Solving a surveying problem by using R and QGIS - Setting out of a land expropriation zone

Milutin Pejović[1*], Zagorka Gospavić[1], Branko Milovanović[1], Igor Arsić[1]

[1]University of Belgrade, Faculty of Civil Engineering, Belgrade, Serbia

Abstract. Application of the open source software when resolving geodetic issues provides limitless possibilities. The present paper will showcase a methodology for setting out of engineering projects through a combination of GIS (Geographic Information System) software QGIS and R software environment. The link between QGIS and R had been established using Sextante toolbox, which significantly extends possibilities for a QGIS application, by establishing link to the other "powerful" tools, such as R. Test example of land expropriation project shows the possibilities of the QGIS application for analytical elaboration of an expropriation zone, and R software package for automating setting out accuracy calculation.

Keywords: QGIS, R, Setting Out, Expropriation

---

[*] Milutin Pejović> mpejovic@grf.bg.ac.rs

1 Introduction

Setting out of an engineering project is one of the most common tasks sets for land surveyors. According to Prof. Asanin [1], setting out of the engineering project is the process of building positioning in the space, within the boundaries of thresholds set for a construction or an assembly. The process of staking out assumes several steps, and each step needs to be carefully planned. First, examining of the project complexity and defining simpler elements of the project (points, lines, etc), which could be set out. This process is called the analytical elaboration. The choice of the methods and approach are often caused by field conditions. Defined construction tolerances impose estimation, which will provide all technical conditions necessary for setting out, within the thresholds prescribed. The entire set represents the setting out of the engineering project. Using the modern software, in the open source field, has no limitations and provides for automated setting out and adjustment to the specific projects. Methodology shown in this paper relies on the application of two open source software solutions: QGIS – GIS software in analytical elaboration process; and R software environment for automated accuracy calculation. Combination of the two software packages is not chosen at random. QGIS is the GIS software, which provides all necessary functionalities for spatial phenomena analysis. Apart from that, due to its functionality and intuitive design, it is gaining popularity among users. R is an integrated software environment for statistical calculations, programming and graphical presentation [14]. R is also the environment with daily growth in population, due to which it is being rapidly developed. Thanks to the modules for spatial data handling development, R is a suitable extension of GIS software functionalities. Good interaction between the two software packages is provided through a very beneficial plug-in *Sextante*, which provides for access and control over all R functionalities from QGIS. In practice, *Sextante* provides for creating proprietary tools in the QGIS environment, by using R functionalities. However, producing proprietary tool within the QGIS using R functionalities requires programming knowledge in the R environment. There is a multitude of beneficial references today, to be recommended to new users – [4] and [7]. The mentioned methodology of setting out of the engineering project is to be presented through the example of the expropriation zone setting out. The present paper is aimed at promoting application and possibility of integration of the software packages above, in order to extend the open source users community, apart from presenting the methodology of setting out of the engineering project.

2 QGIS Software

QGIS is the open source software, computer GIS application providing viewing, editing and analysis of the spatial data. Gary Sherman had started QGIS development early in 2002, as the software for GIS data viewing for Linux. Today, it can be used on various operative systems, including Mac OS X, Linux, UNIX, and Microsoft Windows. QGIS provides for integration of extensions developed using C++ or Python. Some of the important, well-known QGIS extensions, significantly increasing its usability, already developed include GDAL, GRASS GIS, PostGIS and PostgreSQL. Quantum GIS is currently being maintained by the group of volunteers, regularly releasing upgrades and fixes.

Just like all other GIS software systems, QGIS enables users to create multi-layer maps, using various cartographic projections. The map can be used for various purposes and saved in different formats, in the raster or vector format. The vector data can take the form of a point, a poly-line or a polygon, while the raster data need to be geo-referenced beforehand and subsequently saved in the different formats (JPG, TIFF, and BMP). At the official website[1], there are user manuals for the application available, accompanied by examples, exercises, pictures and the data that can be useful when practicing, and the link for desktop application download. Apart from the official QGIS webpage, there is a vast number of websites aimed at helping the application users, where experts publish their findings, experiences or answer to users' questions.

2.1 QGIS Plug-ins

QGIS is designed with a plug-in architecture. This provides for easier adding of new functions or tools in the applications. Numerous add-ons had already been implemented in the program, with a vast online archive constantly expanding, with the possibility to add new plug-ins by the professional programmers:

- Basic plug-ins – maintained by the QGIS development team, being a constitutive part of every QGIS distribution;
- Additional plug-ins – currently all written in

---

[1] www.qgis.org

Python. Available on the official website[2] to be added using "Plug-in Installer".

The following plug-ins can be listed as the most beneficial and the most prospective:

- "Sextante" – to be elaborated in detail below;
- "Spatial Query Plug-in" – used for the location-based selection of the features on the map;
- "Open Layers Plug-in" – providing for addition of various contents available over well-known services, such as Google, Yahoo, Bing...

2.2 "On-The-Fly Projection"

There is a common situation that one project uses layers in raster or vector format with different projection, with the consequence of those being shown at different spots on the map, although they refer to the same location. To resolve this issue, QGIS introduces "On-The-Fly Projection" function, which shows the layers with the different projections in the projection previously selected by the user "on the fly". The only prerequisite for this operation is having coordinate systems clearly defined, for the source and the target system to be used.

3 R

R is a platform-independent, object-oriented interpreter and an interactive programming language, predominantly dedicated to statistical calculations, along with programming and the graphical presentation. R is free and open source software that comes under GNU[3] general public license. In its basic package, R has a wide range of predefined functions for statistical analysis and data modelling. The article [11] provides the best description about popularity and expansion of R as the software tool. R owes its vast popularity to the modular system of organization, i.e. package system under which the functions packages dedicated to the particular thematic are being created by the users themselves [11]. The programming language implemented in R is similar to "S" programming language [4].

R had become particularly interested for users in the field of geo-sciences, when the developers had created packages for handling spatial data [9]. Those packages contain classes and methods for definition of spatial data model, structure and organization [3]. Overview of packages dedicated to the spatial data handling is shown at the webpage[4]. This paper uses some of the most popular packages for spatial data handling – SP and rgdal. rgdal is the implementation of well known GDAL (Geospatial Data Abstraction Library), dedicated to handling (loading and creating) GIS data (formats), with the possibility to handle Coordinate Reference Systems (CRS).

4 SEXTANTE

Sextante is the software environment for spatial data processing, having its full affirmation as the QGIS extension, turning QGIS into powerful GIS tool. Apart from the algorithms implemented in the software itself, Sextante provides for integration of some of the most familiar GIS tools (such as GRASS, SAGA, ORFEO Toolbox, R) under QGIS environment, facilitating their use. Initially, Sextante had been developed at the Faculty of Forestry of Extremadura University (UNEX) in Spain, by the main author Victor Olaya. Sextante is currently a part of QGIS, being developed by the QGIS community as such. Interesting examples of Sextante Plug-in applications are available at author's blog[5], as well as the official QGIS webpage[6].
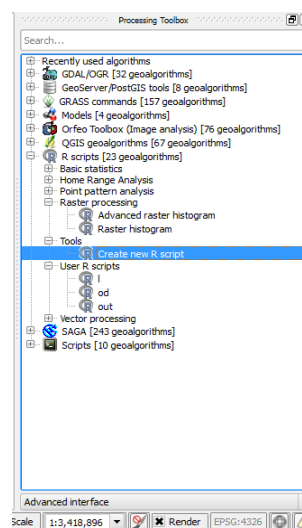
5 R SCRIPT IMPLEMENTATION IN QGIS

Figure 1: (a) Processing toolbox

[2]https://plugins.qgis.org/plugins/
[3]http://www.gnu.org/licenses/gpl-2.0.html

[4] http://cran.r-project.org/web/views/Spatial.html
[5] http://qgissextante.blogspot.de/
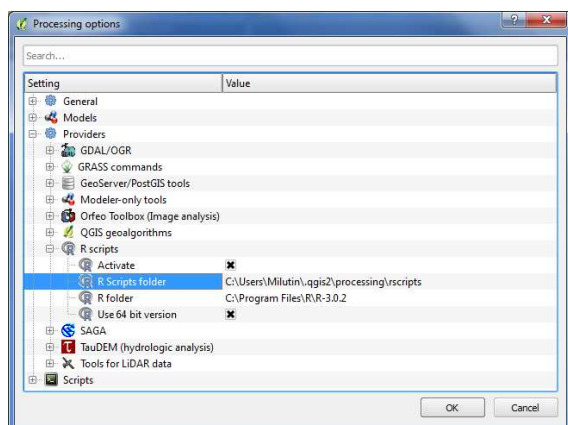[6]http://qgis.org/en/docs/user_manual/processing/commander.html?highlight=sextante

Figure 1: (b) Setting path to R-scripts

R script implementation in QGIS involves several steps. First of all, a file with .rsx extension for storing the script is being created. File is created by double clicking "Create new R script" option in "Processing Toolbox". A file path must also be set, by clicking "Processing" option, followed by "Options and configuration" (Figure 1b).

For the R script to be readable in QGIS, it needs to be structured into two parts:

– Configuration part – where input and output parameters are defined and
– Executable part (code) – containing functions and commands to be executed.

The configuration part contains definitions of input and output parameters' names and types. Input and output parameters are defined using "Python" notation, with two "#" at the beginning of the line. Input and output parameters can be vector, raster, tabular or graphical (only for the output parameters). Apart from the input and output parameters, the configuration part also sets optional parameters, defined by the very function. It is usual to define the group of functions to which the function created belongs (e.g. "##My functions = group"). Table 1 provides short overview of the input and output parameters' definition.

Table 1: Script input and output parameters definition

| ##grupa=group	Setting up group (groups) to which the function belongs
##ulaz=vector	Setting up input layer
##field=field layer	Setting up attributes from a certain layer
##showplots	Setting up graphical presentation option (output parameter)
##izlaz=output vector	Setting up output presentation ("output" vector in this case)
##opcija1=string	Optional parameter (textual)
##opcija2=number	Optional parameter (numerical)
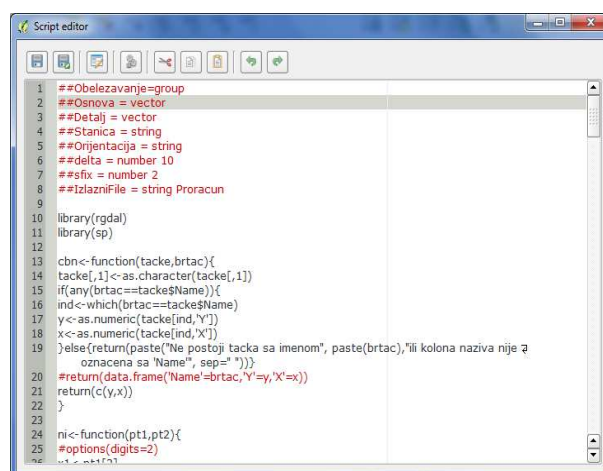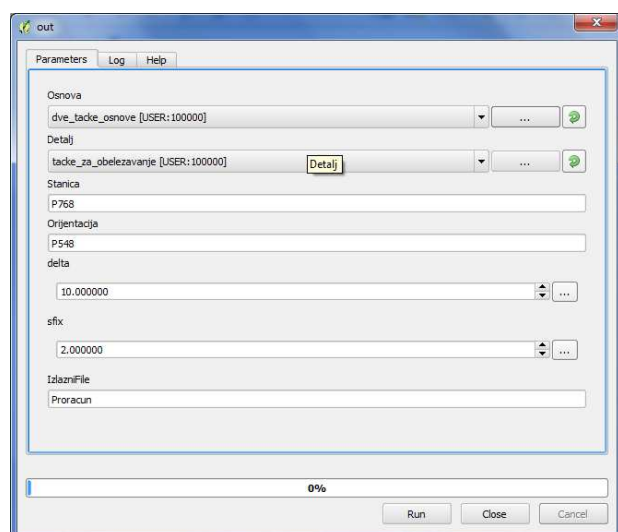
Figure 2: (left) Dialog of R script initiated; (right) Example of script written in Script Editor

The part reserved for code, with functions and commands to be executed, is being filled out with R script. It only takes the following logical sequence of defined functions' execution (loading package – loading functions created – executing commands over data – creating results in desired format). An example of R script written in QGIS "Script Editor" is shown in Figure 2 (right).

When the created function is being initiated (double click), a dialog opens with optional fields defined in the function itself. Creating dialog is an automated routine, making the function created more desirable for wider range of users. After defining optional (input, output and optional parameters), clicking on "Run" executes the function.

6 Methodology and Results

Proposed methodology will be presented using the test example of setting out of engineering project for expropriation zone using the polar method. Under the assumption that Digital Cadastral Map had not been produced for the test area, the task entails setting out of the engineering project for land expropriation purposes for road building, according to a scanned cadastral map and the defined expropriation zone. An algorithm for setting out of the project is shown in the Figure 3.
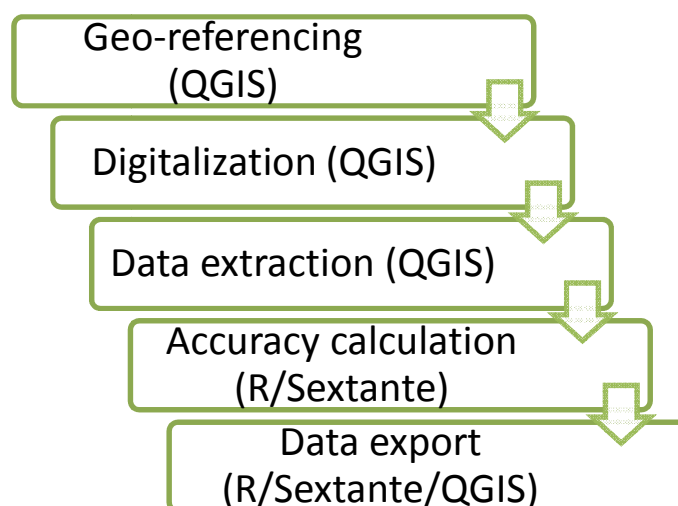
Geo-referencing
(QGIS)

Digitalization (QGIS)

Data extraction (QGIS)

Accuracy calculation
(R/Sextante)

Data export
(R/Sextante/QGIS)

Figure 3: Project setting out methodology

Geo-referencing is usually the first step in raster data handling. Spatial reference determination is the main prerequisite for using the information entailed in the raster. Geo-referencing is usually been performed in two steps. Assigning the coordinates to the raster points and determination of the planimetric transformation parameters, followed by assigning the parameters of the Spatial Reference System to the transformed raster. In QGIS, geo-referencing is being performed by selecting "*Georeferencer*" option from "*Raster*" menu. This opens "*Georeferencer*" dialogue, which provides the option "*Transformation Setting*" for choice of target SRS (Spatial Reference System). "*Target SRS*" is predefined through the parameters of the state coordinate system, which practically defines the coordinate system of the entire project. The coordinate system is being defined by selecting "Custom CRS" option in "Settings" menu. The

coordinate system used in the present paper is defined by the following parameters, written in the notation[7]:

```
+proj=tmerc +lat_0=0 +lon_0=21 +k=0.9999
+x_0=7500000 +y_0=0 +ellps=bessel
+towgs84=574.027,170.175,401.545,4.88786,-
0.66524,-13.24673,0.99999311067 +units=m
+no_defs
```

The verification of properly done geo-referencing is being performed through the control of the residual values obtained.

Digitalization of the scanned cadastral map was necessary for several reasons. First of all, for the purposes of easier implementation of an elementary GIS operation, with the objective of extracting information of interest. This firstly refers to extraction of areas, points, parcel fronts, etc. Selecting "*New Vector Layer*" option in "*Layer*" menu opens dialog

---

[7] http://trac.osgeo.org/proj/

(New Vector Layer), to create new vector layer. In this dialog, there is an option to set the layer type (point, poly-line or polygon), assign the coordinate system and define the layer attributes. In this case, a new polygon layer had been defined, for storage of the digitized parcel. That layer was assigned with the same coordinate system as digitized raster.

After digitalization, necessary data had for setting out of the engineering project for the expropriation zone had been extracted using several linked GIS operations.

– By overlapping and cutting the parcel layers and the expropriation zone using "*Clip*" function, the new layer is obtained, holding all information from the previous layer pertaining to the areas overlapped. The new information refers to the area of the parcel part that "falls" inside the expropriation zone.

– Extracting the parcel points and the expropriation zone points was done using "Extract Nodes" function. For the purposes of the detailed analysis, the points were assigned with the attributes, for "x" and "y" coordinates values, using "Add Coordinates to Points" functions. After extracting the points, duplicated obtained points are noticeable, which can be joint into the single points using "Remove Duplicate Points" function, which provides the final layer of the expropriation zone boundary points shown in the attribute table (Figure 4).

Figure 4: Expropriation zone boundary points layer

– Frontal lines between the points, as the necessary element for the control of setting out, had been extracted as the separate lines using "Polygon to Lines" and "Explode Lines" functions, which assign the attribute with the frontal lines length values (Figure 5):
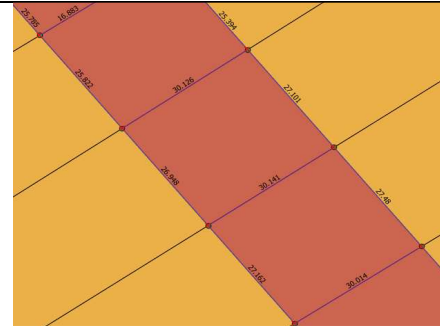

Figure 5: Parcel frontal lines

For the purpose of determination, i. e. technical conditions for the project implementation in line with the criteria required, functions for an automated accuracy calculations had been developed. The accuracy calculation may also be considered as an optimization process, which optimizes the budget of the errors present in the selected setting out method. Influences of certain error sources had been explained in detail in the book written by Prof. Perović [10]. For this purpose, polar points' demarcation method had been chosen. The accuracy calculation function had been prepared pursuant to [2].

By importing previously created R function in QGIS as explained above, a new function within the QGIS had been created, providing for accuracy calculation. The input parameters of this function are:

– Feature points layer (the points for which the demarcation calculation is being performed);
– Geodetic base points layer;
– Permitted discrepancies of the feature points;
– Standard deviation;
– Name of the point for demarcation and the point serving as orientation;
– Output layer name.

When the function is initiated, dialog (Figure 2 (left)) opens, providing for the input parameters entry. Based on the inter-relations between the points and the measurement model including the influences of the individual sources of the errors, maximum permissible values of influence by the individual sources of the errors are being calculated. First of all, the influence of the angle and the distance measurement followed by the influences of the individual sources of the errors, such as an instrument precision influence, the error of instrument and prisms centring and the influence of the predefined values' errors (accuracy of the network points). The output layer is identical to the input layer of the demarcation points, with added information on the previously

noted sources of the errors' influence at each point. Apart from that, the demarcation elements (angle and distance) are also included in the output file. Afterwards, based on the error influencers multitude (for each point separately), the most severe influences are being extracted, as the values meeting all of the requirements. Pursuant to these values, the conclusion on approach during setting out is being chosen, along with the demarcation control measures.

After finishing calculations, all information are easy to **export** to any file using simple write to file operation. Creating that kind of a function, which would single out the important information, relies on "sp" package functionality and the spatial data structure. A simple example of the function that executes summary overview of one layer attributes and stores them in .csv file is given in the Fig. 6:
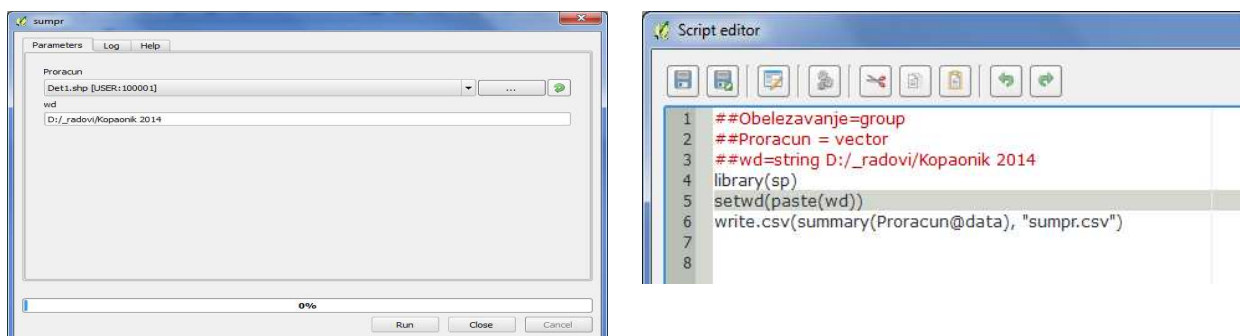

Figure 6: Example of simple function for data summarizing in a single layer

7 Conclusions

Short experiences of the authors regarding R software application within QGIS framework were more than positive. Easy implementation of R functionalities in QGIS had help expanding the vast QGIS users' community with even vaster R users' community. This symbiosis is expected to provide numerous interesting solutions yet to occur. Using the numerous ready-made function packages provides for easy R script development to automate resolving a problem, while QGIS application easily and quickly creates the graphical presentation or implements the GIS analysis. In the present paper, a simple setting out of the engineering project for the expropriation area example had been used to easily present the advantages of the simultaneous R and QGIS usage. The primary intention of the authors was to inform and inspire through this example. Plans for the future work include software development and finding the solutions for other problems in the field of engineering geodesy using R and QGIS, along with the inclusion of these software packages in the lecturing activities at the faculty.

References

[1] Ašanin, S.: „Inženjerska geodezija". Belgrade: Ageo d.o.o.. 2003.

[2] Ašanin, S., Pandžić, S., Gospavić, Z., Milovanović, B.: „Zbirka odabranih zadataka iz Inženjerske geodezije",

Belgrade: Geokarta d.o.o.. 2007.

[3] Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V.: „Applied Spatial Data Analysis with R". Springer. 2008.

[4] Chambers, J. M.: „Software for Data Analysis: Programming with R (Statistics and Computing) ". Springer. 2009.

[5] Ghilani, D.; Wolf, P.: „Adjustment computations - Spatial Data Analysis", John Wiley & Sons. 2006.

[6] Gray, J.: „Getting Started With Quantum GIS". Retrieved from http://www.linuxjournal.com/content/getting-started-quantum-gis, 2008.

[7] Kabacoff, R. I.: „R in Action". Manning Publications Co. 2011.

[8] Olaya, V.: „The SEXTANTE framework". Retrieved from http://sextantegis.com/, 2013.

[9] Pebesma, E. J., Bivand, R. S.: „Classes and methods for spatial data in R". R News, No. 5, pp. 9–13, 2005.

[10] Perović, G.: „Precizna geodetska merenja". Belgrade: University of Belgrade, Faculty of Civil Engineering. 2007.

[11] Smith, D.: „Seven quick facts about R". Retrieved from http://www.r-bloggers.com/, 2014.

[12] R Development Core Team: „R: A language and environment for statistical computing". R Foundation for Statistical Computing. 2008.

[13] Rebelious, S.: „Getting raster histogram in QGIS using SEXTANTE and R". 2013.

[14] Venables, W.N., Smith, D.: „Introduction to R". R Development Core Team. 2010.