GEOMORPHOMETRY 2020 PERUGIA, ITALY JUNE 22 - 26 2020

Automated extraction of areal extents for GNIS Summit features using the eminence core method

Gaurav Sinha[1], Samantha T. Arundel[2]

[1]Department of Geography
122 Clippinger Labs, Ohio University, Athens, Ohio, USA
sinhag@ohio.edu

[2]U.S. Geological Survey
Center of Excellence for Geospatial Information Science
1400 Independence Rd, Rolla, Missouri USA

Abstract— **An important objective of the U.S. Geological Survey (USGS) is to enhance the Geographic Names Information System (GNIS) by automatically associating boundaries with terrain features that are currently spatially represented as two-dimensional points. In this paper, the discussion focuses on experiments for mapping GNIS *Summit* features using the eminence core region-growing method, which maps the area between a peak and its key col (saddle). A secondary goal of this project is to improve the positional accuracy of GNIS *Summit* features, since those locations were derived long ago and need to be snapped to local morphometric peaks detected from analysis of the highest-resolution digital elevation models (DEMs). The eminence cores delineated for a subset of GNIS *Summit* features were compared visually against basemaps and manually digitized polygons created by USGS staff. The comparisons revealed substantial differences between the computationally derived eminence cores and the manually generated polygons. Results clearly suggest that the default core delineation method tested must be modified to "roll back" or truncate growth of unreasonably large cores to smaller extents that would match people's intuitive expectations. However, these results are far more encouraging than any method tested previously, since this method guarantees a 1-1 correspondence between polygons and GNIS *Summit* features.**

I. INTRODUCTION

An important objective at the U.S. Geological Survey (USGS) is to enhance the Geographic Names Information System (GNIS) by associating boundaries with terrain features that are currently spatially represented simplistically as two-dimensional points [1]. The driving force for this work is the need for more realistic representation of features to answer spatial semantic queries that rely on their areal boundary. There is no standard method for delineating landforms. Three broad approaches that are being tested in parallel at the USGS for landform mapping are object-based image segmentation and analysis, hybrid pixel and object-oriented region-growing methods, and machine learning-based image segmentation and object detection.

The authors initially experimented with well-known pixel-based landscape classification and segmentation methods, defined in [2-5], but none of them can guarantee outcomes where each GNIS feature can be mapped exclusively to one specific polygon. A geographic object-based image analysis (GEOBIA) workflow originally proposed in [6] was tested for mapping areal extensions for GNIS terrain features [1]. Out of 16 terrain feature classes recognized in GNIS, only *Summit* and *Valley* feature class members were found to be spatially correlated with only the high and low elevation GEOBIA terrain classes. The segmented terrain objects did not provide a unique polygon for each feature and the shapes of most polygons did not match common sense expectations of boundaries of individual *Summit* or *Valley* features. Thus, the authors concluded that GEOBIA is probably best for general physiographic characterization of terrain, but not for extracting cognitively plausible areal footprints for individual terrain features.

The current approach is on top-down region-growing methods for mapping landforms. The initial focus is on mapping only GNIS *Summit* features because they are shown on most topographic maps, and several algorithms for mapping areal extents of topographic eminences (in other words, convex-shaped landforms such as *Summits*) are available for comparison already [7-11]. In this paper, results are reported from the application of the eminence core region-growing method [11], chosen specifically because of its cognitive and technical simplicity, the guarantee of exclusive discrete regions for every feature, and potential to serve as a general purpose and easily customizable method for mapping the wide variety of features classified as *Summit* features in GNIS.

Gaurav Sinha, Samantha Arundel (2020) Automated Extraction of Areal Extents For GNIS Summit Features Using the Eminence-Core Method:

in Massimiliano Alvioli, Ivan Marchesini, Laura Melelli & Peter Guth, eds., Proceedings of the Geomorphometry 2020 Conference, doi:10.30437/GEOMORPHOMETRY2020_10.

II. METHODS

A. Eminence Core Delineation Method

The eminence core method is a region-growing method proposed originally in [9] and then adapted and extended into a more comprehensive cognitive modeling framework for eminence delineation in [11]. This method must be "seeded" with the location of a known morphometric peak, which can be defined as a local maximum of zero dimension within a defined neighborhood. The method then expands iteratively from the peak to map an eminence core that can be exclusively associated with only one peak. The default conceptualization of the eminence core is the area between the peak and the lowest (base) contour that completely encloses the peak and contains no location higher than the peak [11]. However, a smaller relative drop may also be specified to extract a smaller core. The base contour also supports the peak's key col, which is the lowest saddle between the peak and another higher peak (Figure 1). The difference in elevation between the peak and the key col is the peak's topographic prominence, widely used in the mountaineering community for ranking the attractiveness of peaks for mountain climbing. Prominence is an intuitive and highly effective filtering parameter for selecting topographically salient peaks from the large number of inconsequential peaks that are initially identified from local moving window analysis [11].

In [10], a vector contour-based representation method was used for mapping the core area. The authors prefer a DEM-based region-growing method using a priority queue data structure that ensures O(n) complexity [9]. The core area is mapped through expansion beginning at the peak cell by first adding the immediate eight neighboring cells to the queue, forming the minimal possible core area for a peak. The priority queue automatically sorts cells in descending order of elevation, and then selects the peak cell to promote the next highest cell to the top of the queue. This process is iterated until the cell at the top of the priority queue is at the edge of the study area or higher than the starting peak cell. The selected cells collectively delineate the eminence core area associated with the peak. The lowest elevation cell that reaches the top of the priority queue is the key col cell, which can be used to additionally determine the peak's prominence.

A limitation of this method is that for high-prominence peaks, the key col is at such a large distance that the eminence core grows unreasonably large to enclose cores of all other salient peaks between the prominent peak and its key col [9], [11]. Moreover, in planar areas with few isolated eminences, even a gentle slope can force the core to continue growing until the key col is detected, whereas people would judge the eminence boundary to be far more compact [11]. Two different approaches to solve this problem are presented in the literature [10-11] and may need to be implemented and compared to derive cognitively plausible cores for *Summit* features.

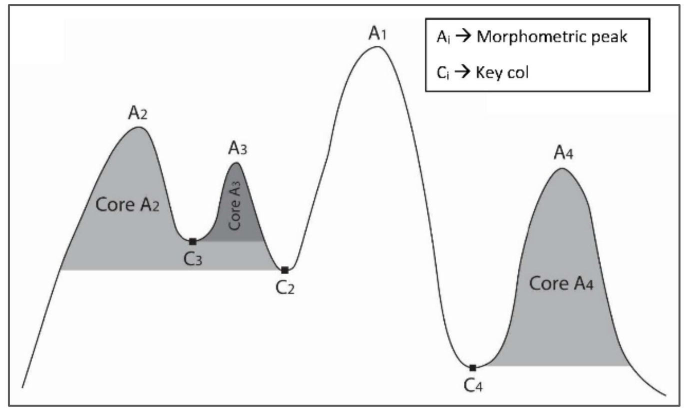

Figure 1. Conceptual diagram for illustrating the core area and key cols ($C_2 - C_4$) of peaks ($A_2 - A_4$). The highest peak A_1's key col is beyond the area shown.

B. Snapping GNIS Summit features to morphometric peaks

Because GNIS terrain point locations were collected for the names as opposed to their location, most GNIS *Summit* features do not coincide exactly with the morphometric peaks detected from high-resolution DEM analysis. For implementing the eminence core delineation method for GNIS *Summit* features specifically, it is essential to relocate or "snap" the GNIS *Summit* feature to the correct morphometric peak nearby. Because there are currently about 70,000 *Summit* features in the GNIS database, a separate secondary project was launched for automating as much of this *Summit* feature location enhancement process as possible. This effort presented its own challenges that are now documented in a separate manuscript [12] and not discussed here for lack of space. For this paper, it is pertinent that the experiments detailed below with GNIS *Summit* features successfully snapped to the correct morphometric peak.

III. RESULTS

The cognitive plausibility of the delineated eminence cores was visually assessed using Esri's ArcGIS Pro software, wherein the extracted core polygons were overlaid on a topographic basemap and a terrain hillshade layer in both two-dimensional map views and three-dimensional scenes. Additionally, the authors used 118 manually delineated polygons corresponding to GNIS *Summit* features in the Blue Ridge mountains of the Appalachian mountain range as reference data (Figure 2). These polygons were digitized manually by USGS staff in the National Geospatial Technical Operations Center's US Topo program to capture the extent of feature labels on historical USGS topographic maps. Labels were originally placed on USGS topographic maps to reflect USGS field surveyors' assessment of the approximate extent of the features. While comparing the polygons overlaid on topographic maps, the authors also found the digitized polygons as cognitively plausible representations of the *Summit* features.

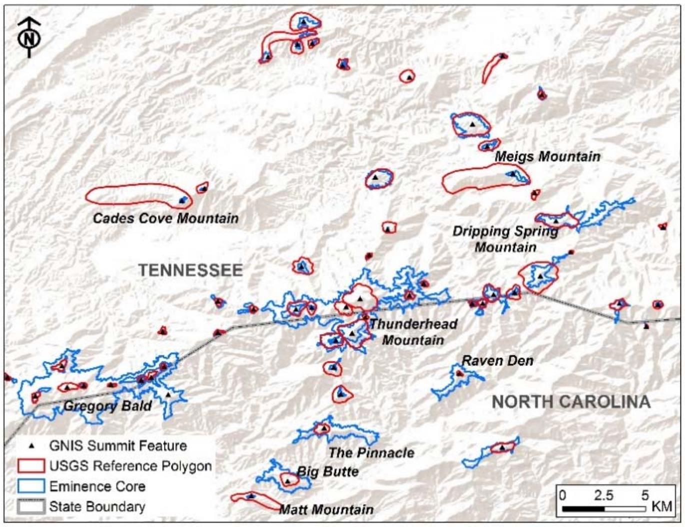

Figure 2. A comparison of a subset of computationally generated eminence cores and manually digitized polygons for GNIS *Summit* features in the Blue Ridge province of the Appalachian mountain range in USA.

A. Visual comparison of computationally extracted eminence cores and digitized polygons

The map in Figure 2 shows a part of the Great Smoky mountain range along the Tennessee–North Carolina border. Visual comparison of the computationally extracted eminence cores (blue polygons) and the manually digitized reference polygons (red polygons) offer several insights. It is obvious that both eminence cores and manual polygons comprise a wide range of shapes between compact and elongated (Figure 2). This map clearly establishes that the GNIS *Summit* feature class is a general category that includes a wide variety of topographic eminences. Mapping areal representations of the individual features will clearly reveal hitherto unknown information about the range of shapes and sizes of eminences, not just in the United States, but anywhere in the world.

The overlap patterns between the cores and polygons are also quite revealing. Sometimes, the manual polygon is contained within or is larger than the core, but the reverse is also true for many features. It is only for smaller and compact eminences that the core and reference polygons look similar. However, there are several cases where the two sets of areal representations are substantially different in shape and extent. For example, Cade's Cove Mountain (west), Meigs Mountain (east) and Matt Mountain (south) were digitized as elongated polygons, but the eminence core growth is truncated quite prematurely. The most likely explanation is that the digitized polygons for these *Summit* features represent larger eminences containing multiple topographic peaks, as shown on the historical USGS topographic map.

Conversely, there are several cases of extremely large and extended cores, which result when the key col is distant from the summit point. In these cases, the core is too large and can contain smaller cores of other nearby *Summit* features. These cores must be shrunk or "rolled back" to match the boundaries of those eminences based on people's common-sense expectations of acceptable feature extents. Two examples are Thunderbird Mountain (center) and Gregory Bald (southwest) in the map in Figure 2. These are *Summit* features with high prominence due to distant key cols, which means their cores will be large and contain the cores of subsidiary peaks. The USGS staff did not interpret these major *Summit* features to have such extremely large extents, since the polygons they delineated were much smaller.

B. Geometric comparison of computationally extracted eminence cores and digitized polygons

Results from the analysis of eminence cores and manually derived polygons are presented in Table 1. The *geodesic area* and *geodesic perimeter* were calculated for both sets of polygons. The *percentage difference* was also calculated. The summary statistics in Table 1 clearly show that the range and variance of area and perimeter measures are substantially higher for the automatically extracted eminence cores. The last column reveals that the difference in percentage in size between the computationally extracted cores and manually digitized polygons is extremely high. This clearly supports the visual assessment that there are many cores that are quite large and are the reason for the extreme percentage change values.

Property	Statistic	Automated Core	Manual Polygon	Percent Difference
Area (m^2)	***Min***	603	11,281	1.4
	Max	17,276,850	9,795,338	74,863
	Mean	1,397,879	1,001,331	1,893
	Std. Dev	3,600,661	1,636,711	3,246
Perimeter (m)	***Min***	142	401	0.3
	Max	101,708	18,192	1,733
	Mean	9,233	3,497	123
	Std. Dev	17,406	3,246	253

Table 1. Statistical summary of geometric measures for computationally and manually generated areal representations of GNIS *Summit* features

Figure 3 shows summary statistics for *Coefficient of Areal Correspondence (CAC)* index, which is a ratio of the overlapping area to the sum of all areas occupied by individual features measured [13]. CAC ranges from a minimum of zero (no overlap) to a maximum value of 1 (perfect overlap between the two sets of features). The mean of the CAC values for the 118 pairs of cores and manual polygons is quite low, underscoring again that there is a substantial difference between the two sets of representations.

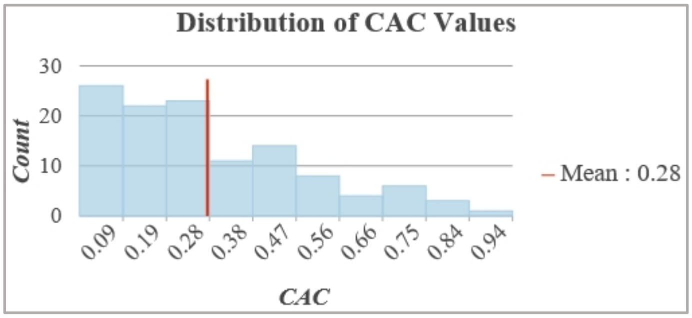

Figure 3. Histogram of Coefficient of Areal Correlation (CAC) values obtained from overlay of manually generated polygons and computationally extracted eminence cores.

IV. DISCUSSION AND CONCLUSION

In this paper, the focus was on testing the eminence core region-growing method's feasibility for creating areal representations of GNIS *Summit* features. Validation of automatically generated eminence cores against manually digitized polygons provided useful insights for advancing this project. The most encouraging finding is that this method of delineating boundaries generates a cognitively plausible solution because there is a clear 1-1 correspondence between features and the delineated areas. This approach is far more encouraging than the GEOBIA workflow tested earlier [1].

However, the default key-col-based delineation produced larger than acceptable polygons for major *Summit* features, suggesting the need for a more complex set of criteria for deciding how to terminate eminence core region growth. The smaller and compact cores correspond well to manually digitized polygons, but there are many extremely large cores that need to be truncated to match people's common-sense assumptions of eminence extents. It is quite difficult to delineate complex eminences with multiple peaks. In [10], a morphological variance reduction method was suggested, whereas in [11] the suggested method is to "roll back" the core based on continually measuring the average boundary slope threshold criterion until it exceeds a threshold. Both approaches need to be tested and compared further.

There is no prescriptive method for mapping the boundaries of topographic eminences. The eminence core approach is easy to implement and was tested first because of expediency, but this does not mean that alternative conceptualizations for mapping the areal extent of eminences will not be supported in this project. As shown in [11], instead of relying on the key col contour, other core mapping techniques (for example, maximum relative drop or minimum elevation threshold, fiat cols, salient apexes, slope inflection points, and landcover change line) may provide more satisfying results.

Whereas it may be sufficient to modify and use the default eminence core method for meeting the short-term mapping goals of the USGS, the alternative methods may be more cognitively appealing in other mapping projects, such as those at other national mapping agencies, mapping of extents of sacred landforms by indigenous communities, and perspective-driven landform visualizations. Thus, the authors hope that this research will lead to a more holistic and comprehensive framework that can support multiple methods and provide guidelines for dynamically extracting context-specific and cognitively appealing boundaries for topographic eminences.

REFERENCES

[1] Arundel, S. T., Sinha, G., 2019. Validating the use of object-based image analysis to map commonly recognized landform features in the United States. *Cartography and Geographic Information Science*, 46(5), 441–455.

[2] Wood, J., 1996. The Geomorphological Characterisation of Digital Elevation Models. Unpublished PhD Dissertation, Leicester University.

[3] Deng, Y., and Wilson, J. P., 2008. Multi-scale and multi-criteria mapping of mountain peaks as fuzzy entities. *International Journal of Geographic Information Science*, 22, 205–218.

[4] Fisher, P., Wood, J., and Cheng, T., 2004. Where Is Helvellyn? Fuzziness of Multi-scale Landscape Morphometry. *Transactions of the Institute of British Geographers*, 29, 106–128.

[5] Jasiewicz, J., and Stepinski, T. F., 2013. Geomorphons — a pattern recognition approach to classification and mapping of landforms. *Geomorphology* 182: 147–156.

[6] Drăguţ, L., and Eisank, C., 2012. Automated object-based classification of topography from SRTM data. *Geomorphology*, 141–142, 21–33.

[7] Graff, L., and Usery, E. L., 1993. Automated Classification of Terrain Features in Digital Elevation Models. *Photogrammetric Engineering & Remote Sensing* 59: 1409-1417.

[8] MacMillan, R. A., Pettapiece, W. W., Nolan, S. C., and Goddard, T.W., 2000. A Generic Procedure for Automatically Segmenting Landforms into Landform Elements Using DEMs, Heuristic Rules and Fuzzy Logic. *Fuzzy Sets and Systems* 113/1: 81-109.

[9] Wood, J., 2004. A New Method for the Identification of Peaks and Summits in Surface Models. *Proceedings of GIScience 2004*.

[10] Chaudhry, O. Z., and Mackaness, W. A., 2008. Creating Mountains out of Mole Hills: Automatic Identification of Hills and Ranges Using Morphometric Analysis. *Transactions in GIS*, 12(5): 567-589.

[11] Sinha, G., 2008. Delineation, Characterization, and Classification of Topographic Eminences. *Unpublished PhD Dissertation*, Department of Geography, University at Buffalo, NY.

[12] Arundel, S. T., Sinha, G., and Chan, A., 2020. Automated location correction and spot height generation for named summits in the coterminous United States. *International Journal of Digital Earth. Accepted*.

[13] Taylor, P. J., 1977. *Quantitative methods in geography: An introduction to spatial analysis*. Areal association, Houghton Mifflin, Boston, p. 396.