Raport Badawczy

Research Report

RB/17/2017

Multi-criteria classification of locations on the map with applications in archaeology

R. Bieńkowski, K. Leśniewski, W. Radziszewska

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. dr hab. inż. Antoni Żochowski

Warszawa 2016

IMEKO International Conference on
Metrology for Archaeology and Cultural Heritage
Lecce, Italy, October 23-25, 2017

Multi-criteria classification of locations on the map with applications in archaeology.

Rafał Bieńkowski[12], Krzysztof Leśniewski[13], Weronika Radziszewska[1]

[1] *Systems Research Institute PAS, ul. Newelska 6 01-447 Warszawa Poland, Rafal.Bienkowski@ibspan.waw.pl, Krzysztof.Lesniewski@ibspan.waw.pl, Weronika.Radziszewska@ibspan.waw.pl*
[2] *Institute of Archaeology and Ethnology PAS, Al. Solidarności 105, 00-140 Warsaw, Poland*
[3] *Warsaw University of Technology, Faculty of Mathematics and Information Science, Pl. Politechniki 1, 00-661 Warsaw, Poland*

***Abstract* – The computer multi-criteria classification of the locations on the map can help in the predefining the areas for archaeological survey. The assumption is that settlements were located considering a number of rational criteria. The criteria considered in this paper are geographical criteria: the walking distance from drinkable water and arable areas. The method calculates the areas that are in reachable distance from each criteria then combines the data using Simple Additive Weighting method. The method was tested on real known location settlement from Venetian period in south coast Crete.**

I. INTRODUCTION

Archaeological survey is a time consuming and labor-intensive task of searching and mapping of archaeological remains of the ground surface. One of the types of remains are traces of settlements. Predefining the interesting areas can lead to faster finding of the site of the interest. Up to now the areas to be searched were defined by archaeologists who, using their specialized knowledge, could evaluate how likely it is to find some remains of an archaeological site in the area. Although the specialized knowledge will be always necessary there might be possibility to partially automate the initial selection of areas.

The settlements are up till now created with some reasoning, like distance from other city, access to roads, climate, etc. Recognition of such criteria can help with determining what was the function of settlements and their location. There are two main types of such factors: the geographical and the anthropological. The anthropological factors are much more influential but very difficult to define, they are factors such as distance from the nearest settlement, trade route or defensible location. The geographical factors are the ones connected to the location and terrain, as e.g. short (walking) distance from water source or arable areas, difficulty of the terrain. Other of such factors is the proximity to the drinking water – the people would generally consider living near the rivers, but not too near to them if the river regularly floods the area. The arable land is the source of food, but creating the settlement on the most fertile land is wasting of farming potential. In this work we focus on geographical factors that favour the establishment of a settlement. The considered area is Crete, and the settlements that are in our example are from Venetian period. Crete is rich in archaeological material and the settlements were established not too far from each other. That gives a good test case for our algorithm and the data source for determining the parameters in our method. The next section is presenting the method of calculating the interesting areas considering each criteria. In section III the approaches to combine those data are presented. Final section concludes the paper.

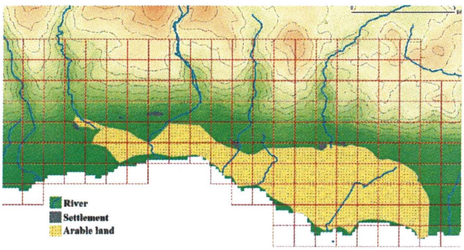

Fig 1. The sample of the data - south-west coast of Crete, Greece, in yellow are marked arable lands, grey areas indicate the known settlements, blue colour represents the rivers, different shades of green and beige represent the height over the sea level.

II. THE MULTICRITERIA CLASSIFICATION

The method calculates the cost of walking from the defined resource to each other point reachable on foot. Distance is represented by the cost of walking in the terrain. The cost is defined by the topological relief function that represents the difficulty of the terrain. The difficulty is calculated from the raster representing the height above sea level, by first calculating the normalized average if the height over some area z_i (in our case 10x10 cells), and then each cell of the topological relief is calculated: $s(x_i) := \sum |p_i^k - z_i|$, where p_i^k is the height above sea level and x_i is the aggregated cell of the raster.

The algorithm for calculating the cost of reaching each point of the map is based on Bellman-Ford algorithm [2] is used. This algorithms starts at a point with the required resource, then calculates the cost of accessing all of the neighbouring cells, then neighbours of those cell and so on. When it finds the cheaper route to the cell it recalculates again the costs and the paths ensuring that the cheapest route is found. It calculates till it reaches the threshold cost that is representing the distance that is too far or/and too difficult to access regularly on daily basis. The cost is calculated for every cell of the raster map and for every criteria separately. The areas unreachable on foot are excluded. The final classification of points on the map is done by calculating the weighted sum of the costs for all of the criteria.

The calculated costs for each criterion are presented: in Fig. 2 for proximity to arable land and in Fig. 3. for proximity to sweet water.

III. SIMPLE ADDITIVE WEIGHTING

We consider weighted function to combine the values of a decision-making options for given criteria. The weights for the additive weighting methods are determined using the real example. The considered example is a real map of remains of settlements and archaeological sites in the area of south-west coast of Crete, Greece (presented in Fig. 4 in black). To calculate weights the two different methods were used: the simple calculation of the sum of the costs and the method of auxiliary vectors.

Our aim to calculate the function that indicates how suitable the area is for establishing the settlement, which means where archaeologists should focus their search:

$f(x_i) = w_1 f_1(x_1) + \ldots + w_d f_d(x_i)$,

where $f_i(x_i)$ is a value of a decision-making option for the given criterion, w_i are weights such that $\sum_{i=1}^{d} w_i = 1$, d represents the number of different criteria. We want to classify points on the map to get the settlement points using function f. Of course weights may be given by decision maker but we want to estimate them by using known locations of the settlements (Fig. 2. and Fig. 3.). The problem is that for the given data and the given values of function f the solution set for w_i may be empty.

Let us assume that we have some prepared data with information about localization i.e. we have points $o_1, o_2, \ldots, o_p$ which are the settlements points in Fig. 1. For each point o_i, $i = 1, 2, \ldots, p$ we have information about value of each criterion f_i, $i = 1, 2, \ldots, d$. For example in Fig. 2. and Fig. 3. we can see the values for f_1 and f_2, respectively.

The first method to quickly approximate the weights was to calculate the average distance and maximum distance of the known settlements from the resources as sweet water and arable land. It is done by calculating weighted function for w_1 from 0.05 to 0.95 with interval of 0.05. Then for the outcome weighted grid the average (AVG) and maximum (MAX) distance is calculated:

$$AVG = \frac{\sum_{i=1}^{p}\sum_{j=1}^{d} f_j(o_i)}{p}, \quad MAX = max(f_j(o_i)).$$

The results of such approximation are presented in Fig. 4. The best result give the weight of $w_1 = 0.2$ and $w_2 = 0.8$. This solution is presented in Fig. 5

Such calculation of weights is very simple, so the method of auxiliary vectors was used. Let us calculate auxiliary vectors $W_1, W_2, \ldots, W_d$ in the following way

$$W_1 = [\frac{f_1(o_1)}{\sum_{i=1}^{d} f_i(o_1)}, \ \ldots, \ \frac{f_d(o_1)}{\sum_{i=1}^{d} f_i(o_1)}]$$

$$\ldots$$

$$W_d = [\frac{f_1(o_d)}{\sum_{i=1}^{d} f_i(o_d)}, \ \ldots, \ \frac{f_d(o_d)}{\sum_{i=1}^{d} f_i(o_d)}].$$

Let us denote W^l_k as l-th element from the vector W_k. We can define our weights in the following way

$$w_i = \frac{\sum_{k=1}^{d} W_k^i}{\sum_{i=1}^{p}\sum_{k=1}^{d} W_k^i}, \quad i = 1, 2, \ldots, d.$$

This method gave the values $w_1 = 0.543069$ and $w_2 = 0.456931$. This solution is presented in Fig. 6. The real difference between those weights are very small, but it seems that at least one settlement is better defined with the auxiliary vector method.

The values of the weighted cost give the raster defining the most interesting places from archaeological point of view.

IV. CONCLUSION

This approach creates a classification of the areas from the most promising for archaeological survey to ones where it is the least likely to find some interesting discoveries. The method does not include anthropological factors, which means some settlements might be outside of the preferred areas, e.g. a port. Method is calculating the cost of reaching the resources then using simple additive weighting method it combines the results for each criterion.

The method will be extended by adding different criteria and using expert knowledge, as e.g in [5,6].

REFERENCES

[1] E.Balestrieri, L.De Vito, S.Rapuano, D.Slepicka, "Estimating the uncertainty in the frequency domain characterization of digitizing waveform recorders", IEEE Trans. on Instrum. and Meas., vol.61, No.6, June 2012, pp.1613-1624.

[2] R. Bellman. On a Routing Problem. Quarterly of Applied Mathematics, 16:87{90, 1958.

[3] N.Björsell, P.Daponte, L.De Vito, S.Rapuano, "Automatic signal recognition for a flexible spectrum management," Proc. of 19th IMEKO World Congress, 2009, vol.1, pp. 444–449.

[4] S.Mallat, "A wavelet tour of Signal Processing. The Sparse Way", 3rd edition, Academic Press, New York, USA, 2009.

[5] J. Verstraete, "The Spatial Disaggregation Problem: Simulating Reasoning Using a Fuzzy Inference System," in *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 3, pp. 627-641, June 2017.

[6] J. Verstraete, "Solving the map overlay problem with a fuzzy approach", in Climatic Change, vol. 124, issue 3, pp 591–604, February 2014.

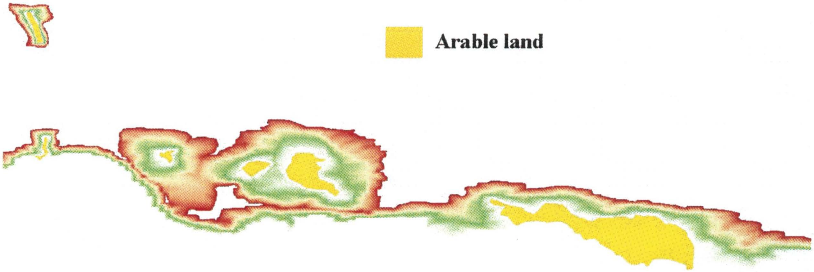

Fig. 2. The calculated cost (colours from green to red) of reaching the arable land (in yellow).

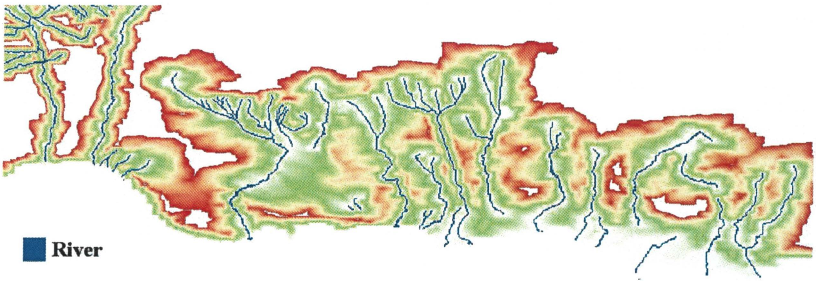

Fig. 3. The calculated cost (colours from green to red) of reaching the rivers (in blue).

Fig. 4. The simple approximation of the best weights, w1 is the weight for the arable land criterion, w2 for the proximity to the sweet water.

Fig. 5. The outcome for w1 = 0.2 (for arable land) and w2= 0.8 (water).

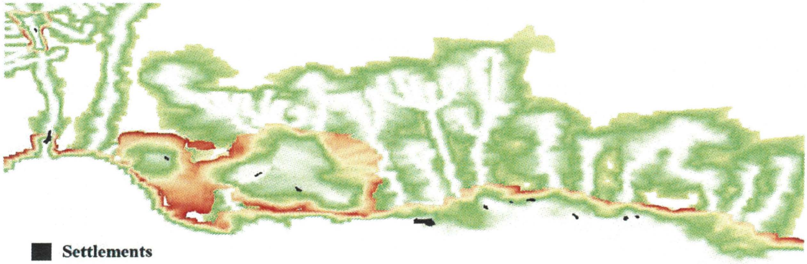

Fig. 6. The outcome for w1 = 0.55 (for arable land) and w2= 0.45 (water).