

## Exploration of Semantic Uncertainty in Parameterized Land Use/Land Cover Classifications

**Ola Ahlqvist and James Macgill**

GeoVISTA Center, Department of Geography  
The Pennsylvania State University  
State College, PA 16802, USA  
oka1@psu.edu, jmacgill@psu.edu

**Diansheng Guo**

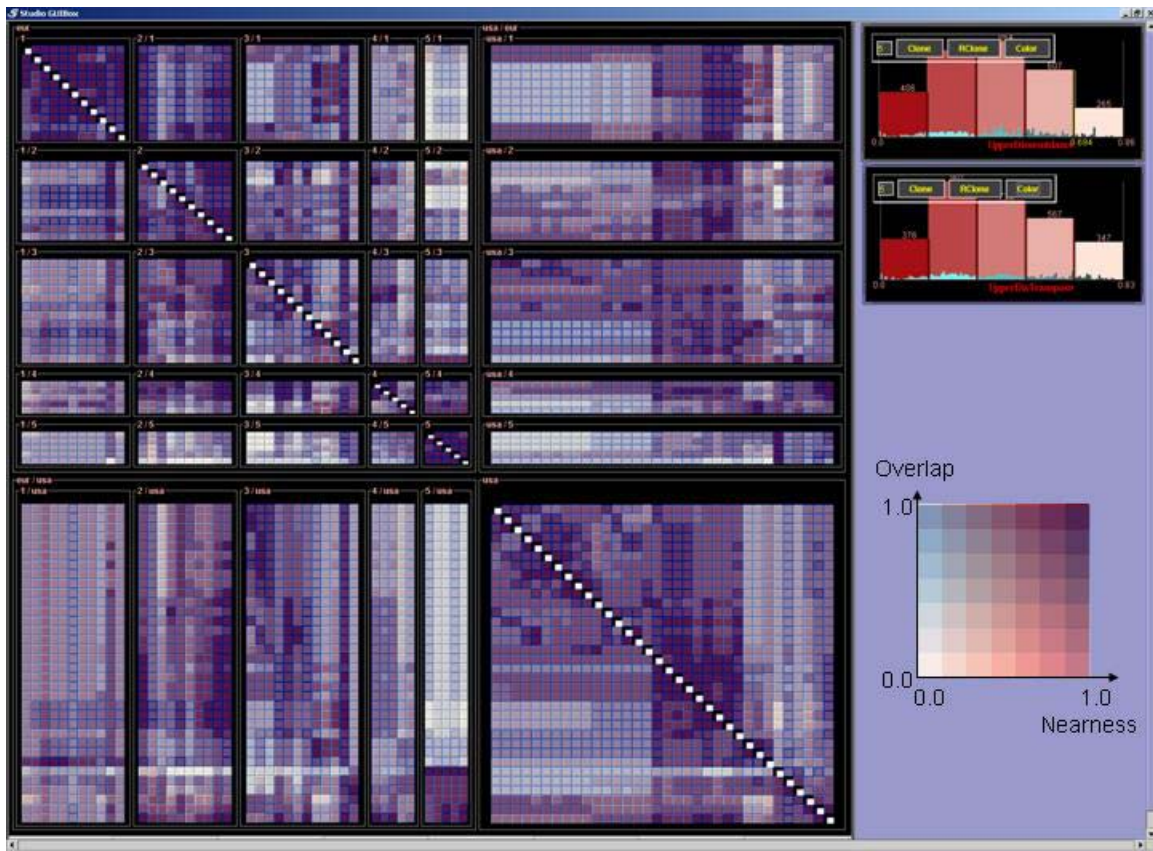
Department of Geography  
University of South Carolina  
Columbia, SC 29208, USA  
guod@sc.edu

Any attempts to use GIS to integrate data from environmental databases and to use models or analytical tools upon data need a full understanding of the origin and context of each data set used. Thus, taking data using different conceptualizations from different contexts, a successful GIS integration process relies upon a transformation of this information into the currently desired conceptualization and context. Problems with different data semantics when trying to mediate heterogeneous databases have stimulated research on interoperability (c.f. Goodchild et al. 1999).

Efforts to address semantic interoperability have proposed conceptual models, often grounded on some cognitive principles, using various representational approaches (Nyerges, 1991; Usery, 1996; Bishr, 1997; Mennis 2003). There are also some recent suggestions how to merge categorical systems or ontologies (Kavouras and Kokla, 2002; Worboys and Duckham, 2002, Feng and Flewelling, 2003), but these assume that crisp, atomic entities can be defined. Unfortunately geographic categories often possess aspects of semantic uncertainty (Salgé, 1995) such as vagueness or indiscernibility or both (Fisher and Wood, 1998; Bowker, 2000).

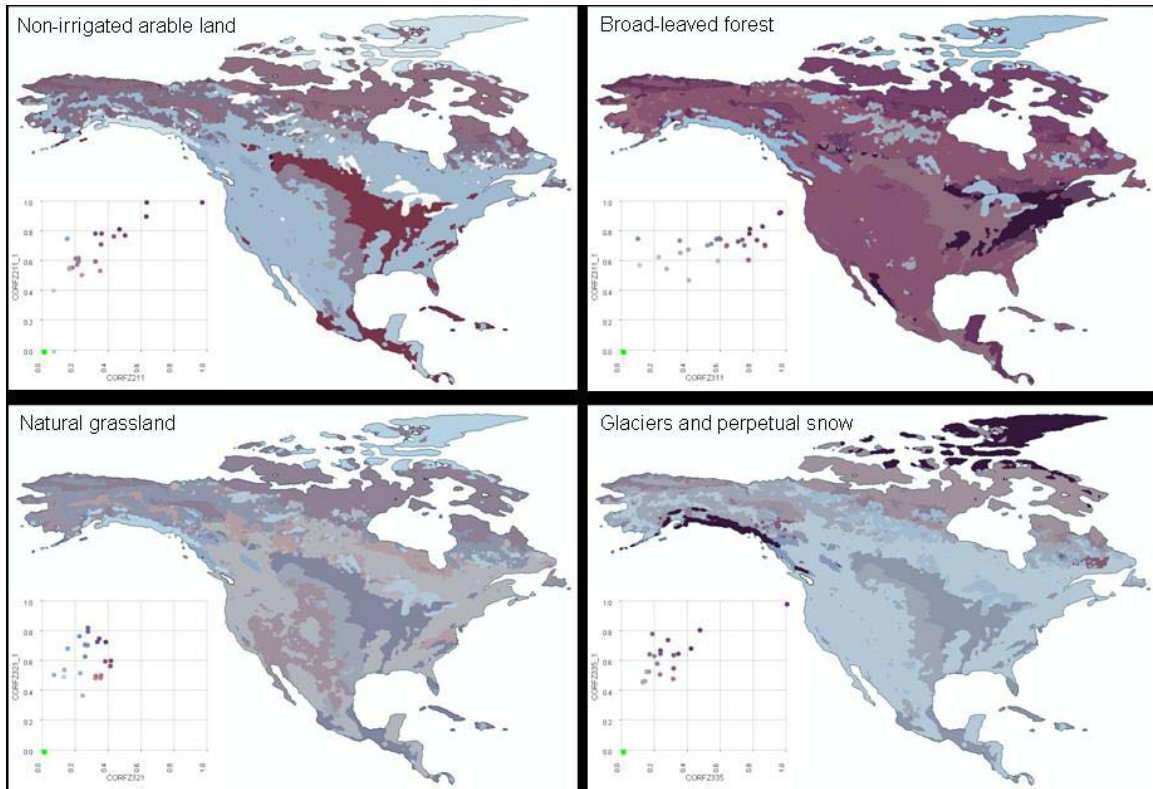
This work models geographic categories through parameterized concept definitions, similar to the frame-based object models in Faucher (2001) and Mennis (2003). But it differs in the way it synthesizes a hybrid representation of vague and granular semantic uncertainty (Ahlqvist et al, 2003) with a cognitive theory of conceptual spaces (Gärdenfors, 2000). Following Ahlqvist (in press) a category  $C$  is defined by three vectors  $(S_i, R_i, W_i)$  each holding a collection of approximation spaces  $S_i$  with corresponding property values  $R_i$  given as rough fuzzy set definitions, and accompanying salience weights,  $W_i$ . Using this model we give an example of interoperating land cover taxonomies used in the Land Cover Database of North America (Latifovic et al., 2002) and the European CORINE Land Cover (CEC, 1995). Both taxonomies are defined using the conceptual space representation, and we use metrics of semantic similarity to perform a translation between the two systems.

There are a number of approaches to estimate the similarity of categories. From psychology two common approaches uses either the proportion of shared features (Tversky, 1977) or the psychological distance between related properties (e.g. Nosofsky, 1986). Two recent examples of comparing geospatial entity classes using semantic similarity both used a shared feature approach (Feng and Flewelling, 2004; Rodriguez and Egenhofer, 2004). Our conceptual space representation interprets feature based similarity as a graded class-subclass relation or “overlap”, and distance based similarity as graded category “nearness”. Computation of similarity values for all possible pairs of categories in our study is summarized in a similarity matrix (Figure 1) where each cell has a value of similarity from comparing the column category with the row category.



**Figure 1** Example of semantic similarity matrix visualized in GeoVISTA studio (<http://www.geovistastudio.psu.edu>) where each cell has a value of similarity from comparing the column category with the row category. Low similarity values are shown in light shades and higher values with increasingly darker shades. The categories can be grouped to enhance readability of the matrix.

Two similarity matrices containing the two different metrics form the basis for exploring the semantic relationships categories in our study. Our illustrations demonstrate several exploratory techniques such as colored bi-variate large matrix displays (Figure 1) and bi-variate cartographic displays and scatter plots (Figure 2). The bi-variate color scheme was created by mixing a blue color scheme (for the overlap similarity) and a red color scheme (for the nearness similarity).



**Figure 2** Bi-variate similarity maps that display a combination of feature based and distance based similarity of some selected Corine land cover categories (CEC, 1995) with the categories in the Land Cover Database of North America (Latifovic et al., 2002). Scatter plots show the similarity values of each North American category compared to the chosen European category.

The techniques can be used to explore both intra- and inter-ontology semantic relationships. Intra ontology relationships can be used to discover the semantic separation between categories in a system as a way to predict possible confusion during classification. Inter ontology relationships are useful in efforts to translate between classification systems such as the example here.

One interesting feature of the used similarity metrics are their non-symmetric property, which basically say that the similarity of A to B is not necessarily the same as the similarity of B to A. This property poses a challenge from a visualization viewpoint because each comparison can be viewed from (at least) two different perspectives.

### Acknowledgements

This material is based upon work supported in part by the National Science Foundation under Grant No. BCS-9978052 and by the National Cancer Institute under Grant No. R01 CA95949-01.

## References

- Ahlqvist O., in press, A parameterized representation of uncertain conceptual spaces, *Transactions in GIS*.
- Ahlqvist O., Keukelaar J., and Oukbir K. 2003 Rough and fuzzy geographical data integration, *International journal of geographic information science*, 17(3): 223-234.
- Bishr Y. 1997 *Semantic aspects of interoperable GIS*, PhD Thesis, ITC Publications Series No. 56, Enschede, The Netherlands, 154p.
- Bowker G. C. 2000 Mapping biodiversity, *International Journal of Geographical Information Science*, 14 (8): 739-754.
- CEC (Commission of the European Communities), 1995, *CORINELand Cover*, 163p., available at <http://reports.eea.eu.int/COR0-landcover/en>
- Faucher C. 2001 Approximate knowledge modeling and classification in a frame-based language: The system CAIN, *International Journal of Intelligent Systems*, 16: 743-780
- Feng C.-C. and Flewelling D.M. 2003 Assessment of semantic similarity between land use/land cover classification systems, *Computers, Environment and Urban Systems*, 27: 229-246.
- Fisher P. and Wood J. 1998 What is a Mountain? – Or the Englishman who went up a Boolean Geographical Concept but Realized it was Fuzzy, *Geography*, 83(3): 247-256.
- Goodchild M. F., Egenhofer M. J. Fegeas R. and Kottman C. A. (Eds.) 1999 *Interoperating Geographic Information Systems*. Boston: Kluwer Academic Publishers.
- Gärdenfors P. 2000 *Conceptual spaces : the geometry of thought*, MIT Press, Cambridge, Mass., 307p.
- Kavouras M. and Kokla M. 2002 A method for the formalization and integration of geographical categorizations, *International journal of geographic information science*, 16(5): 439-453.
- Latifovic R., Zhu Z.-L., Cihlar J., Giri C. 2002 *Land cover of North America 2000*, Natural Resources Canada, Canada and US Geological Service EROS Data Center.
- Mennis J.L. 2003 Derivation and implementation of a semantic GIS data model informed by principles of cognition, *Computers, Environment and Urban Systems*, 27: 455-479.
- Nosofsky R.M. 1986 Attention, similarity, and the identification-categorization relationship, *Journal of Experimental Psychology: General*, 115: 39-57.
- Nyerges T.L. 1991 Geographic information abstractions: conceptual clarity for geographical modeling, *Environment and Planning A*, 23: 1483-1499.
- Rodríguez M.A. and Egenhofer M.J. 2004 Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure, *International Journal of Geographical Information Science*, 18(3): 229-256.
- Salgé F. 1995 Semantic accuracy, in: Guptill S.C. and Morrison J.L. (Eds), *Elements of spatial data quality*, 1st ed., Elsevier Science Ltd, Oxford: 139-151.
- Tversky A. 1977 Features of similarity, *Psychological review*, 84: 327-352.
- Usery E.L. 1996 A conceptual framework and fuzzy set implementation for geographic features, in: P.A. Burrough and A.U. Frank (Eds), *Geographic, Objects, With Indeterminate Boundaries*, Taylor & Francis, London, 345 p. : 71-85.
- Worboys M. F. and Duckham M. 2002 Integrating spatio-thematic information. In: Egenhofer M. J. and Mark D.M. (Eds.), *Geographic information science—second international conference, GIScience 2002* Berlin: Springer: 346–361.