Beyond the Surface: Current Issues and Future Directions in Uncertainty Visualization Research

Jennifer Smith*, David Retchless*, Christoph Kinkeldey, Alexander Klippel***

* The Pennsylvania State University, Geography - University Park, United States

** HafenCity University Hamburg, Lab for Geoinformatics and Geovisualization (g2lab) - Hamburg, Germany

Abstract. As people both reason and make decisions with uncertain geospatial data every day, it is important to understand the complexity of uncertainty, how it propagates through each dataset, and how to best visualize uncertainty to support reasoning and decision-making. When decisions are made from visualized geospatial data without the uncertainty explicitly mentioned or depicted with the dataset, it can lead to an inaccurate or misleading understanding of spatial patterns and processes. While this research area has expanded over the past two decades, much of the attention in evaluating uncertainty visualizations has been focused on identifying and measuring lower level perceptual processes. In contrast, we argue for a shift in attention towards assessing higher-level cognitive processes by keeping the focus on the user from the onset of visualization design to the final uncertainty visualization output. This article outlines two emerging research topics that will allow researchers to evaluate and promote these higher-level cognitive processes: identifying **appropriate levels of precision** when visualizing uncertainty and **supporting a deeper comprehension** of uncertainty. Currently, there is a lack of comprehensive empirical work that attempts to cognitively assess uncertainty visualization and decision-making through a human factors standpoint. A behavioral research approach addressing uncertainty classification and deeper uncertainty can provide a necessary and significant contribution for more intuitive representations in current uncertainty visualization and research.

Keywords: Uncertainty, Visualization, Decision-Making, Cognition

1. Introduction

Utilization of geospatial information within analysis and decision-making without attention to uncertainty makes the assumption that the data in hand is accurate and reliable (Beard et al., 1991). However, uncertainty is inherent in all geospatial data (Duckham et al., 2001), including outputs derived from a geographic information system (Hope & Hunter, 2007). As people both reason and make decisions with uncertain geospatial data every day, it is important to understand the complexity of uncertainty, how it propagates through each dataset, and how to best visualize uncertainty to support reasoning and decision-making. When decisions are made from visualized geospatial data without the uncertainty explicitly mentioned or depicted with the dataset, it can lead to an inaccurate or misleading understanding of spatial patterns and processes. Thus, recent efforts have attempted to support the decision-maker through integration of uncertainty in data visualization. Hunter and Goodchild (1993) state that lack of proper attention to uncertainty can lead to the "use of wrong data, in the wrong way, to arrive at the wrong decision" (p. 55). It is essential that researchers address this complex topic that propagates within all geospatial data and research in order to develop better solutions and visualizations to support the user.

Many uncertainty visualization techniques draw upon existing cartographic methods. However, due to a small number of empirical studies, little is known about the effectiveness of these techniques for visualizing data uncertainty. Additionally, whatever progress has been made on the empirical front has been split across several disciplines (MacEachren et al., 2005), and researchers have not been effective in systematically assessing and developing formal ontologies that put individual evaluations into perspective. This lack of empirical testing may partially be due to the numerous and conflicting definitions of uncertainty (Aerts, Clarke, and Keuper, 2003; Pang, Wittenbrink and Lodha, 1997). For instance, Deitrick and Edsall (2008) find that the term uncertainty is an issue across multiple disciplines, where it is often defined using numerous terms: data quality, accuracy, precision, error, vagueness, ambiguity, etc. For many authors, uncertainty is simply a configuration of many different views and definitions (Griethe & Schumann, 2005; Thomson et al., 2005) such as imprecision, error, subjectivity, and non-specificity. These different definitions are a result of the lack of a formal taxonomy or ontology that would synthesize the topic across various domains and contexts.

While this research area has expanded over the past two decades, much of the attention in evaluating uncertainty visualizations has been focused (perhaps unintentionally) on identifying and measuring lower-level perceptual processes. In contrast, we argue for a shift in attention towards assessing higher-level cognitive processes by keeping the focus on the user from the onset of visualization design to the final uncertainty visualization output. This article outlines two research topics that will allow researchers to evaluate and promote these higher-level cognitive processes: supporting a deeper comprehension of uncertainty and identifying appropriate levels of precision when visualizing uncertainty. This article is not meant to provide a comprehensive overview but rather serves to identify two emerging areas of uncertainty visualization.

2. Level of Precision

Traditionally, classification of geospatial data uncertainty is often chosen and visualized by the map maker based upon the characteristics of the data without lending proper attention to the natural classes users may have for uncertainty or their requirements for making a decision. Depending upon the context, variations in the level of precision (i.e. number of classes and their respective ranges) may be necessary to better support the user in a more informed decision-making process. Potential research questions in this area may ask: Are there natural levels of precision for conceptualizing about geospatial data uncertainty? What levels of uncertainty are important for the user in making decisions? At what point does increasing the number of categories not make a difference in how they make a decision? Do we alter user behavior if we change the contextual scenario? To assess these questions, researchers need to incorporate behavioral studies with the end user in mind.

Some users may prefer to reason about and make decisions with an unclassed color scheme (e.g. light to dark red representing lower to higher levels of uncertainty) that assigns a specific variation of the color to a single value of uncertainty. There is a tradeoff perceptually, however, as the points in the color spectrum can become impossible to discern. For example, as noted in Olson (1981), the use of more than nine categories in a bivariate choropleth color scheme quickly becomes confusing for the user. On the other hand, aggregating a range of uncertainty values into classes may not offer enough precision or the right distinctions the user needs to make a more informed decision. Ultimately, finding the right tradeoff will depend upon the requirements of the user. In Figure 1, Aerts, Clarke, and Keuper (2003) visualize uncertainty of urban growth using two different classification schemes (i.e. one contains two classes and the other five). Incorporating the end user may allow cartographers to choose the level of precision and classes in a way that is beneficial for the end user to reason and/or make decisions from the visual result.

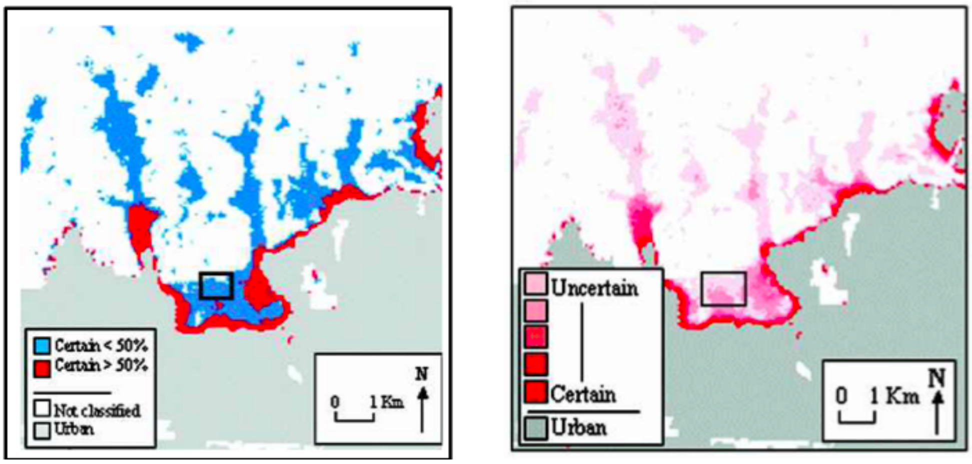

Figure 1. Images from Aerts, Clarke, and Keuper (2003) revealing two different classification schemes.

As a simple example, someone may be interested in the uncertainty associated with the chance of rain in order to decide whether to bring an umbrella the following day. Perhaps this person is only interested in two classes where 70% is the critical value separating both classes. In another scenario, imagine a wildfire threatening your property. What classes would you likely want in order to make a decision of whether you evacuate the property? These considerations suggest that users may appreciate uncertainty classifications that are flexible and interactive; in support of this notion, Kandlikar, Risbey, and Dessai (2005) note that “simple schemes that attempt to represent uncertainty in a uniform manner across many different contexts...may lead to biases...depending on how much detail is presented in the information” (p. 446). While several researchers have considered how the choice of classification system and color scheme can affect judgments and decisions made using maps (see, e.g., Brewer and Pickle 2002), more research is needed to extend these results to visualizations of uncertainty.

Uncertainty may also be implicit in maps that do not explicitly represent uncertainty information. For example, map users' understanding of the physical processes associated with a hazard might shape their understanding of how the certainty of this hazard varies across a map, even if the map does not explicitly show uncertainty information. Flood maps may be one case where uncertainty is implicit. As many people understand that locations closer to the water and have lower elevations are more likely to flood, they may intuitively assign higher certainty to flooding predictions for these areas than to adjacent areas that are higher or further from the water, even if the map shows both areas within the same hazard zone. Thus, uncertainty may already be implicit in many maps that do not explicitly include uncertainty representations. This implicit uncertainty may need to be quantified before the effect of explicitly visualizing uncertainty can be considered.

Classification of uncertainty for visualization may prove particularly challenging when the uncertainties that are being visualized are poorly defined, that is, when the probabilities or confidence intervals are themselves uncertain, or are completely unknown. The next section explores some of the challenges of visualizing these deeper uncertainties.

3. A Call For Visualizing Deeper Uncertainty

A large number of uncertainty visualization studies employ evaluations in order to assess the effectiveness of the visualization techniques. We argue, however, that a large majority only assess whether participants can attain a surface understanding of uncertainty. In this approach, visualization techniques are evaluated based upon a participant's map reading skills, where the user attempts to identify the level of uncertainty on a map by visually matching the map graphics with the corresponding values in the legend. These types of assessments do not evaluate whether users obtain useful knowledge or utilize higher-level cognitive processes important in decision-making scenarios. When faced with important decisions, it is still under determination as to whether users actually grasp the complexity of uncertainty, or the deeper implication and intricacies of the data. Future research should attempt to support a deeper comprehension of data uncertainty by transitioning beyond current surface visualization techniques and evaluation methodologies. This section outlines a few case studies that evaluate surface level tasks and presents more complex evaluations and graphics in other research areas that have helped users grasp these deeper levels of uncertainty.

In a study on visualizing uncertainty of urban growth, Aerts, Clarke, and Keuper (2003) recruited planners and decision-makers and asked them to make decisions based upon approximations and preferences related to the chance of urban growth. Uncertainty is displayed through color lightness for two separate display methods: static side-by-side comparison and animation. In this study, levels of uncertainty were aggregated and placed into set classes. On the simple approximations, participants only need to employ basic low-level perceptual processes to match colors between the map and legend to identify the corresponding uncertainty value. While users appear capable in associating particular numerical values with a specific level of uncertainty, it does not imply that they actually grasp the deeper meanings and intricacies of the data itself. Similarly, Hope and Hunter (2007) created a thematic map of land suitability for airport sites. Each areal region was shaded based on its suitability and an overlaid glyph bar indicated the amount of uncertainty. Users were tasked with deciding the optimal region for a new airport based on the combination of suitability and certainty, comparing two different areas. While this task forced participants to combine two types of data to make a decision, we argue, however, users still only need to perform a fairly simple map-reading task by matching the uncertainty representation with the value in the legend.

These studies highlight the need to move towards research that captures a more profound understanding of the complex interaction among diverse uncertainties, cartographic communication, and decision making -- in short, towards the study of what has been called ‘deep uncertainty’. In discussions of deep uncertainty, “deep” can refer to profound indeterminacy in the data itself, to irreducible ignorance about the data, or – in the context of uncertainty communication – to the complexity and completeness with which an audience understands information about uncertainty that has been communicated to them. Most discussions of deep uncertainty focus on ignorance or indeterminacy, both of which define limits to what can be known. In cases of indeterminacy, the fault is ontological (the true state of the world limits what can be known) while in cases of ignorance, the fault is epistemological (the limits of human reasoning bound our knowledge). According to Bankes (2002), the term “deep uncertainty” was first used by Nobel laureate and economist Kenneth Arrow in a talk on the Economics and Integrated Assessment of Climate Change that he gave at the Pew Center Workshop in 1999. While “deep uncertainty” is apparently of fairly recent coinage, the idea of multiple levels of uncertainty has a long history in economics, stretching back to at least Knight (1921), who distinguished uncertainty that is quantifiable using probability and statistics (which he termed “risk”) from uncertainty that is not measurable (which he considered uncertainty in the proper sense of the term). Perhaps following the lead of Dr. Arrow, much of the recent work on “deep uncertainty” has been in the field of climate change, with discussions considering ignorance and indeterminacy in both natural systems (Kandlikar et al. 2005) and human dimensions (Moser 2005). Researchers have also extended the discussion of deep uncertainty to other hazards and risks (see, e.g., Spiegelhalter and Riesh 2011).

Several researchers have explored the challenges inherent in the visual communication of uncertainty in general and deep uncertainty (ignorance and indeterminacy) in particular. For example, Spiegelhalter et al. (2011) suggest techniques for communicating probability using pie charts, bar charts, icon arrays and other techniques. In cases where probabilities are uncertain due to sampling error or other quantifiable limitations, they suggest using continuous fading or blurring to communicate this uncertainty. However, they also caution against trying to visualize truly deep uncertainty; noting that making a depiction of deep uncertainty more attractive may lead people to "believe it represents the whole truth rather than being a construction of limited knowledge and judgment," they contend that "the greatest challenge is to make a visualization that is attractive and informative, and yet conveys its own contingency and limitations" (Spiegelhalter et al. 2011, 1400). Similarly, Kandlikar et al. (2005, 449-51) divide uncertainty about probabilities into six different levels, ranging from situations where a "robust, well-defended probability distribution" can be identified to "effective ignorance"; intermediate levels include situations where minimum-maximum bounds, order of magnitude estimates, or the sign/trend for the probability distribution can be identified with confidence. As understanding of the probability distribution becomes less precise, the authors recommend a corresponding reduction in the granularity or dimensionality of the representation (e.g., from a continuous probability density function, to a confidence interval, to a single order of magnitude estimate). While these examples do not explicitly consider the communication of uncertain spatial data on maps or other geographic representations, some of their general guidance on how to visually depict various levels of uncertainty can be transferred to a cartographic context.

These attempts to formulate guidelines for communicating deep uncertainty do not explicitly consider how visualizations can be used to deepen understanding of uncertainty. The distinction between surface and deep understanding of uncertainty is closely associated with dual process theories of cognition, which suggest that humans process information using either heuristics and associations (quick and effortless but prone to errors and biases) or using a rule based, analytic approach (slow and effortful but less likely to commit logical errors). It also draws on levels of processing theory, which contends that deeper processing (which relates new information to prior knowledge) is more likely than shallower processing (which focuses on the surface or merely perceptual aspects of stimuli) to encourage engagement, understanding, and learning (Rapp 2005). While not explicitly addressing uncertainty, Rapp (2005) suggests that visualizations can encourage deeper understanding of complex scientific information – particularly when they are designed to be engaging and interactive.

In a study by Micallef et al. (2012), the authors create visual representations to help participants grasp a deeper and more complex understanding of uncertainty and probability. Through crowdsourcing, the authors gather participants' understanding of six representations to assist in Bayesian reasoning that include area-proportional Euler diagrams, frequency grids, and combinations of both. Each of the six representations visualize five aspects: 1) a woman with a positive mammography, 2) the population within this positive mammography that have breast cancer, 3) the population within this positive mammography that do not have breast cancer (false positive), 4) the population with a negative mammography who actually do have cancer (false negative), and 5) the rest of the population who correctly test negative for breast cancer. This multilevel representation allows users to better grasp the complexity of the context. More research is needed to transition useful visualizations like these into geospatial representations to assist users in gaining a deeper knowledge of uncertainty.

4. Conclusion

In a weekly uncertainty visualization reading group, we have identified two largely missing areas in research (as outlined in this paper) that can benefit current approaches to assess and create intuitive and helpful uncertainty visualizations to promote reasoning and decision-making. There is a lack of comprehensive empirical work that attempts to cognitively assess uncertainty visualization and decision-making through a human factors standpoint. A behavioral research approach addressing uncertainty classification from the user's requirements and point of view as well as promoting a deeper comprehension of uncertainty can provide a necessary and significant contribution. When faced with important decisions, it is still under determination as to whether users actually grasp the complexity of uncertainty information, or the deeper implication and intricacies of the data. Many of the attempts to formulate guidelines for communicating a deeper uncertainty do not explicitly consider how visualizations can be used to deepen understanding of uncertainty."It is important that techniques that aid probabilistic reasoning and decision making can benefit untrained people with different backgrounds and age ranges and, remain effective in situations where little time and attention are available. " (Micallef et al., p. 2542)

References

Aerts, J. C. J. H., Clarke, K. C., & Keuper, A. D. (2003). Testing popular visualization techniques for representing model uncertainty. Cartography and Geographic Information Science, 30(3), 249-261.

Bankes, S. C. 2002. Tools and techniques for developing policies for complex and uncertain systems. Proceedings of the National Academy of Sciences, 99 (90003):7263–7266.

Beard, M. K., Buttenfield, B. P. & Clapham, S. B. (1991). NCGIA Research Initiative 7: Visualization of Spatial Data Quality. Technical Paper 91-26, National Center for Geographic Information and Analysis.

Brewer, C. A. and Pickle, L.: Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series, Ann. Assoc. Am. Geogr., 92, 662–681, 2002.

Deitrick, S., & Edsall, R. (2008). Making Uncertainty Usable: Approaches for Visualizing Uncertainty Information, in Geographic Visualization: Concepts, Tools and Applications. In M. Dodge, M. McDerby & M. Turner (Eds.), Geographic Visualization: Concepts, Tools and Applications (pp. 277-291). Chichester, UK.: John Wiley & Sons, Ltd.

Duckham, M., Mason, K., Stell, J., & Worboys, M. (2001). A formal approach to imperfection in geographic information. Computers, Environment and Urban Systems, 25(1), 89-103.

Griethe, H., & Schumann, H. (2005). Visualizing Uncertainty for Improved Decision Making. Paper presented at the 4th International Conference on Perspectives in Business Informatics Research, Skovde, Sweeden.

Hope, S., & Hunter, G. J. (2007). Testing the effects of positional uncertainty on spatial decision-making. International Journal of Geographical Information Science, 21(6), 645-665.

Hunter, G. J., & Goodchild, M. F. (1993). Managing Uncertainty in Spatial Databases: Putting Theory into Practice. Journal of Urban and Regional Information Systems Association, 5(2), 52-62.

Kandlikar, M., J. Risbey, and S. Dessai. 2005. Representing and communicating deep uncertainty in climate-change assessments. Comptes Rendus Geoscience 337 (4):443–455.

Knight, Frank H. Risk, Uncertainty, and Profit. 1921. Library of Economics and Liberty. Retrieved April 10, 2013 from the World Wide Web: http://www.econlib.org/library/Knight/knRUP.html

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., & Hetzler, E. (2005b). Visualizing Geospatial Information Uncertainty: What we know Know and What We Need to Know. Cartography and Geographic Information Science, 32(3), 139-160.

Micallef, L., Dragicevic, P., & Fekete, J.-D. (2012). Assessing the Effect of Visualizations on Bayesian Reasoning Through Crowdsourcing. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2536 - 2545.

Moser, S. C. 2005. Impact assessments and policy responses to sea-level rise in three US states: An exploration of human-dimension uncertainties. Global Environmental Change 15 (4):353–369.

Olson, J. M. 1981. Spectrally Encoded Two-Variable Maps. Annals of the Association of American Geographers 71 (2):259–276.

Pang, A. T., Wittenbrink, C. M., & Lodha, S. K. (1997b). Approaches to Uncertainty Visualization. The Visual Computer, 13(8), 370-390.

Rapp, D. 2005. Mental Models: Theoretical Issues for Visualizations in Science Education. In Visualization in Science Education, Models and Modeling in Science Education., ed. J. Gilbert, 43–60. Springer Berlin Heidelberg. http://www.springerlink.com.ezaccess.libraries.psu.edu/content/mv61l837385h5531/abstract/ (last accessed 5 September 2012).

Spiegelhalter, D. J., and H. Riesch. 2011. Don't know, can't know: embracing deeper uncertainties when analysing risks. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 369 (1956):4730–4750.

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing Uncertainty About the Future. Science, 333(6048), 1393-1400.

Thomson, J. R., Hetzler, E. G., MacEachren, A. M., Gahegan, M. N., & Pavel, M. (2005). A Typology for Visualizing Uncertainty. Paper presented at the Conference on Visualization and Data Analysis, San Jose, California.