Areal Differentiation Using Fuzzy Set Membership and Support Logic

Michael N. DeMers

ABSTRACT: Qualitative thematic maps are graphic representations of the process of areal differentiation, a process recognized as one of the hallmarks of the geographer's craft and one of geography's grand traditions. Traditionally the process can best be described as classification based on Boolean Set Theory wherein the area bounded by the cartographic line is considered to be in that class for the target polygon, or stated differently it is considered to be contained within the boundary of the polygon. Conversely then, any content not contained in the polygon is not of that class of polygon. There are two fundamental confounding issues with the use of Boolean logic for such classification of geographic areas however. First is that the determination of class membership itself is often less than 100% obvious, either because of the lack of visual cues in interpretation from a source document, or by the subjective nature of the linguistic variables and classes selected. This suggests that the process more closely resembles fuzzy set membership than it does Boolean set membership. Second, the categories, whether discretely defined or more subjective will, when cartographically depicted result in polygonal adjacency, where the line represents two or more categories simultaneously, and where any attempt at adding something like an epsilon band of uncertainty for such lines results in conflicts because of assumed complementarity of the two set values. This paper hypothesizes that one alternate method of dealing with the uncertainty of such lines be based first on Fuzzy Set membership values to account for the classification uncertainty itself. Because Fuzzy Set Memberships do not assume complementarity, by extension support logic – an interval-based probabilistic reasoning is applied to the fuzzy set memberships to create a modified version of the epsilon band normally applied to the uncertainty of the graphic line itself. The support logic extends the fuzzy set membership of polygonal containment to include the degree of logical support there is for the category. In this support logic epsilon band, which I call an SLE-Band, you are able to include both graphic line uncertainty (the Epsilon part) with the maximum support of membership and the minimum necessary support of membership. While still at a conceptual stage of development, the paper proposes the conceptual mathematical framework supporting the development of this Fuzzy Set based support logic strategy that will enable the determination of fuzzy set membership (polygonal inclusion) as well as providing the support logic for the membership value of any polygon. Further, while not the major thrust of this paper I will propose some possible methods of graphic display and briefly describe how such cartographic layers may be compared through alternative overlay operations not currently implemented in contemporary commercial GIS software.

KEYWORDS: Areal Differentiation, Map Boundary Determination, Fuzzy Set Polygon Classification, Fuzzy Set Support Logic

Introduction

If one considers a major objective of geography the observation, description, analysis, and understanding of spatial pattern, of necessity the correctness of the formalization of the subject matter is paramount. Couclelis (1992) posited that geographic phenomena could be formalized as either field or object and that these distinctions were fundamental to understanding our world. Rather than deliberate that there might be a general theory of geographic representation forthcoming, as has been proposed by Goodchild, Yuan, and Cova (2007), this paper assumes the distinction between field and object, and focuses on the object, specifically the differentiated area, especially as represented in geospatial databases and cartographic representations. Early work by Sauer (1921) enumerated a classic example of the issues related to the problems of classifying land areas. Arguments about the very existence of areas as unitary objects in geography were later enumerated by Hartshorne (1939a), which led to discussions of the characteristics of a systematic geography (Hartshorne 1939b). Years later, Hartshorne (1962) provided more detail of the debate between the importance of the contents of areas and the areas themselves. Ultimately the two are integral to one another as the study of an area assumes one can define that area by its contents.

Unlike areal maps composed of well-defined typological regionalized boundaries, such as political or land ownership boundaries, areal differentiation of the land is a vague, if not highly subjective process whereby even supposed objective numerical classification methods rely on the same general geographical approach of regional core and zones of transition (Johnston, 1968). All classification can be thought of as consisting of collections of items that do or do not belong to a set. Thus, a land cover class of, for example, grassland would contain those portions of the earth that consist of that category. Stated differently, the portions of the earth (polygons) would be members of the set of grassland polygons. If grassland were always to exist as containing 100% grass cover and nothing else, the decisions to include a polygon within the set of grassland polygons would be an easy one because they would either be 100% grass or they would not. As such they would either be members of the set of grassland polygons (denoted as 1 or present) in Boolean Set mathematics, or they would not (denoted as 0 or absent). In reality such pure conditions are rare, in that the land may have a proportion of grass (say 90%) and other types of vegetation like trees (10% in this case). Fundamentally what this suggests is that areal differentiation will produce results of variable purity. In some cases, such as our grassland circumstance, the purity can be directly observed or measured, while in other situations a determination of purity may depend on multiple factors, such as the properties of soils that change continuously yet are represented as discrete polygonal objects. The problem we face is that discounting these vagaries results in categorical maps whose classifications are of various, undocumented quality. The discrete lines representing such classified polygons are likewise variable. Even more disturbing is that in some cases the support for selecting a particular category, such as those related to integrated terrain units might be equally uncertain.

Focusing on the grouping of ecoregional classifications, Bittner (2012) employed an ontological analysis of classification vagueness and showed that the argument over classification is ongoing and ranges between those who believe that many geographic regions can be differentiated by scientific means of delineation while others believe that classification is an artificial construct and often policy-based. He defined three aspects of the ontology of vague regions as (1) granularity – such as the scale-dependent relationships of the setting for the ecoregions, (2) the vagueness of the classifications based on the qualities of the setting – e.g. landforms, vegetation, and others, and (3) the vagueness that affects the boundary lines separating regions of different qualities. His conclusion was not promising as he suggested that the vagueness and granularity are unavoidable and that the need to use geometrically precise boundaries is in direct conflict with such vague classifications. From an analytical cartographic and GIS perspective he suggests that there is no current method to accommodate these tradeoffs. This paper proposes that research into Fuzzy Set Theory, Fuzzy Logic, and Support Logic Programming might provide solutions.

Literature Review

There have been many attempts at defining how decision-making can be quantitatively derived even dating back to the late 1950's (Zobler, 1958, Berry 1958). A substantial proportion of the remote sensing literature has focused both on how best to accomplish quantitative methods of categorical analysis by evaluating the properties of groups of pixels and grouping them. These latter methods have and many others might be adapted to allow for the development of Fuzzy Membership values of categories rather than their Boolean equivalents (Foody 1996). In the vector domain most of the focus of classification in GIS has been to consider the categories as exhaustive (Beard, 1988, Chrisman, 1982, Frank, et al. 1997, Goodchild, Guoqing, and Shiren, 1992, Mark, and Csillag 1989) and attempts to find the best equivalent classes for selected categories. With this predilection it stands to reason that the analysis of the error associated with lines bounding such classes would be a logical focal point.

Its importance placed it as a major research initiative of the National Center for Geographic Information And Analysis (NCGIA) [http://www.ncgia.ucsb.edu/research/initiatives.html] (last visited July15, 2016). While the integrity of the cartographic line to represent the boundary of a classified polygon remained, an early attempt at reducing its rigid nature based on the error associated with changes is scale was based on the idea that the line itself introduced a band separating adjacent categories. The width of the band was scale-dependent. Thus the smaller the scale of the map, the more impact the line would have on regional accuracy between any two categories. This Epsilon Band (Blakemore 1984) was then considered a zone or buffer in which neither category could claim ascendance.

An important consideration of the Epsilon Band is that its function is not to define classes, nor to provide a means of developing those classes, but rather makes the case that as two or more differentiated areas come together, the line itself can contain neither category as there is no logical support for one category being more dominant than another. In essence, regardless of whether any two areas are classified as Boolean set members or as Fuzzy Set members with membership values, there is no basis for assigning either class to the Epsilon Band. Skidmore and Turner (1992) even expanded the Epsilon Band idea to include tests of line accuracy compared to a control line, thus generalizing the concept to include not just scale but line sampling as well (Figure 1).

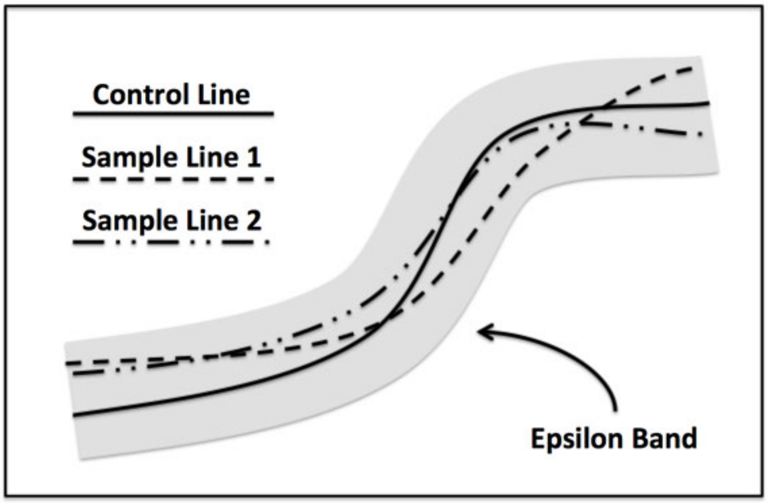

Figure 1. Epsilon band used to evaluate area of uncertainty of sample lines.

The epsilon band concept allows one to acknowledge the effect inherent in the scale-dependent nature of the drawn line or even the degree to which these boundary lines conform to some agreed upon control. Unfortunately, neither of these addresses the issue of how to model the degree of support one might have for the categories selected, based on the criteria used.

The problem with the epsilon band then, is that it exists as a discreet entity while the relationship between any two adjacent categories is more likely a gradation. Kronenfield (2011) provided an alternative to the Epsilon Band model, one that characterizes the areas as gradational (Figure 2). His polygonal modeling approach adds zones of uncertainty drawn perpendicularly from a medial axis outward to each adjacent polygonal line (Figure 3). This approach has the unique advantage over probabilistic methods (Honeycutt 1987) in that the polygons created can be used for analytical operations in GIS, especially important among them are any of the overlay approaches. Because the transition zones are themselves however Boolean fails to characterize either the degree of category on either side (Fuzzy set membership values) or the degree of transition itself within the transition zone polygons.

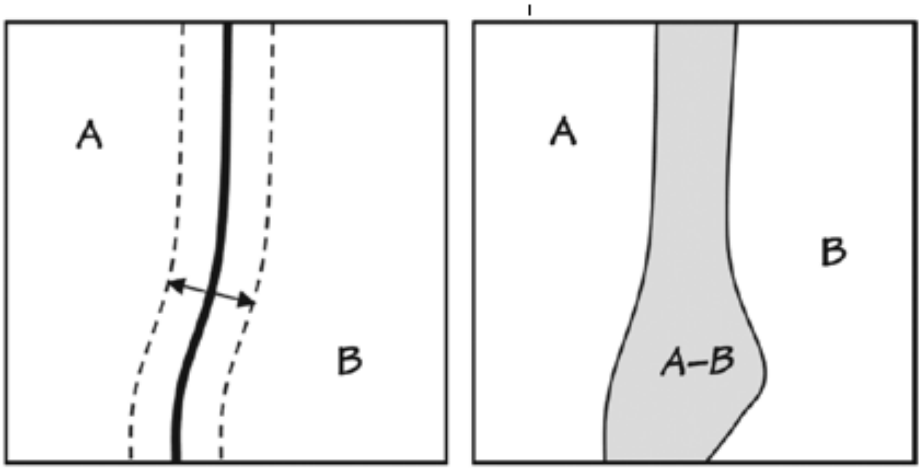

Figure 2. Traditional discrete epsilon band (left) and a polygonal epsilon band with transition (right). Modified from Kronenfeld (1992)

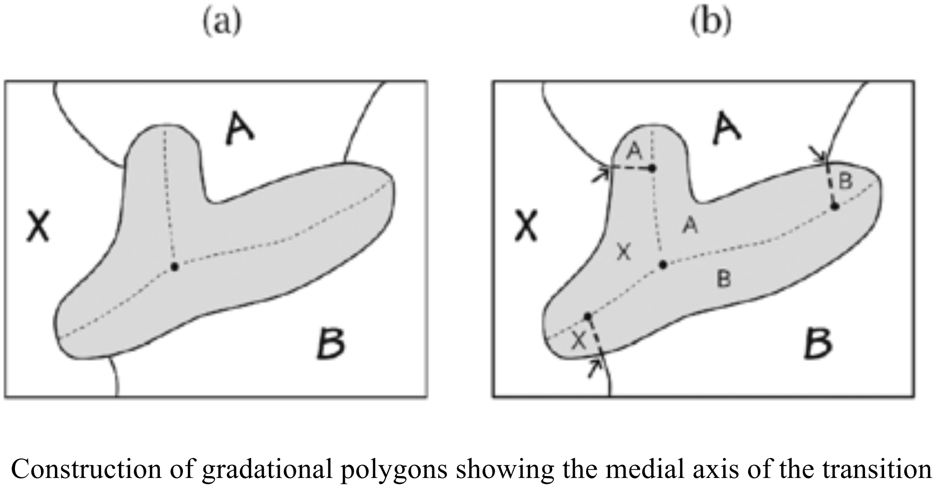

Figure 3. Construction of gradational polygons showing the medial axis of the transition zone (a) and the extended medial axis with segments (heavy dashed) added connecting triple-boundary points(denoted by arrows) to the medial axis. Modified from Kronefeld (1992).

Fuzzy Sets and Support Logic Programming

I propose that Fuzzy Set Theory and Fuzzy logic can be applied not only to the classification process, but more importantly for the current consideration, as a means of modeling the imprecision of adjacent boundaries. Fuzzy logic is a form of imprecise reasoning or reasoning in the absence of crisp logic (i.e. Boolean Logic). In Boolean Logic an object, convention, or condition is either present or absent – that is, it either belongs or does not belong to a particular set. Fuzzy Set Theory extends Boolean Logic to incorporate the degree of membership *from* 0 to 1 and, by extension provides a means of representing support for the degree of membership of a contention, condition, or object within a defined set. The approach in Fuzzy Logic is to pose a question regarding set membership. So, as a geographic situation one might, for example, ask the question "is the polygon bounding the town of Lovington, New Mexico (Figure 4) a polygon of ecoregion Arid Llano Estacado?" Now looking at the map it is currently classified as such, but let us assume that it has impurities in it, and your best guess is that about 15% of that area might not correctly be classified as Arid Llano Estacado. One could map this polygon and include a notation that says something like "Arid Llano Estacado 0.85" meaning that it belongs to the category Arid Llano Estacado with a membership of 0.85. So rather than saying it is (1) or is not (0), it is qualified with a set membership value indicating the degree to which it belongs in that ecoregion classification set.

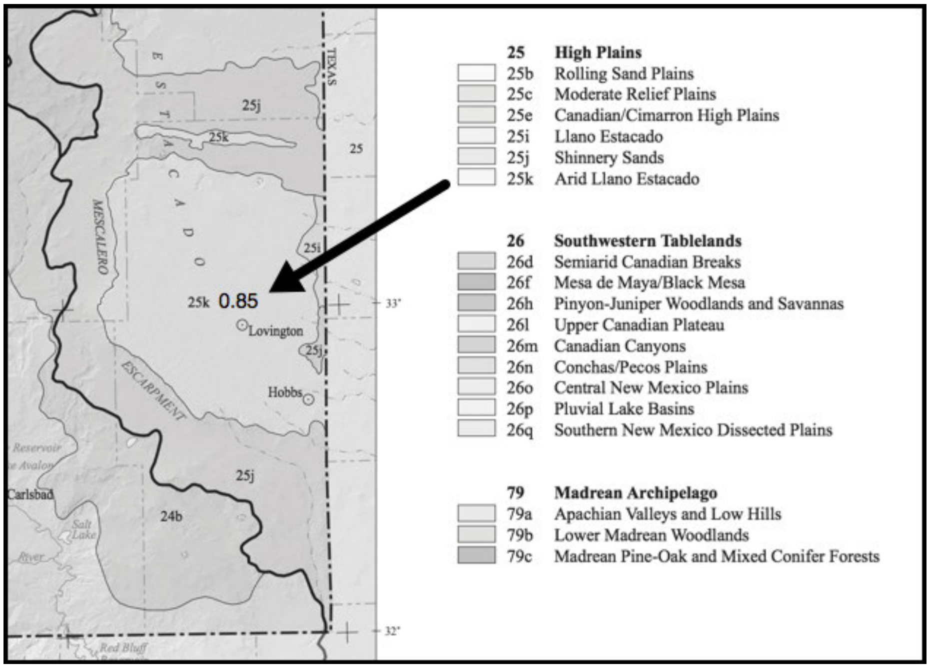

Figure 4. A portion of the EPA Ecoregions Map showing the southwest corner of New Mexico. Note the addition of 0.85 signifying the degree to which it belongs in its category 25k.

We have now seen that Fuzzy Set Theory provides us with an ability to catalog set membership (the degree to which a polygon belongs to a given category), and this is a substantial improvement over Boolean Set Theory, however, it lacks the ability to document situations in which evidence for a particular contention, condition, or object may exist in equally variable amounts. For example, there may be considerable evidence that a particular polygonal map category of ecoregion exists with an associated membership of, for instance 0.7, under optimal circumstances – i.e. well documented variables of landform, slope, vegetation cover, etc. Under other circumstances evidence for the same contention existing with the same 0.7 membership value may be present, but in less than optimum; even minimal conditions. For example, the vegetation present is difficult to interpret, the nature of the landforms is unclear, and the slope is highly variable. While one might want to deem these as the probabilities of the conditions existing with the specific membership value, this degrades the purity of relying entirely on Fuzzy Set Theory and Fuzzy Logic. This is critical given that there is no substantial body of theory supporting the overlay of probability surfaces – a common use of categorical maps in GIS. A probabilistic interpretation further complicates matters because it relies on the rules of probability theory that may not be available in non-stochastic modeling environments, particularly where variable evidentiary support exists. A method is needed to account for this.

Baldwin (1984) addressed this problem when he developed a method to account for this range of support for Fuzzy Membership values called support logic programming (SLOP), eventually formalized as a part of a Prolog-like logic-based programming language called the Fuzzy Relational Inference Language (FRIL) (Baldwin, et al. 1987, 1988). Based on a combination of Fuzzy Set Theory (Zadeh 1978) and Shafer's (1976) theory of evidence, that method employs a set of support pairs that acknowledge a range of *possible* support where optimum circumstances can be met by the evidence, if it exists, and a *necessary* support level where an absolute minimal level of evidentiary support required for the particular set membership value exists. The general form of the support clause is:

(<atom>):(<L> <U>)

where <L> is the necessary support for the atom and <U> is the possible support for the atom. The atom, in our case is the selected category defined by its differentiating factors.

In this way, one can have a range of support for any given object, contention, or condition (or map category). While considering this remember that areal features do not occur in isolation, but are most often adjacent to other areal features whose classification may be likewise variable. If one employs the concept of regional core and transitional areas one might assume that the farther one is from adjacent areas, the more support one would have for the category selected and as one moves closer to adjacent polygons, the less support one would have for that same category. It is logical to assume that this relationship is non-linear, and that there is little chance of categorical confusion until one is quite close. In the Boolean case, the chance for confusion, in fact, only exists along the line drawn between adjacent categories. And as we have already seen, the Epsilon Band defines the line as a region of categorical uncertainty.

Method

Given adjacent polygons, each of which has its own set of categories and each with a level of membership one can define the evidence supporting the categories themselves. If we are considering ecoregions as our example, and further focusing on the boundary between any two such polygons one can define the support for each polygon's category based on the following.

Given polygon A with ecoregion class X and an adjacent polygon B with ecoregion class Y:

1. Membership value of ecoregion category X in polygon A.
2. Membership value of ecoregion category Y in polygon B.

Assumption: the higher the set membership value the greater its categorical purity. The greater the categorical purity the greater the inertial quality of the category.

These factors are represented in FRIL as the following for polygon A:

((Is category A)): (0.7 0.9)

which means that the necessary evidence that the polygon is category A is 0.7 (which means that the selected category is supported to the degree 0.7), and the possible evidence is 0.9 indicating that there is a possibility that the selected category is supported to a higher degree of certainty).

((Is category B)): (0.6 0.8)

which means that the necessary evidence that the polygon is category A is 0.7 (which means that the selected category is supported to the degree 0.7), and the possible evidence is 0.9 indicating that there is a possibility that the selected category is supported to a higher degree of certainty).

Because the two are near each other there is the possibility that Polygon X might contain some of the adjacent category B and that Polygon Y may contain some of adjacent category A. It should be noted that the two need not be complementary as they are different questions.

Using the medial axis of the boundary line as a starting point, as was proposed by Kronenfeld, one can move outward toward each category the zone is based on the support logic. As such for category A, the range of support ranges from 0.7 at the median to 0.9 at the end of contact of the epsilon band and the content of category A (Figure 5). In this case, however the median is adjusted in the direction opposite of the polygon with the greatest necessary support by the difference in necessary support. So in this case the median is moved 0.6 – 0.7 = 0.1 or 10%. Figure 5 shows what would result from this approach.

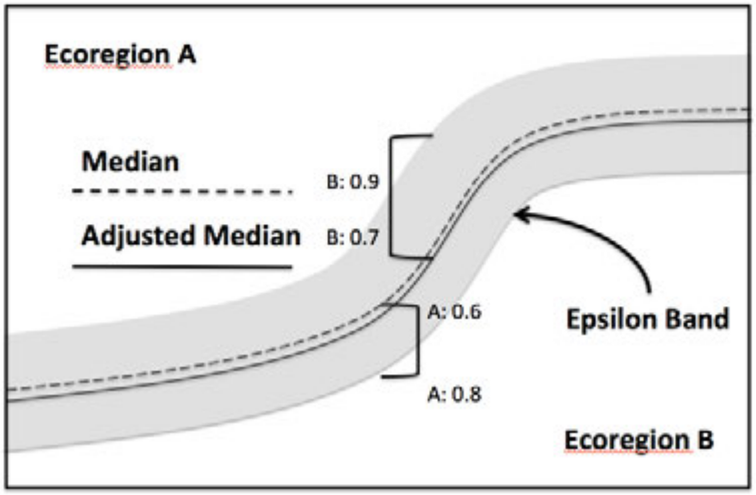

Figure 5. Epsilon band modified to incorporate a range of supports. The Median is adjusted by 10% from ecoregion A toward ecoregion B as the minimal support for ecoregion A is 10% higher (0.7) than for ecoregion B (0.6).

As displayed in figure 5 there is no difference between the visual appearance of the ecoregion bands but a variety of graphical techniques could be employed for their display. Essentially a gradational range for each would be presented, but rather than going from 0% to 100% the ranges are those that represent the support pairs. As such a range of shades of 90% would be used near the ecoregion A edge and graduated until a value of 70% at the adjusted median. Alternatively, the epsilon band would range from 60% at the adjusted median to 80% at the margin with ecoregion B.

Conclusions

This paper presented a conceptual model to add support logic that I call the SLE Band (Support Logic Epsilon Band) to the traditional epsilon band. Rather than a continuous, linear gradation from 0% to 100% moving from the median outward to the edge of each category, the range uses Baldwin's support logic based on Fuzzy Set Theory and Shafer's mathematical theory of evidence. The median itself is also adjusted to account for the impact of greater evidence, if present, for one category over another. The concept requires testing and research is required to determine if it is able to be implemented and how to best graphically display the results.

References

Baldwin, J.F. (1987) Evidential Support Logic Programming. *Fuzzy Sets and Systems* 24:1-26.

Baldwin, J.F. and S.Q. Zhou (1984) A Fuzzy Relational Inference Language. *Fuzzy Sets and Systems* 14:155-174.

Berry, B. (1958) A Note Concerning Methods of Classification. *Annals of the Association of American Geographers* 48(3): 300-303.

Bittner, T. (2011) Vagueness and the Tradeoff Between the Classification and Delineation of Geographic Regions – An Ontological Analysis. *International Journal of Geographic Information Science* 25(5): 825-850.

Couclelis, H. (1992) People Manipulate Objects (But Cultivate Fields): Beyond the Raster-Vector Debate in GIS. In A.U. Frank, I. Campari, and U. Formentini (Eds.) *Theory and Methods of Spatio-Temporal Reasoning in Geographic Space*, pp. 65-77.

Foody, G.M. (1996) Approaches for the Production and Evaluation of Fuzzy Land Cover Classifications from Remotely Sensed Data. *International Journal of Remote Sensing* 17(7): 1317-1340.

Frank, A.U., G.S. Volta and M. McGranaghan (1997) Formalization of Families of Categorical Coverages. *International Journal of Geographical Information Science* 11(3): 215-231.

Goodchild, M.F., S. Guoqing and Y. Shiren (1992) Development and Test of an Error Model for Categorical Data. *International Journal of Geographical Information Systems* 6(2): 87-104.

Goodchild, M.F., M. Yuan and T.J. Cova (2007) Toward a General Theory of Geographic Representation in GIS. *International Journal of Geographical Informational Science* 21(3): 239-260.

Hartshorne, R. (1939a) The Character of Systematic Geography. *Annals of the Association of American Geographers* 29(4): 413-436.

Hartshorne, R. (1939b) Is The Geographic Area A Concrete Unitary Object. *Annals of the Association of American Geographers* 29(4): 262-276.

Hartshorne, R. (1962) On the Concept of Areal Differentiation. *Professional Geographer* 14(5): 10-12.

Honeycutt, D.M. (1987) Epsilon Bands Based on Probability. *Abstract presented at Auto-Carto VIII, Baltimore, MD*, 30 March–2 April.

Johnston, R.J. (1968) Choice in Classification: The Subjectivity of Objective Methods. *Annals of the Association of American Geographers* 58(3): 575-589.

Kronenfeld, B.J. (2011) Beyond the Epsilon Band: Polygonal Modeling of Gradation/Uncertainty in Area-class Maps. *International Journal of Geographical Science* 25(11): 1749-1771.

Mark, D.M. and F. Csillag (1989) The Nature of Boundaries On 'Area-Class' Maps. *Cartographica* 26(1): 65-78.

Sauer, C.O. (1921) The Problem of Land Classification. *Annals of the Association of American Geographers* 11(1): 3-16.

Shafer, G. (1976) *A Mathematical Theory of Evidence*. Princeton, NJ, Princeton University Press.

Skidmore, A.K. and B.L. Turner (1992) Map Accuracy Assessment Using Line Intersect Sampling. *Photogrammetric Engineering and Remote Sensing* 58(10): 1453-1457.

Zadeh, L. (1978) Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems* 1: 3-28.

Zhang, J. and R.P. Kirby (1999) Alternative Criteria for Defining Fuzzy Boundaries Based on Fuzzy Classification of Aerial Photographs and Satellite Images. *Photogrammetric Engineering and Remote Sensing* 65(12): 1379-1387.

Zhang, J., M. Goodchild and P Kyriakidis (2007) A Conceptual Framework for Categorical Mapping and Error Modeling. *Acta Geodaetica et Cartographica Sinica* 3(8): 296-301.

Zobler, L. (1958) Decision Making in Regional Construction. *Annals of the Association of American Geographers* 48(1):140-148.

Michael N, DeMers, Professor, Department of Geography, New Mexico State University, Las Cruces, NM 88003