A comparison of morphometric and web prominence of mountain features

Ian Greatbatch, Jo Wood, Peter Fisher

Room A520
Information Science
School of Informatics
City University
London
EC1V OHB
Telephone: +44 20 7040 8381
Fax: +44 20 7040 8584
Email, iang jwo pff @soi.city.ac.uk

1. Introduction

The new field of ethnophysiography was first proposed at the COSIT conference in 2003 by David Mark and Andrew Turk (Smith and Mark, 2003). The argument put forward was that there should be an ethno-science for the earth sciences, as landscape is as important to some cultures as plants and animals. The problem is that the classic model of an ethnoscience typically consists of a comparison between an indigenous typology and a scientific one. In the case of landscape morphometry there is no clearly defined scientific standard to compare. It is also more difficult to collate an indigenous typology of landscape than of plants or animals, as landscape features are typically less well defined. In any British context it has become even more difficult as the concept of "indigenous" is very hard to define, and so a slightly different approach is required.

This work will investigate the relationship between cultural and physical prominence of mountain features, using the number of web pages a Google search reports as being relevant to a search term as the level of cultural prominence. There are a number other ways in which one could describe the cultural prominence of a peak and any value of prominence is likely to change throughout time and with different cultural groups. The groups typically involved in the creation of webpages are English speaking hill walkers and so it could be said that any measure of "culture" is limited to that or similar groups.

It is worth observing that this methodology would not be suitable for measuring any local cultural significance in a place with little or no internet access. It can only measure the global results as considered relevant by Google, although some filtering of results by IP address could be considered. Although could be said to have some drawbacks, one advantage is that language is not considered a hindrance in this case as any page containing the place name would be included regardless of the language of the page.

The data collected from the webmining experiment will then be compared to some measures of physical prominence derived from digital data.

2. Method

2.1 Landscape

The creation of ratios for prominence all involve the creation of an area that delimits the peak in some way, which in combination with its relative drop will give a value that will serve as a measure of physical prominence.

The peaks known as the “Wainwrights” were chosen, these being those described in Alfred Wainwrights’ series of books that comprise *A pictorial guide to the Lakeland fells*. The important thing about the Wainwrights, as far as this research is concerned, is that they are not confined by specific topographic measurements, but are entirely subjective.

The locations of the Wainwright summit peaks were downloaded onto a GPS for a hill walking website (Stevens, 2005). A 10m DEM of the area was downloaded from the Digimap site and mosaiced to cover the entire area. Using the *Landserf* software, a peak classification exercise was carried out on the surface which produced a series of peak contributing areas, altitude and relative drops for each peak. The names of each Wainwright point could then be joined to the peak contributing area that they fell within. This contributing area divided by the relative drop of the peak gives the first ratio for prominence, referred to in this work as the P ratio.

The second method of attributing areal values to peaks was to use inverse watersheds. Watersheds are defined as the contributing area to a drainage point, where any drop of water falling within a watershed will drain down to a single point. Watersheds in combination define draining systems and have also been used to effect as a system of delimiting zones within other surfaces, such as detecting contours in images (Beucher S and Lantuejoul C, 1979). If one were to invert a drainage surface, a watershed algorithm applied to a surface would produce contributing areas that “drain” upwards towards a peak and this was effectively what was done. The elevation surface was multiplied by -1 and then added to by 1000 (so that there are no negative integers) resulting in an inverse DEM. The areas created were divided by the relative drop of the peak and the resulting value is known as the W ratio.

The third method of ascribing area to peaks was to create Voronoi polygons around each summit point. The area of each Voronoi polygon divided by the relative drop is known as the V ratio.

2.2 Google Web API

The Google Web API (now known as Google Code) is a public interface for expanding the functionality of Google. Using Simple Object Access Protocol (SOAP), a user can write services for search and data mining that can be automated or repeated and the results sent to a text file. An API application was written in VBscript that took a comma delimited text file containing search terms, and then searched for those terms and returned the number of pages that Google considered relevant to the search term.

The 214 Wainwright peaks were searched for and then those names together with one additional search term. These additional search terms were selected using all of the words associated with “mountain” from an online version of *Roget’s Thesaurus* (www.thesaurus.com, 2006), along with three geographical terms, that referred to the region. The full list of additional terms were Wainwright, Cumbria, Lake District, Mountain, Hill, Summit, Peak and Climb. For this list only, an additional random word was added - in this case the word “envelope” – to see if a word not associated with mountain features would produce any change in correlation.

For ease of understanding this work will refer to the name of the hill as the *search term* and the additional term as the *additional search term*. For example, for the peak “Buckbarrow” the following list of search terms and additional search terms would be produced:-

"Buckbarrow"
"Buckbarrow" Wainwright
"Buckbarrow" Cumbria
"Buckbarrow" "Lake District"
"Buckbarrow" mountain
"Buckbarrow" hill
"Buckbarrow" summit
"Buckbarrow" peak
"Buckbarrow" climb
"Buckbarrow" envelope

The results of this automated search were repeated one month later and the results compared, to test how reliable the relationships between terms and numbers of relevant pages. There was no significant difference between the two sets of results, although some of the numbers of relevant pages had understandably changed.

The correlation for every paring of the normalised values for Altitude, Voronoi polygon ratio, inverse watershed ratio, peak contributing area ratio and the web mining results was found using the Pearson Correlation Coefficient. It can be seen (in Table 1) that the relationships between Lake District, and the Inverse Watershed Ratio, Voronoi Polygon Ratio, Altitude and relative drop were significant at least the 0.05 level.

This method is biased towards the highest peaks which are high outliers. To counteract this, data was transformed to the log of each value. The results of the correlation carried out on the log transformed values can be seen in Table 2.

3. Results

Correlation of Non-transformed Results

		Peak Contributing Area Ratio	Inverse Watershed Ratio	Voronoi Polygon Ratio	Altitude	Relative Drop
Name Only	Pearson Correlation	-.039	.006	.010	.127	.032
	Sig. (2-tailed)	.570	.929	.886	.063	.645
Wainwright	Pearson Correlation	-.014	.071	.035	.084	.088
	Sig. (2-tailed)	.839	.304	.611	.222	.197
Cumbria	Pearson Correlation	-.013	.016	.015	-.017	.041
	Sig. (2-tailed)	.848	.817	.823	.805	.553
Lake District	Pearson Correlation	.105	.167	.221	.174	.262
	Sig. (2-tailed)	.127	.014	.001	.011	.000
Mountain	Pearson Correlation	-.035	.051	.031	.150	.056
	Sig. (2-tailed)	.614	.460	.648	.028	.418
Hill	Pearson Correlation	-.039	.043	.027	.124	.062
	Sig. (2-tailed)	.567	.528	.698	.070	.365
Summit	Pearson Correlation	-.028	.034	.024	.143	.041
	Sig. (2-tailed)	.680	.622	.726	.036	.555
Peak	Pearson Correlation	-.035	.003	.011	.133	.028
	Sig. (2-tailed)	.609	.966	.877	.052	.688
Climb	Pearson Correlation	-.028	.008	.015	.146	.021
	Sig. (2-tailed)	.683	.904	.831	.032	.765
Envelope	Pearson Correlation	-.037	-.012	.002	.124	.013
	Sig. (2-tailed)	.595	.867	.979	.070	.845

Table 1: Results of Pearson Correlation Test

(dark grey)	Correlation is significant at the 0.01 level (2-tailed).
(light grey)	Correlation is significant at the 0.05 level (2-tailed).

Logarithm Correlations

		Peak Contributing Area Ratio	Inverse Watershed Ratio	Voronoi Polygon Ratio	Altitude	Relative Drop
Name Only	Pearson Correlation	0.105	.199	.177	0.145	.245
	Sig. (2-tailed)	0.125	0.003	0.009	0.034	0
Wainwright	Pearson Correlation	-0.002	0.069	0.049	0.077	0.075
	Sig. (2-tailed)	0.979	0.318	0.472	0.265	0.273
Cumbria	Pearson Correlation	.181	.231	.247	0.134	.300
	Sig. (2-tailed)	0.008	0.001	0	0.051	0
Lake District	Pearson Correlation	.279	.323	.340	.248	.430
	Sig. (2-tailed)	0	0	0	0	0
Mountain	Pearson Correlation	0.091	.185	.199	0.172	.202
	Sig. (2-tailed)	0.185	0.007	0.003	0.012	0.003
Hill	Pearson Correlation	0.096	0.167	.181	0.136	.205
	Sig. (2-tailed)	0.161	0.014	0.008	0.047	0.003
Summit	Pearson Correlation	0.105	.186	.209	.187	.221
	Sig. (2-tailed)	0.125	0.006	0.002	0.006	0.001
Peak	Pearson Correlation	0.099	.181	.200	0.164	.206
	Sig. (2-tailed)	0.149	0.008	0.003	0.016	0.002
Climb	Pearson Correlation	0.126	.214	.232	.186	.238
	Sig. (2-tailed)	0.066	0.002	0.001	0.006	0
Envelope	Pearson Correlation	0.069	0.164	0.14	0.164	.183
	Sig. (2-tailed)	0.317	0.017	0.041	0.016	0.007

Table 2: Results of Pearson Correlation for Log Transformed results

(dark shading)	Correlation is significant at the 0.01 level (2-tailed).
(light shading)	Correlation is significant at the 0.05 level (2-tailed).

4. Some conclusions from the results

The results show that there are relationships between both the normal and log transformed sets of values and that it appears to be the case that the more physically prominent a peak is that the more culturally prominent it will be. There are some possible reasons for the patterns we see in the correlation tables. For example, in the non-transformed table, the strong relationship shown between altitude and all of the physical search terms could be a reflection of the “peak bagging” culture of British climbing – that is that the absolute elevation rather than the distinctiveness of a peak is what is driving the creation of web pages.

In the transformed data, there is a general pattern of correlation, with the exception of the search term “Wainwright”. One reason for this may be that any page that mentions the name of a Wainwright peak and the term "Wainwright" is likely to contain a list of all of the other Wainwrights. It is likely to be a page for walkers which show the peaks as a group, with either links to the other peaks or a list of the peaks that constitute the group.

The fact that there is some correlation does not necessarily mean that this is the most appropriate method for comparing cultural and physical prominence. One could argue that comparisons of this kind are as inappropriate as comparing apples and oranges. Further investigation is required to understand why these patterns are found and whether they have any significance.

7. References

Beucher S and Lantuejoul C, 1979; *Use of Watersheds in Contour Detecion*

Mark, D. and Turk, A., 2003; *Ethnophysiography* Conference on Spatial Information Theory (COSIT)'03, Ellicottville, NY, USA

Smith, B. and Mark, D. M., 2003; *Do mountains exist? Towards an ontology of landforms* Environment and Planning B-Planning & Design, Vol. 30(3)**,** pp. 411-427.

Stevens, S., 2005. Bagging. Available at: http://www.simonhjstevens.com/sml/bagging/ (last accessed 7th November 2005)

Biography

Ian Greatbatch is in the third year of a PhD funded by City University.