieces; abbreviated as "pcs") is much lower than the number of the owners (427 pcs), so the situation is conspicuous.

The planning was carried out based on two variables, in four ways. The very first variable is the base unit of the allocation, which can be the land parcel or the lamella (1 ha unit area). The second variable is the starting point of the allocation which can be the centroid of the area belonging to the owner or a freely-given premise coordinate. Dividing the land parcels into unit areas has resulted in increasing the number of the areas (508 to 3728) that can be used in planning. Furthermore, the accuracy of allocation can get better. Lamella based planning is displayed in Fig.6.

Figure 6
Lamella based planning on the sample area

During land parcel based planning, the sample area was allocated according to the mGC values determined in a previous study [9]. During lamella based planning, the evaluation was carried out again, but the shape and size factors were not present due to the artificially produced shape of the lamellas.

4.3 Results

To compare the planning versions, the land parcel indices (Simmons, Januszewski and Igozurike) and the accuracy index of the allocation were used (Table 1).

It turns out from the analysis that all of the planning versions have more favourable index numbers compared to the preliminary situation. (While for Simmons and Januszewski index 1 stands for the best value; according to Igozurike, index 0 means the best one.) According to the indices, there is no unambiguous difference between the land parcel based planning and lamella based planning. According to the Simmons-index, lamella based planning is favourable, but Igozurike finds land parcel based planning to be better. The accuracy of allocation is more favourable especially in lamella based planning, but the number of parcels per farmer is lower in the case of land parcel based planning.

Table 1
Indices of the planning versions carried out on the sample are

	Starting point	Planning based on parcel		Planning based on lamella	
		with centroid	with coord. of estate	with centroid	with coord. of estate
Number of parcels	508	309	243	497	325
Parcel per capita	6.120	3.723	2.928	5.988	3.916
Index according to Simmons	0.678	0.720	0.750	0.725	0.756
Index according to Januszewski	0.749	0.785	0.814	0.764	0.791
Index according to Igozurike	1.081	0.600	0.299	0.942	0.531
Accuracy of distribution		0.994	0.919	0.999	0.996

The hypothesis, which says that lamella based planning using premise coordinates is more optimal than land parcel based planning with the use of centroid coordinates, has been proven by the analysis. Considering the accuracy of allocation and the indices of fragmentation, the lamella based planning has been accepted.

In terms of the evaluation of the developed method it would have been useful to test different kinds of methods on the sample area. However, there were not any possibilities for it due to the lack of IT support.

In order to take the personal demands of the farmers and the importance of their participation in land consolidation into account, the accepted planning version has been worked on with the use of shape and size factor and mGC. The data was normalised in a way that all shape and size factors vary between 0 and 1. The need of normalization arose from the different number of parcels. The area calculation under the normalised function was carried out by numeric integration the results of which:

$$T_{bLC} = \int_0^1 f_{bLC}(x) = 0{,}504 \quad (6)$$

$$T_{aLC} = \int_0^1 f_{aLC}(x) = 0{,}529 \quad (7)$$

It turns out from the analysis that after the change the shape factor of the areas is cumulatively more favourable. However, the narrow parcels have come into existence in limited number as well. The formation of such parcels can be improved by the optimal spatial distribution of premises (e.: minimalizing the centroid of the areas distributed into one block).

The mGC values counted before and after the change can contribute to the persuasion of the farmers and the effectiveness of land consolidation. The grade mGC values are shown in Fig. 7, in logarithmic scale. To compare the different data, it is not necessary to normalize the values because mGC values are not specific. The analysis needs cumulated summation.

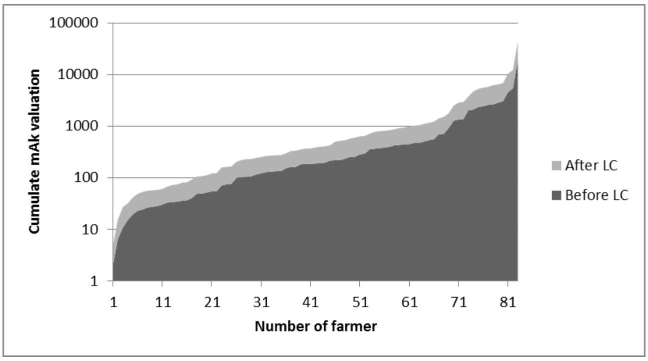

Figure 7

Lamella based planning on the sample area

The results of the summation:

$$\sum_{i=1}^{n} mGC_{bLC} = 67578.05 \qquad (8)$$

$$\sum_{i=1}^{n} mGC_{aLC} = 81771.07 \qquad (9)$$

It turns out from the analysis that the cumulated mGC value of the sample area has grown by 21% which is mainly as a consequence of the shape factor. Based on the land fragmentation indices and the mGC values (cumulated by shape and size factors), it could be said that the land consolidation was successful.

Conclusion

The present paper has dealt with the problems and solutions of land consolidation, prioritizing Geoinformatics. The method of land fragmentation calculation, the necessary data of planning, and the analysis of the DigiTerra land consolidation module and further development opportunities have been introduced. Land consolidation has been carried out on a sample area; the results have been proven by land fragmentation indices, the shape factor and mGC calculations. According to the analysis, the planning on the basis of lamella with coordinate determination has been proposed. The results have proved the efficiency of the proposed method since it resulted in a 21% increase of the cumulative mGC of the sample area. The validation of the method can be achieved by further tests on differently scattered areas. The methodology can be flexibly shaped according to the participants' demands since the aim is their satisfaction. Therefore, land consolidation can contribute to the competitive production, the sustainable production, and to the correction of the quality of life in rural areas.

Acknowledgement

The research was supported by the Department of the Land and Geoinformatics of the Ministry of Agriculture and Regional Development Hungary. We thankfully, acknowledge the Óbuda University Alba Regia Technical Faculty.

References

[1] J. Ripka: Az agrártárca hosszú távú földbirtok-politikája, Háttéranyag, Vidékfejlesztési minisztérium Földügyi Főosztály, FF/1744/2011, 11 p

[2] A. J. Simmons: An index of farm structure, with a Nottingahmsire example, East Midlands Geographer, 3. 1964, pp. 255–261

[3] F. Dovring: Land and labour in Europe in the twentieth century (3rd ed.) Hague: The Nijhoff. 1965

[4] D. Demetriou: The Development of an Integrated Planning and Decision Support System (IPDSS) for Land Consolidation, Doctoral Thesis accepted by the University of Leeds, UK, ISBN 978-3-319-02346-5, 2014, 351 p

[5] J. Januszewski: Index of land consolidation as a criterion of the degree of concentration. Geographia Polonica 14, 1968, pp. 291–296

[6] M. U. Igozurike: Land tenure, social relations and the analysis of spatial discontinuity. Area, 6, 1974, 132–135

[7] R. Keeney, H. Raiffa: Decision with multiple objectives: Preferences and nvalue trade-offs. Cambridge: Cambridge University Press. 1993

[8] A. Sharifi, M. Herwijnen, W. Toorn: Spatial Decision Support Systems. Lecture Notes. ITC, International Institute for Geo-Information Science and Earth Observation, The Netherlands. 2004

[9] J. Katona: The Application of Fuzzy Logic in the Field of Land Consolidation, 10th International Symposium on Applied Informatics and Related Areas (ISBN:978-615-5460-49-4), P15. 2015, 4 p

[10] P. Gáspár: Birtokrendezési feladatok megoldása matematikai programozással. Budapesti Műszaki és Gazdaságtudományi Egyetem Általános- és Felsőgeodézia Tanszék, 2003, 12 p

[11] M. M. Csordásné: Matematikai modell a birtokrendezés támogatására, Geodézia és Kartográfia, Budapest, 57. évf. 2. szám. 2005, pp. 24-30

[12] K. Czimber: Képfeldolgozási és geoinformatikai algoritmusokon alapuló birtokrendezési eljárás kifejlesztése, Geodézia és Kartográfia 65. évf. 2013/11-12. pp. 15-18