

## STATISTICAL ANALYSIS OF A DIGITAL ELEVATION MODEL USING ARCGIS

Timea GERGELY<sup>1</sup>, Georgiana OPRITA<sup>1</sup>, Grațiela PASCAL<sup>1</sup>

Scientific Coordinators: Assoc. Prof. PhD. Florica MATEI<sup>1</sup>, Lect. PhD. Eng. Tudor SALAGEAN<sup>1</sup>

<sup>1</sup>University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, 3-5 Calea Manastur, 400372, Cluj-Napoca, Romania, Phone: + 40264596384, Fax: + 40264593792

Corresponding author email: gratielapascal94@gmail.com

### Abstract

*The aim of the paper is to eliminate the redundant data included in a geographical information system (GIS) involving surface analysis. Because the GIS applications are used to solve problems from miscellaneous domains, is recommended that the spatial information and the particularities of the area involved to be combined in one raster layer. Principal component analysis applied to raster layers gives such a facility. The studied surface is about 2.5hectares and is a part of the campus of University of Agricultural Science and Veterinary Medicine from Cluj-Napoca. The digital elevation model for the studied area will be generated and the slope and aspect raster layers derived from DEM will be combined using the principal component analysis in order to eliminate the redundant values from the new raster but to keep the particularities of the surface.*

**Key words:** digital elevation model, principal component analysis, ArcGIS.

### INTRODUCTION

The ability to model the Earth surface is very important in order to understand the phenomena derived from environmental science (Griffith & Peres-Neto, 2006), (Verfaillie, Du Four, Van Miervenue, & Van Lancker, 2009) and to provide a more sustainable management for environmental resources as well as to understand complex life science problems, (Abson, Dougill, & Stringer, 2012). The common point of all these, is the importance of the spatial references of the data in order to analyze the mentioned problems; in most of the cases the elevation is one of the factor that will influence the studied phenomena. Spatial references will be included in the studied model through digital elevation model (DEM) that is a gridded array of elevations. (Mihai, Radu, & Cazanescu, 2009). No matter what software will generate the DEM, there will be involved a large amount of data. Most of the analysis performed on DEM will be based and will include the elevation. From this point of view it appears the problem of data redundancy from raster images derived from DEM. The slope and aspect for the surface use the elevation as primary source.

The most suitable way to manage large amount of data is to use a statistical analysis of them and the statistical method used to eliminate redundant value is principal component analysis (PCA) (Jolliffe, 2002), (Demser, Harris, Brunson, & Stewart Fotheringham, 2013).

In this paper the DEM will be generated from points obtained from survey measurements using ArcGIS software and the raster images provided by the Spatial Analyst tools for surface analysis. These will be studied using PCA in order to eliminate the data redundancy due to the fact that all raster images are based on elevation.

### MATERIALS AND METHODS

The approach proposed by this paper combines the spatial aspect of data with statistical analyses of those data in order to obtain a geographical information system that eliminates the redundancy of data. The software application used for this is ArcGIS 10. The spatial aspect of data is included in DEM and the raster and statistical analyses are made using Spatial Analyst extension of ArcGIS.

The studied surface is included in the University of Agricultural Sciences and Veterinary Medicine from Cluj-Napoca campus. In order to represent our terrain in ArcGIS we took measurements using the GNSS technology with GPS receiver Trimble R10, which has a precision of  $\pm 8\text{mm} + 1\text{ppm}$  in horizontal plane and  $\pm 15\text{mm} + 1\text{ppm}$  in vertical plane, using the kinematic method. In this way we obtained the coordinates of the points used to create the digital elevation model (DEM). Using the ArcGIS software, we created the geodatabase `usamv.gdb` in which is included a feature dataset where we imported our coordinates and also created other feature classes for details like roads and polygons that delimited the studied area as well as the building. Using the above elements was created the DEM for the investigated surface and were generated raster images for the slope and the aspect of the land using the 3D Analyst toolbar (Price, 2014). Furthermore, we used the PCA to compress data by eliminating redundancy. This statistical technique is widely used in the social and physical sciences to identify the spatial dominant patterns that will be the principal component and all other spatial data will be reconstructed using the principal component. (Demsar, Harris, Brunson, & Stewart Fotheringham, 2013).

The main idea of PCA is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set, (Jolliffe, 2002). This will be obtained by transforming to a new set of variables, named principal components (PCs), which are uncorrelated and which are ordered so that the first few retain most of the variation present in all of the original variables.

Applying this method to raster, PCA from the input multivariate attribute space transforms the data into a new space in which those axes are rotated respect the original space. In our case, the raster images involved are elevation, slope and aspect. Since slope and aspect are usually derived from elevation, most of the variance within the study area can be explained with elevation. All the procedure is made in ArcGIS software using Multivariate Tools.

## RESULTS AND DISCUSSIONS

As a first result we obtained the TIN Layer (Figure 1), that was created from our points, based on terrain tools. Every node is joined with its nearest neighbors by edges to form triangles. One or more polygons, which contain the entire set of data points, form the hull of the TIN. The hull polygons define the zone of interpolation of the TIN. As a result, the TIN generator creates a convex hull to define the bounding edges of the TIN.



Figure 1. TIN layer

Based on the TIN Layer we created the Raster Image (Figure 2), of the terrain, represented in the image below. The darker partitions show higher elevations and the brighter ones show lower elevations. We can notice that the elevation varies between 48.8 m and 402.4 m.

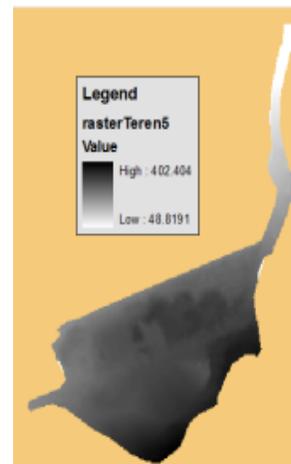


Figure 2. Raster Image

Using the slope tool implemented in surface analysis it was obtained a raster image, Slope

Layer illustrated in (Figure 3). Analyzing this we can observe the slope inclination of the surface. The inclination varies between 0.19% and 89.45%. However after reclassifying the slope raster, the mean value is 20.1 and the standard deviation is 20.32, (Figure 6.b).

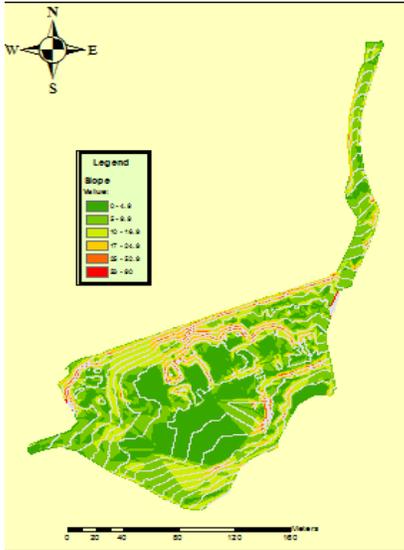


Figure 3. Slope

From elevation raster was created the Aspect Layer (Figure 4) of the surface, in order to identify the slope direction or the compass direction of hill faces. In our case, we can observe that the north-facing and northwest-facing slopes are predominant, in comparison with the southeast-facing slopes which are less visible on the map.

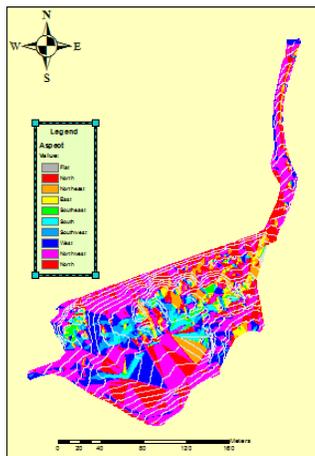


Figure 4. Aspect

For a better view of the surface we generated the contours (Figure 5), lines that connect

points of equal height (Z). The distribution of the lines shows how values change across this surface: where there is little change in a value, the lines are spaced farther apart; where the values rise or fall rapidly, the lines are closer together.

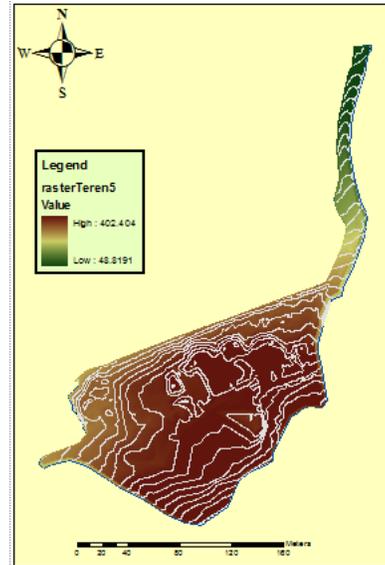


Figure 5. Contours

Using the Band Collection Statistics implemented in Multivariate tools of Spatial Analyst in the next figure (Figure 6) emphasize the variance involved in each generated raster.

(a)

bandcol_teren5_2					
# STATISTICS of INDIVIDUAL LAYERS					
#	Layer	MIN	MAX	MEAN	STD
#	1	26.4721	401.8434	391.0721	12.2930
#	=====				

(b)

bandcol_slope_t1					
# STATISTICS of INDIVIDUAL LAYERS					
#	Layer	MIN	MAX	MEAN	STD
#	1	0.1857	89.4496	20.0986	20.3184
#	=====				

(c)

bandcol_aspect_1					
# STATISTICS of INDIVIDUAL LAYERS					
#	Layer	MIN	MAX	MEAN	STD
#	1	0.0634	359.9373	237.8685	108.8054
#	=====				

Figure 6. (a) statistics for elevation (b) statistics for slope (c) statistics for aspect

Applying PCA to elevation, aspect and slope raster was obtained the raster image given in (Figure 7) which does not contain redundant

data but keeps all the characteristics of the patterned surface.

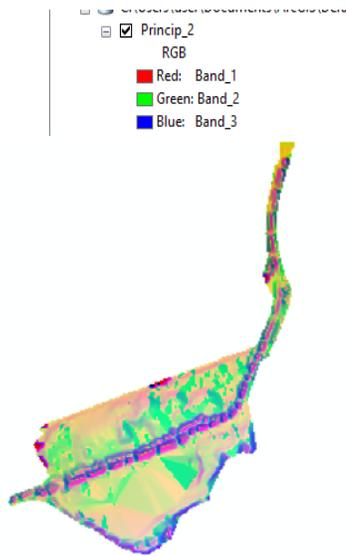


Figure 7. Investigated surface after the PCA

Analyzing the output data file obtained after the PCA implementation results that the cumulative value of the variance is given by the first layer of the multiband raster obtained from PCA in 95%. (Figure 8).

# EIGENVALUES AND EIGENVECTORS			
# Number of Input Layers	Number of Principal Component Layers		
3	3		
# PC Layer	1	2	3
# Eigenvalues			
	3090.53781	110.41004	33.85117
# Eigenvectors			
# Input Layer			
1	0.99953	-0.03037	0.00509
2	-0.03062	-0.96272	0.26874
3	0.00326	0.26877	0.96320
# PERCENT AND ACCUMULATIVE EIGENVALUES			
# PC Layer	EigenValue	Percent of EigenValues	Accumulative of EigenValues
1	3090.53781	95.5403	95.5403
2	110.41004	3.4132	98.9535
3	33.85117	1.0465	100.0000

Figure 8. Output file for PCA

## CONCLUSIONS

When PCA is applied to raster data the principal component are calculated for a data set where data are raster's cells. The result is a new raster where for each cells is assign a value based on new principal components space. The new raster image is suitable to be used in subsequent spatial analysis based on map algebra such as combining raster from meteorology with soil type or census data. In all of this cases the amount of data is considerable large and it is important to work with one layer that includes the characteristics and particularities of all layers considered input data for PCA. In the case considered for this paper the investigated surface has 2.5 hectares and the initial number of points is 1631.

## REFERENCES

- Abson, D., Dougill, A., & Stringer, L. (2012). Using Principal Component Analysis for information-rich socio-ecological. *Applied Geography*, 515-524.
- Demsar, U., Harris, P., Brunson, C., & Stewart Fotheringham, M. S. (2013). Principal component analysis on spatial data: an Overview. *Annals of the Association of American Geographers*, 106-128.
- Griffith, D., & Peres-Neto, P. (2006). Spatial modelling in ecology: The flexibility of eigenfunction spatial analysis. *Ecology*, 2063-2016.
- Jolliffe, I. (2002). *Principal Component Analysis 2nd ed.* Berlin: Springer Verlag.
- Mihai, D., Radu, M., & Cazanescu, S. (2009). Highlighting, Rehabilitation and Monitoring of Land Reclamation Works. *Bulletin UASVM*, 596-602.
- Price, M. (2014). *Mastering ArcGIS Sixth Edition.* New York: McGraw Hill.
- Verfaillie, E., Du Four, I., Van Miervenne, M., & Van Lancker, V. (2009). Geostatistical Modelling of sedimentological parameters using multiscale terrain variables: Application along the Belgian part of the North Sea. *International Journal of Geographical Information Sciences*, 135-185.