INFRASTRUKTURA I EKOLOGIA TERENÓW WIEJSKICH
INFRASTRUCTURE AND ECOLOGY OF RURAL AREAS

No IV/3/2017, POLISH ACADEMY OF SCIENCES, Cracow Branch, pp. 1887–1902
Commission of Technical Rural Infrastructure

DOI: http://dx.medra.org/10.14597/infraeco.2017.4.3.143

COMPARISON OF THE RESULTS OF SPATIAL ANALYSES PERFORMED IN ARCGIS AND QGIS SOFTWARE

Piotr Benduch
AGH University of Science and Technology

Abstract

The geoinformation software market is becoming increasingly competitive. New improvements and tools are created, which are provided to the users with subsequent updates. Additionally, completely new software is appearing as well. Nevertheless, for the past several years, ArcGIS and QGIS have been the most popular software in Poland. There are many handbooks and manuals describing their possibilities. Both systems offer numerous functions. This does not mean, however, that the results of identical spatial analyses performed in ArcGIS and QGIS will be identical as well. The comparison of the results of such analyses performed on vector and raster data is the main objective of the conducted study. The research demonstrates, that selected analogical tools implemented in discussed systems provides different results especially in the range of spatial analyses performed on raster data.

Keywords: ArcGIS, QGIS, spatial analyses, comparison of results

INTRODUCTION

In response to the high demand for geoinformation software in recent years, an increase was observed in releasing new, upgraded (according to manufacturers) versions of the existing software and completely new applications available both as free software and closed software. Innovative solutions and functionality, together with the increased efficiency, are the main objectives for implementing and developing these systems.

This article aims to compare the results of vector and raster analyses carried out in ArcGIS 10.4 and QGIS 2.8 software. The problem of using geographical spatial information systems for various purposes has already been the subject of several publications (Anselin 1992, Basista 2015, Chang 2006, Głowienka and Wójcik-Leń 2015, Litwin and Szewczyk 2012, Wheatley 1995), as well as the issue of the implementation of certain functions in the discussed software (Mašíček, Kozlovsky Dufková, Stejskal 2016, Fotheringham and Rogerson 2013, Hugentobler 2008, Theobald 2007, Wong and Lee 2005). A large number of articles, handbooks and book titles related to the subject prove its validity and importance. Different versions of ArcGIS and QGIS software were also compared, but the analysis was focused on how they can be used for specific tasks (Friedrich 2014, Kuka and Bushati 2014, Lis, Mikrut, Guzik 2007, Rosca, Chelaru, Plescan 2013). The influence of the input data quality on the final results of the spatial analysis was also researched (Hejmanowska 2005, Malczewski 1999). However, so far not many publications have been identified which would concentrate on a direct confrontation of the results of the identical analyses performed in two, currently most popular, geoinformation software systems in Poland. Such research studies will be useful especially to commercial users using these tools on a large scale.

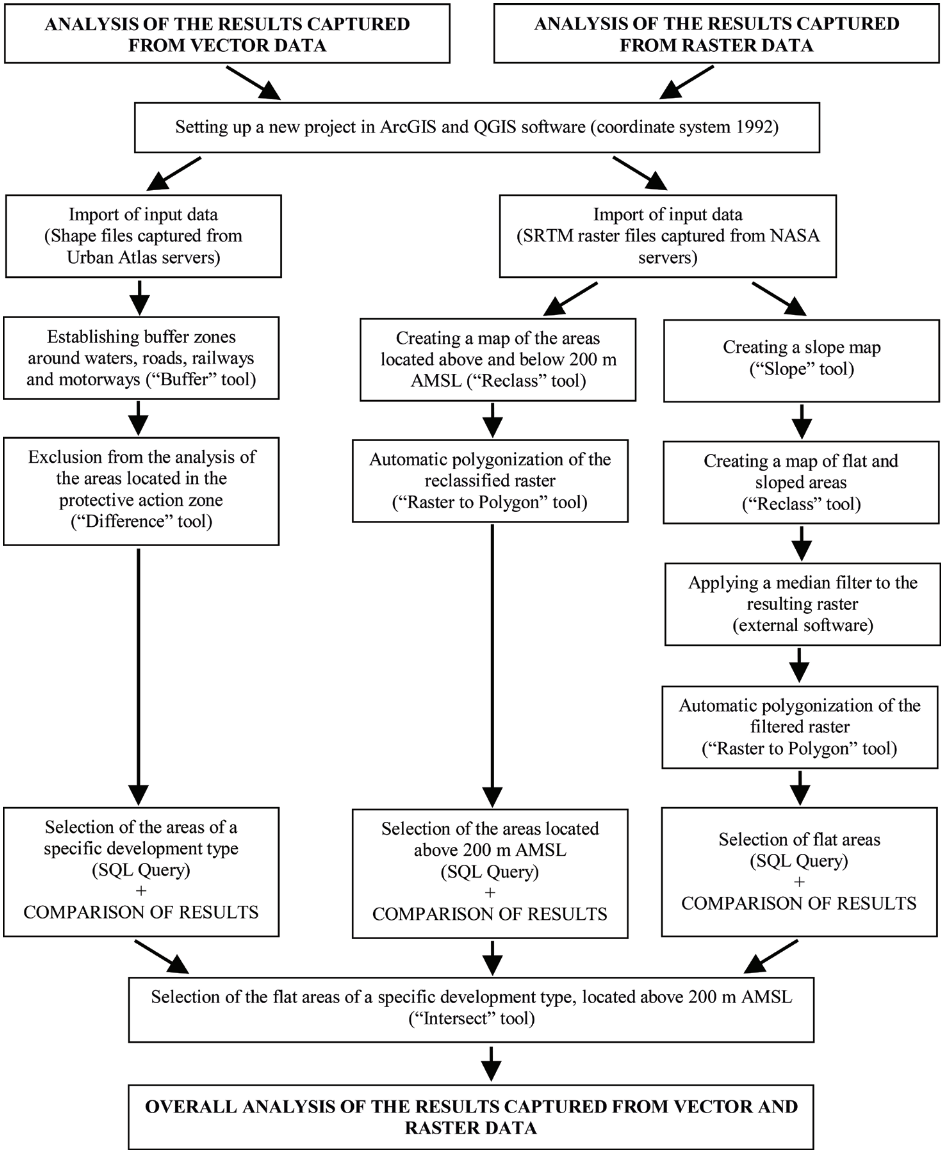
RESEARCH METHODOLOGY

For the purpose of conducting the comparative analysis of the results obtained in ArcGIS 10.4 and QGIS 2.8 software, hereinafter referred to as ArcGIS and QGIS, respectively, identical operations were performed on the collected data in both systems. The basic research tools available in the specific software were used, such as buffering, intersection, difference or reclassification. The procedure was divided into three main stages:

- analysis of the results captured from vector data;
- analysis of the results captured from raster data;
- comprehensive analysis of the results captured from vector and raster data.

The whole process was demonstrated in Figure 1.

Confrontation of the results was carried out directly in ArcGIS software. Using the statistics tools, the surface areas and the quantity of the resulting polygons were compared. A visual assessment was conducted as well. The results of the studies have been presented in the following subsections.



Source: own study

Figure 1. Diagram presenting the course of the performed analyses

RESULTS OF THE ANALYSES ON VECTOR DATA

Spatial analyses on vector data, conducted according to the diagram presented in Figure 1, were performed on the materials developed as part of the European Urban Atlas project. These data represent functional zones of urban areas. The Shape file stored on the Urban Atlas server, prepared for the city of Suwałki (Podlasie Province), was used.

First, in ArcGIS and QGIS software, using the default settings (in the case of ArcGIS – the planar method), a buffer tool was used to determine protective action zones of predetermined width around the areas of a specific use (Table 1).

Table 1. Width of protective action zones around the selected areas

Land development type	Buffer [m]
Water bodies	100
Other roads and associated land	25
Railways and associated land	25
Fast transit roads and associated land	50

Source: own study

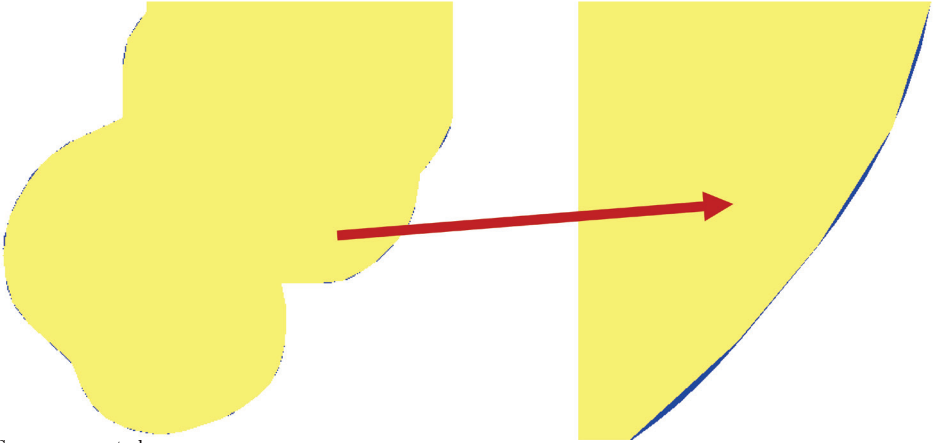
Already at this stage, the first differences between the results obtained in the analyzed geoinformation systems were noted (Table 2).

Table 2. Comparison of polygon surface areas having generated protective action zones

Polygon	Polygon surface area [ha]		
	Water bodies	Other roads, railways and associated land	Fast transit roads and associated land
ArcGIS	6001.36	6913.67	20.76
QGIS	5990.31	6913.22	20.75
Difference	11.05	0.45	0.01
Discrepancy	**0.2%**	**0.0%**	**0.0%**

Source: own study

It was found that the main factor causing the discrepancy between the results obtained during buffering are the generalization algorithms implemented in ArcGIS and QGIS. This is confirmed by the visual analysis, demonstrated in Figure 2.



Source: own study

Figure 2. Visual assessment of the differences occurring while generating buffer zones in ArcGIS (blue) and QGIS (yellow)

It can therefore be concluded that the algorithms for simplifying the polygons generated during the use of the buffering tool are more intensive in QGIS. This is illustrated by the cut edges of the resultant polygons. Moreover, in all the cases, larger surface areas of the objects generated in ArcGIS were noted.

It should also be mentioned that, despite the larger surface area of the buffer zone generated around roads, railways and the connected areas, the differences in the analyzed software systems are smaller than in the case of the resultant polygons generated around water-covered areas. It is due to the use of the smaller buffer (Table 1) and the linear geometry of the polygons representing roads and railways. Fewer turn points of the boundaries of input polygons result in the objects generated in the buffering process having a less complex shape. Then, the generalization algorithms used by ArcGIS and QGIS have a smaller impact on the final result of the analysis.

The last stage was to select the areas outside the protective action zones, of a specific land use. The „Difference" tool was used, followed by an appropriate SQL query. Table 3 demonstrates the information on the surface area of the resultant polygons and the number of objects.

The resulting discrepancies are a consequence of the differences which have arisen at the stage of generating buffer zones around the areas occupied by water. As a result of the selection, many areas bordering these territories were rejected. For this reason, the final discrepancies are smaller than those demonstrated in Table 2. Larger surface area was observed for the resultant polygons generated in QGIS software, which, in connection with the results demonstrated in Table 2, proves the correctness of the performed tests and calculations. Using the surveying method of generating buffer zones in ArcGIS yielded the results, for which similar conclusions can be drawn.

Table 3. Comparison of the surface areas of the polygons and the number of objects after using the „Difference" tool and after the selection

Software	Polygon surface area [ha]	Number of objects
ArcGIS	35921.03	1054
QGIS	35924.87	1055
Difference	3.84	1

Source: own study

It is worth noting that different generalization algorithms used by the analyzed geoinformation systems during buffering operations led to the differences in the number of the resultant objects. In the case of the analyses aimed at selecting objects threatened by the influence of a certain factor, this issue may be of major importance.

RESULTS OF THE ANALYSES ON RASTER DATA – RECLASSIFICATION AND AUTOMATIC VECTORIZATION

Spatial analyses on raster data were based on the raster files from the NASA servers, demonstrating a terrain model with a grid mesh of 90 m, created during the international Shuttle Radar Topography Mission (SRTM). It is worth mentioning that the SRTM is the first case of using single-pass radar interferometry to obtain a digital elevation model from the Earth's orbit (Bamler 1999).

After the relevant raster had been adopted, it was reclassified to identify the areas located above and below 200 m AMSL (Figure 3).

Old values	New values
0 - 200	1
201 - 350	2
NoData	NoData

Source: own study

Figure 3. Reclassification window in ArcGIS software

Reclassification itself is the assignment of new values to individual pixels. It gave identical effects in both cases. Then, „Raster to Polygon" tool was used.

In this way, the raster model was converted into a vector model. Table 4 demonstrates the comparison of the polygons.

Table 4. Comparison of the polygons generated by automatic vectorization of the reclassified raster

Software	Surface area of land exceeding 200 m AMSL [ha]	Number of objects
ArcGIS	78746.85	374
QGIS	78746.85	374
Difference	0	0

Source: own study

The obtained result demonstrates that the subject tools in ArcGIS and QGIS produced identical results in terms of the number and the surface areas of the polygons. This means that the same algorithms are used for automatic vectorization. The possibility of simplifying the resultant polygons was not used. It should be emphasized that for such comparisons it is important to apply the same definitions of the coordinate systems. Otherwise, there may be differences in the surface areas of the generated objects calculated by the software, which may lead to the distortion of the study results and incorrect conclusions.

RESULTS OF THE ANALYSES ON RASTER DATA – A SLOPE MAP

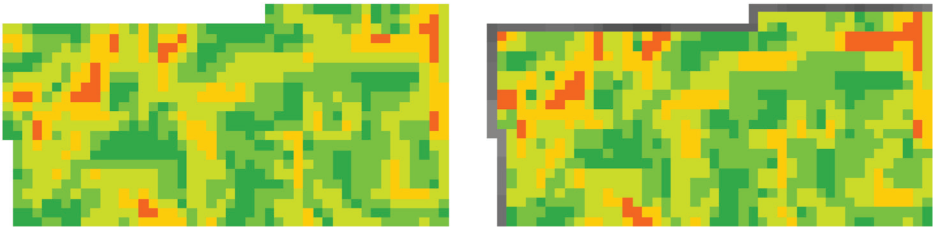
The next stage of the analysis involved the use of the input raster file to generate a slope map. For this purpose, it was necessary to use the „Slope" function available in ArcGIS software in the ArcToolbox, and in the case of QGIS – in the GRASS toolkit. The slope maps, with the slopes expressed in degrees, were prepared with default settings (Z factor = 1) and the same coordinate system (National Geodetic Coordinate System 1992). The resultant rasters were subjected to reclassification. The land with less than or equal to 1° slope was considered flat land.

First of all, it is worth noting that the slope map generated in QGIS has 6,880 fewer pixels than the same map generated in ArcGIS (Table 5). This is a result of the algorithm to calculate a slope, used in QGIS software, which decreases the resultant map by exactly one row of pixels on each of its edges relative to the source image. In other words, the edge pixels are removed. This is not the case with ArcGIS software, as demonstrated in Figure 4.

Table 5. Analysis of the number of pixels and the maximum slope values of the generated maps

Software	Number of pixels representing flat land (slope ≤ 1°)	Number of pixels representing sloped land (slope >1°)	Maximum slope value [°]
ArcGIS	685292	1400386	22.47
QGIS	645226	1433572	25.64
Difference	40066	33186	3.17

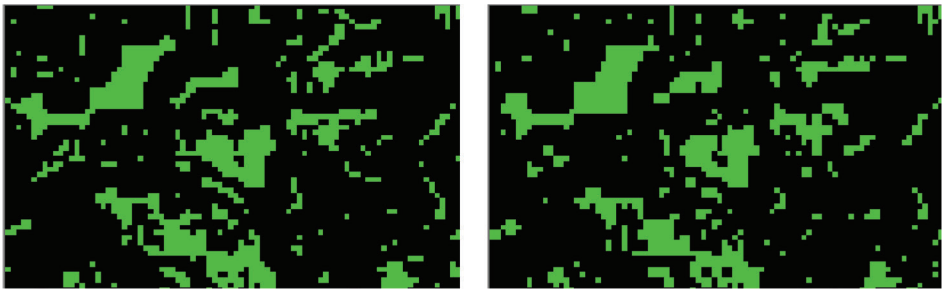
Source: own study



Source: own study

Figure 4. Fragments of the generated slope maps prior to reclassification in ArcGIS (left) and QGIS (right)

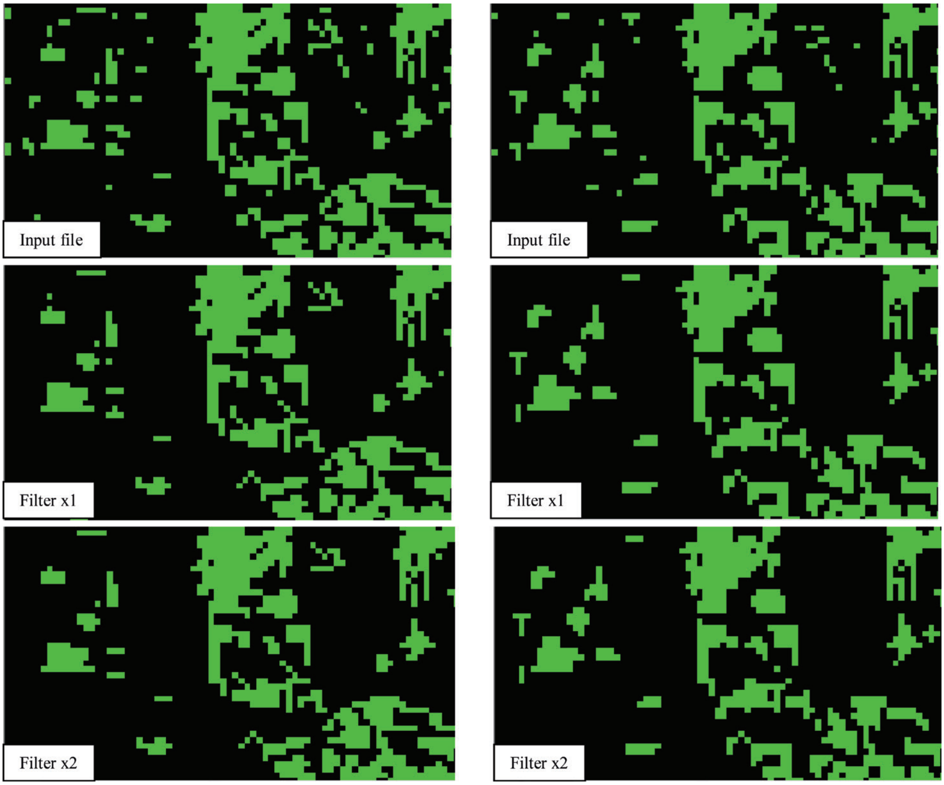
In the case of the reclassified maps generated in QGIS, a greater number of pixels representing the sloped land and a higher maximum slope value were observed. Therefore, it was finally confirmed that different algorithms were responsible for generating the slope maps in both systems. This is also confirmed by the visual assessment of the obtained results (Figure 5).



Source: Own study

Figure 5. Visual assessment of the reclassified slope maps generated in ArcGIS (left) and QGIS (right)

The observed differences are important and may have a significant influence on the final outcome of spatial analyses. The next step involved the application of median filters to the generated elevation maps for the purpose of their smoothing and noise elimination (Figure 6). Such a procedure was used to make it easier to implement the polygonization of the raster. The median filter was used twice.



Source: own study

Figure 6. The influence of applying the median filter to raster files generated in ArcGIS (left) and QGIS (right)

It was found that further filtering of the raster files in question was unnecessary. Further iterations do not introduce significant changes, and the use of another, more aggressive algorithm could lead to distortion of the final results.

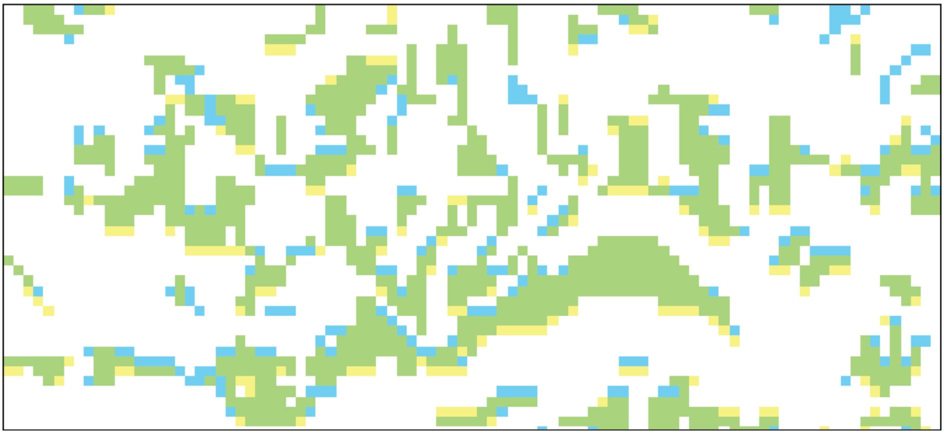
Double-filtered, reclassified slope maps were converted to vector files by using the „Raster to Polygon" tool. This process was performed in ArcGIS. Then, flat land was selected. The surface areas and the number of resultant objects are demonstrated in Table 6.

Table 6. Surface areas and number of selected objects after the transformation of the slope maps into vector form

Software	Surface areas of the resultant polygons [ha]	Number of objects
ArcGIS	155234.48	28427
QGIS	144067.02	25857
Difference	11167.46	2570
Discrepancy	**7.2%**	**9.0%**

Source: own study

The results of the analyses performed on raster data are significantly more divergent than the results of the analyses carried out on vector data. This is mainly the effect of the various algorithms implemented in ArcGIS and QGIS software to generate slope maps. This means that in the case of the analyses requiring higher accuracy and reliability, it would be necessary to get acquainted with internal algorithms used for generating different types of products on the basis of raster files in individual systems. Uncritical acceptance of the results of such analyses could have significant adverse consequences.



Source: Own study

Figure 7. Flat land selected in ArcGIS (blue), QGIS (yellow), and flat land selected by both systems (green)

Depending on the nature of the performed analyses, one of the solutions to the problem described could be the selection of specific areas (flat land in this case) that met certain criteria in both analyzed geoinformation systems. Such an approach would increase confidence level regarding the outcome of a given process, as well as improve the quality of the final product. Figure 7 demonstrates a fragment of overlaid resultant maps which illustrate the land with a slope of less than 1°.

The land marked in green, with identical input parameters, was defined as flat land by both algorithms (in ArcGIS and in QGIS). It can be stated therefore that their use in further analyses will be less uncertain than the use of the areas identified by one of the studied programs only.

RESULTS OF THE ANALYSES ON VECTOR AND RASTER DATA

The last stage of the study was to intersect the results of the analyses performed on vector and raster data to capture the information on the location of flat land of a specific development type, located higher than 200 m AMSL. The „Intersect" tool was used. Table 7 demonstrates the information on the number of objects and the surface areas of the resultant polygons.

Table 7. Surface areas and number of polygons after intersection of the results of the analyses on raster and vector data

Software	Surface areas of the resultant polygons [ha]	Number of objects
ArcGIS	2323.54	1171
QGIS	2173.63	1046
Difference	149.91	125
Discrepancy	**6.5%**	**10.7%**

Source: own study

The differences presented in Table 7 are primarily the consequence of the operations performed on raster data. Surface area variation of 6.5% should be considered as significant. It is worth noting that the use of the „Intersect" tool yielded identical results in both programs.

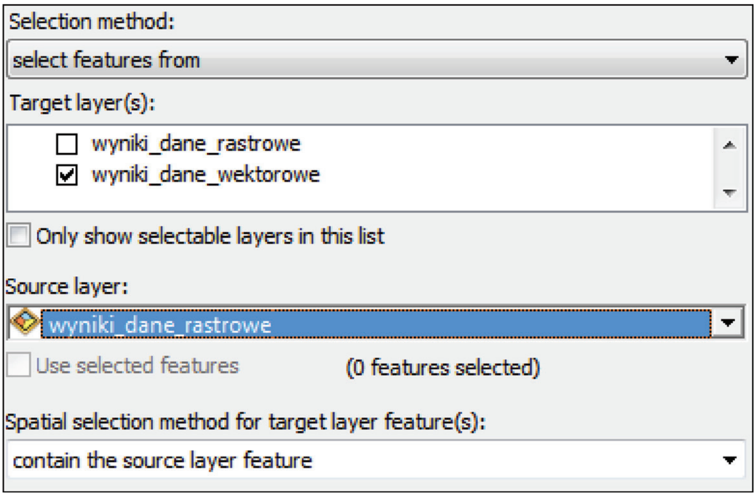
The study also used the approach mentioned at the end of the previous chapter. Only the common part of the flat areas specified by ArcGIS and QGIS was used for the final selection. The effects are presented in Table 8.

Table 8. Surface areas and number of polygons generated by intersection of the common part of the results of the analyses on raster data with the results of the vector analyses obtained in each program

Software	Surface areas of the resultant polygons [ha]	Number of objects
ArcGIS	1846.28	922
QGIS	1846.31	922
Difference	0.03	0

Source: own study

The obtained results in both programs are almost identical. The only differentiating factor are the algorithms for generalizing buffer zones. Nevertheless, when comparing the information contained in Table 7 and Table 8, it can be stated that both the number of the objects and their total surface area were significantly reduced as a result of the implementation of the suggested method. However, in the case of the analyses aimed to select the location of specific areas, while minimizing the risk of error, the proposed approach may produce satisfying results, especially during operations performed on raster data. The final decision always belongs to the user, though. To ensure the highest quality of the final result, the aim of the analyses and the accompanying circumstances should be taken into account.



Source: own study

Figure 8. Selection by the location performed in ArcGIS

Finally, an attempt was made to estimate the influence of the discrepancies between the results of the conducted analyses in the studied programs on the result of the selection by location. The most popular selection types which were used included the following:

- target layer feature are within the source layer feature;
- target layer feature intersect the source layer feature;
- target layer feature contain the source layer feature.

The layer created from the results of the analyses on raster data, representing flat land located higher than 200 m AMSL, was chosen as the source layer. The target layer, on which the specific objects were selected, was the layer containing the results of the analyses on vector data, i.e. the polygons representing the land of a specific development type, located outside the predetermined area of the protective action zones. Figure 8 illustrates the selection procedure in ArcGIS software, taking these layers into account. Table 9 presents the results of performed operations.

Table 9. Surface area and number of objects selected by different methods

Software	Target layer feature are within the source layer feature		Target layer feature intersect the source layer feature		Target layer feature contain the source layer feature	
	Surface areas of the resultant objects [ha]	Number of objects	Surface areas of the resultant objects [ha]	Number of objects	Surface areas of the resultant objects [ha]	Number of objects
ArcGIS	1.28	5	17311.25	221	13484.03	98
QGIS	6.20	7	17519.52	221	13214.38	95
Difference	4.92	2	208.27	0	269.65	3
Discrepancy	**79.4%**	**28.6%**	**1.2%**	**0.0%**	**2.0%**	**3.1%**

Source: own study

In all of the analyzed cases, the difference in the number of objects did not exceed 3. This result does not seem to be fully satisfying, though. It proves that after performing a series of similar operations on the same data in two geoinformation systems, the obtained results may be different due to the implemented internal algorithms. This is also confirmed by the analysis of the surface area of the selected polygons, where the greatest discrepancy of almost 80% was noted for the first of the selection types. In the context of the spatial analysis of high risk, and the responsibility of those conducting it, this is very important information. Uncritical acceptance of the results of the analyses as well as considering them certain and reliable can have serious consequences.

It should be emphasized that the input raster data used in the tests was of average accuracy, and the operations performed in ArcGIS and QGIS were simple, basic operations, which could be performed even by less advanced users. Nevertheless, the final results demonstrate that the tools implemented in the software, especially those used for analyzing raster data, can ultimately lead to divergent results. This problem should be covered by further, more extensive studies.

SUMMARY AND CONCLUSIONS

The subject of the study was a direct comparison of the results of the spatial analyses performed on vector and raster data in ArcGIS and QGIS software. Their implementation made it possible to formulate the following conclusions:

1. Most of the vector data operations performed in both systems yielded identical results. The exception was the „Buffer" tool which, due to different generalization algorithms, led to the small but possibly significant discrepancies.
2. The more complicated geometry of the polygon, the greater differences in the buffer zones around it, generated in ArcGIS and QGIS.
3. Much more varied results were obtained in the case of the spatial analyses performed on raster data. The exception was the reclassification of the raster, which yielded identical results in both geoinformation systems, and so did the automatic vectorization without simplifying the resultant polygons.
4. The biggest discrepancies were recorded for the „Slope" tool, used to generate slope maps. These differences significantly affected the final results of the analyses performed jointly on vector and raster data, as well as the results of the selection of the polygons according to the location.
5. Uncritical acceptance of the results of the spatial analysis may have negative consequences, depending on the complexity of the analysis and the risk of error. Such operations should be performed by experienced users who have the knowledge of the particular algorithms implemented in the given geoinformation system.
6. Using the common part of the results of specific analyses performed on raster data in several programs seems to be an effective way to minimize the risk of error.
7. The studied problem remains relevant and should be covered by further research with other geoinformation software taken into consideration, as well as the available analysis tools.

REFERENCES

Anselin L. (1992). *Spatial data analysis with GIS: an introduction to application in the social sciences.* Technical Report 92-10.

Bamler R. (1999). *The SRTM Mission-A World-Wide 30 m Resolution DEM from SAR Interferometry in 11 Days.* Photogrammetric Week 1999: 145-154.

Basista I. (2015). *Przykłady wykorzystania narzędzi GIS w procesie scalania i wymiany gruntów.* Infrastruktura i Ekologia Terenów Wiejskich 4: 1047-1055. DOI: 10.14597/infraeco.2015.4.1.083

Chang K.T. (2006). *Geographic information system*. John Wiley & Sons, Ltd.

Fotheringham S., Rogerson P. (2013). *Spatial analysis and GIS*. CRC Press.

Friedrich C. (2014). *Comparison of ArcGIS and QGIS for applications in sustainable spatial planning*. University of Vienna, PhD Thesis.

Głowienka E., Wójcik-Leń J. (2015). *Application of GIS Analyses to Identify the Problematic Agricultural Areas in the Course of Land Consolidation*. Geomatics and Environmental Engineering, 9(4): 45-55. DOI: 10.7494/geom.2015.9.4.45

Hejmanowska B. (2005). *Wpływ jakości danych na ryzyko procesów decyzyjnych wspieranych analizami GIS*. AGH Uczelniane Wydawnictwa Naukowo-Dydaktyczne.

Hugentobler M. (2008). *Quantum GIS. In: Encyclopedia of GIS*. Springer US: 935-939.

Kuka S., Bushati J. (2014). *On the role of GIS technology in Geospatial Education (QGIS vs ARGIS)*. CSIS & MER: 98-102.

Lis N., Mikrut S., Guzik M. (2007). *Możliwości wykorzystania darmowego oprogramowania w budowie bazy danych GIS dla Tatrzańskiego Parku Narodowego.* Archiwum Fotogrametrii, Kartografii i Teledetekcji, Vol. 17b: 463-472.

Litwin U., Szewczyk R. (2012). *Morfologia działek przyczynkiem kształtowania krajobrazu.* Infrastruktura i Ekologia Terenów Wiejskich 2/II: 39–48.

Malczewski J. (1999). *GIS and multicriteria decision analysis*. John Wiley & Sons.

Mašíček T., Kozlovsky Dufková J., Stejskal B. (2016). *Analysis of erosion vulnerability of agricultural land in the catchment of Luh stream using the GIS tools*. Infrastruktura i Ekologia Terenów Wiejskich 4: 1789–1810. DOI: 10.14597/infraeco.2016.4.4.134

Rosca B., Chelaru D.A., Plescan S. (2013). *The Analysis Of Landscape Morphology In Lower Bistrita Valey Using Grass And Quantum GIS*. International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management, 1: 951.

Theobald D.M. (2007). *GIS concepts and ArcGIS methods*. Conservation Planning Technologies.

Wheatley D. (1995). *Cumulative viewshed analysis: a GIS-based method for investigating intervisibility, and its archaeological application.* Archaeology and GIS: A European Perspective: 171-186.

Wong W.S.D., Lee J. (2005). *Statistical analysis of geographic information with ArcView GIS and ArcGIS*. Wiley.

Eng. Piotr Benduch, MSc
AGH University of Science and Technology
Faculty of Mining Surveying and Environmental Engineering
Department of Geomatics
Al. A. Mickiewicza 30
30-059 Kraków
Tel: +48 603782700
E-mail: piotr.benduch@agh.edu.pl

Received: 10.12.2017
Accepted: 29.12.2017