TOWARDS A SYSTEMS APPROACH TO THE VISUALIZATION OF SPATIAL UNCERTAINTY

Phaedon C. Kyriakidis
Department of Geography
University of California Santa Barbara
Ellison Hall 5710, Santa Barbara, CA 93106-4060
E-mail: phaedon@geog.ucsb.edu

Introduction

Uncertainty is endemic in spatial data due to the imperfect means of recording, processing, and representing spatial information (Zhang and Goodchild, 2002). The characterization (modeling and portrayal) of spatial uncertainty, as well as its propagation to geographical modeling and its impact on decision-making, has been identified as a critical research priority in Geographic Information Science (GIScience), e.g., UCGIS (1996).

Early cartographic work on the portrayal of spatial uncertainty focused on the adaptation of Bertin's visual variables (location, size, value, shape, texture, color and orientation) for representing uncertainty measures, such as standard errors of predictions in an interpolation setting, or of posterior probabilities of class occurrence in a classification setting. Bertin's variables, along with additional ones such as color saturation, were thus used in the context of static verity visualization, i.e., the simultaneous depiction of the original map product and its accompanying uncertainty (Beard and Buttenfield, 1991; MacEachren, 1992; Pang et al., 1997). Although the above research efforts developed appropriate cartographic techniques for graphically conveying spatial uncertainty in map products, no systematic framework exists that allows a *quantitative* evaluation of the impact of uncertain maps on vision-related tasks, and consequent decision-making.

Consider, for example, the construction of a precipitation map from a sparse set of rain gauge measurements. A typical GIS operation would consist of an interpolation of the sample values to create a map of precipitation predictions, along with a map of the standard error of the predicted values. Assume for the sake of the argument that the objective of spatial analysis is the visual assessment of uncertainty regarding mapped storm fronts, and that storms are defined as heavy precipitation events with rainfall exceeding a given threshold. A storm front in this case consists of a set of pixels at which predicted precipitation exceeds that threshold value, i.e., a storm front constitutes a boundary (edge) between areas of heavy rainfall and areas of little or no precipitation. Assessment of uncertainty regarding the existence of the front, i.e., the existence of a map feature, calls for evaluating the probability that all pixels within that potential front zone exceed *simultaneously* the precipitation threshold or not. Even if map users can visually determine areas on the map where uncertainty is large, a single precipitation map and the corresponding pixel-wise uncertainty measures cannot be readily used to characterize uncertainty regarding the storm front as a whole, unless one incorrectly assumes lack of spatial correlation.

Existing approaches of verity visualization do not explore (and consequently do not depict) spatial uncertainty incurred from spatial interactions. In other words, a map of the standard error of precipitation estimates at all pixels comprising a hypothesized storm front does not allow the quantification of uncertainty regarding the existence of that storm front, no matter how sophisticated a cartographic technique is employed for the depiction of such errors.

Consider, as another example, the case of a groundwater flow and contaminant transport model, which can be viewed as a system that operates on various inputs, e.g., reservoir architecture (including boundaries and faults), hydraulic conductivity, and contaminant source characteristics, to yield predictions, e.g., plumes of contaminants in the subsurface. The shape and extent of such plumes, as well as the location of their center of gravity, are uncertain, due to the uncertainty in the input values (also termed input parameter uncertainty). If one ignores for a moment model uncertainty, i.e., the fact that the groundwater flow and transport model at hand might not be an adequate representation of reality (e.g., in terms of physics), a complete characterization of the uncertainty in the characteristics of the predicted plumes calls for a characterization of the *joint* uncertainty in the input parameters. In addition, the impact of uncertain plume characteristics on the decision to locate or not a remediation well at a particular point, calls for repeated realizations of: (i) a possible set of parameter maps, i.e., a realization of a possible reservoir with its hydrogeological properties, (ii) a model evaluation on these parameter maps to produce a possible contaminant plume, (iii) an action of placing a well or not had this simulated contaminant plume be real, along with the (possibly monetary) consequence of such an action. By repeating the above steps in a Monte Carlo framework, one can obtain a distribution of possible actions, and consequently a distribution of possible losses or gains. The optimal decision would then correspond to that particular action that minimizes the expected (average) loss or maximizes the expected utility.

Most existing visualization approaches do not offer a formal means for quantifying the impact of uncertain spatial data to decision-making, since they focus on the visualization of a single error-prone map, and its associated uncertainty. What is needed is a set of alternative maps, a set of actions that depend on the visual perception of these maps, and a formal evaluation of the optimal decision after examining the consequence of each action.

It should be noted here that the impact of uncertain spatial information on decision-making depends on the particular application at hand, i.e., on the analysis objectives of each map user. Small errors in a digital elevation model (DEM), for example, are more consequential when determining a hydrographic network in low-relief areas than when determining watershed boundaries. No matter the map analysis objective, however, any such analysis requires a realistic numerical representation of the spatial distribution of the attribute of interest. A single precipitation map obtained, say, from interpolation of rain gauge measurements does not provide a realistic representation of the underlying spatial distribution of precipitation, and as a consequence suppresses the variability of, say, associated river-stage forecasts. In addition, when visualizing a smooth interpolated precipitation map, even when this map is accompanied by cartographic depictions of standard errors of interpolation, the user is not exposed to the correct spatial variability of precipitation as it is inferred from sample rain gauges.

A Research Agenda

To overcome the above limitations, it is proposed that the research frontier of spatial uncertainty visualization in GIScience be advanced through the synergy of spatial statistics, systems analysis, computer vision, and decision-theory. More precisely, a formal quantitative framework for the visualization of spatial uncertainty that builds on an analogy from engineering systems is advocated. A system is a model of some aspect or process of the real world, often approximated by a set of (possibly non-linear) mathematical equations. The system is excited by a set of inputs to produce a set of outputs (model predictions). In a similar fashion, a map user can be viewed as a system: his/her visual perception and cognition are extremely complex operations (typically non-linear and multi-resolution) that lead to various outcomes/decisions depending on the level of training, the conditions of interaction with the map, and purpose of map analysis. In this spatial metaphor of an engineering system, the input is the map product and the output is the spatial analysis, decision, or action based on that map product. In analogy with the engineering systems, quantification of the impact of an uncertain input map on vision-related tasks requires that these tasks be applied on a set of alternative inputs, all of which are processed by the map user to arrive at a set of potential answers. From this viewpoint, approaches for evaluating the impact of alternative depictions of a single uncertain map on decision-making via human subject testing, amount to replacing the unavailable realizations of input maps with realizations of model evaluations (each different human subject being essentially a "model run") for each depiction tested.

Previous work along the same lines of multiple map realizations includes Fisher (1993, 1996), Ehlschlaeger et al. (1997), and Davis and Keller (1997). None of these efforts, however, capitalized explicitly on the combination of disciplines outlined above and its formal application to visual processing. In addition, the stochastic models employed in the above approaches are computationally prohibitive for large grids, cannot reproduce complex spatial patterns, and do not reproduce known attribute measurements (possibly of different resolutions) at their sampling locations.

The components of the proposed research agenda include: (i) the generation of realistic simulated maps possibly of very large dimensions, e.g., millions of pixels, (ii) the minimization of the probability of simulating redundant realizations, (iii) the ranking of these simulated realizations prior to visualization, (iv) the application of computer vision algorithms e.g., edge detection and image segmentation, as proxies for early- and mid-level vision operations on each simulated realization, and the joint multivariate analysis of the resulting outcomes, e.g., size and shape of connected regions, and (v) the formal evaluation of the impact of spatial uncertainty on decision-making using decision-theoretical methods.

(I) Construction of realistic simulated maps

Given a stochastic model for the spatial distribution of the attribute of interest (inferred from sample measurements, related imagery and/or prior knowledge), the system (map user) should be excited by a suite of alternative synthetic maps (numerical spatial models), all of which are compatible with that stochastic model. Geostatistical simulation, e.g., Chilès and Delfiner (1999), Kyriakidis et al. (1999), can be employed to construct such alternative synthetic (yet as realistic as possible) realizations, which are constrained to reproduce a given set of measurements at their sampling locations, as well as their histogram and a model of their spatial correlation. In addition, possible relationships between the attribute of interest and ancillary indirect information (possibly available at different resolutions) can also be accounted for.

(II) Improving the representativity of simulated realizations

A realistic uncertainty assessment of a system's output is only possible if a representative sample (distribution) of alternative output values is available. To achieve such a representative sample, the system must be excited by a set of stimuli that span maximally the input state space, i.e., one needs a set of inputs that constitutes a representative sample of the input parameter population. In the context of uncertainty visualization, map users should be exposed to a set of simulated realizations that maximally span the set of possible attribute states, i.e., the possible spatial arrangements of attribute values in the study area. With simple random sampling Monte Carlo techniques, such a representative sample is achieved only by generating a large number of realizations, a task which is both time and computer resources intensive.

To address the above problem, efficient Monte Carlo simulation using an adaptation of Latin Hypercube (LH) sampling to a spatial setting can be employed to ensure that the finite set of alternative images span maximally the population of all possible images within the random field model adopted. This entails minimizing the generation of redundant realizations for a given set of possible such realizations.

(III) Ranking alternative realizations prior to visualization

The visualization of spatial uncertainty could be more efficient if the sequence of simulated realizations is ranked according to some meaningful criteria before presented to the map user. In temporal animation, for example, the natural ranking index is time, although other variables can also be used for ranking (DiBiase et al., 1992). In a stochastic simulation context, ranking cannot be simply performed by sorting the realizations according to their univariate statistics (mean or variance), because for large grids (and under ergodic conditions) all simulated realizations will have exactly the same statistics.

Alternative schemes for ranking multiple simulated realizations into an ordered sequence for visualization could include: (i) ranking according to their deviation from the ensemble average map, i.e., likelihood ranking, (i) ranking according to characteristics derived from low- or mid-level vision operations, such as edge detection, image segmentation, or from shape characteristics of objects found in the image, and (iii) a hierarchical application of schemes (i) and (ii), whereby realizations are first classified into rank classes and sorted according to their likelihood ranking, and then realizations within each rank class are further sorted according to the visual criteria defined in (i).

(IV) Uncertainty in early- mid-level vision operations

In all rigor, a bottom-up approach to the visualization of spatial uncertainty in a systems framework requires the availability of a mathematical model of the brain and its visual and cognitive functions, based on biological, psychological, and neuroscientific principles. Efforts towards such a mathematical model are underway (Wilson, 1999; Dayan and Abbot, 2002), but there is no existing one that accurately characterizes the enormous complexity of the joint activity of neurons and its relation to human visual perception and cognition. Instead, digital image processing and computer vision algorithms (Shapiro and Stockman, 2001; Forsyth and Ponce, 2002) can readily be used as quantitative proxies for early and mid-level human vision tasks. Edge detection (a low-level vision operation) and image segmentation (a mid-level vision operation identifying image components or subregions) can be performed on each simulated realization using traditional gradient operators or say image thresholding via morphological watershed segmentation (Gonzalez and Woods, 2002). This approach differs from the traditional use of stochastic simulation for uncertainty visualization that is limited to just a visual display of alternative realizations, or to the depiction of per-pixel summary statistics (mean and variance) of simulated values at each pixel. A notable exception is the work of Kao et al. (2001), which also proposed the display of characteristics of objects (connected pixels) derived from individual realizations of continuous spatial variables.

Quantitative visual uncertainty propagation can be accomplished by first applying the vision proxies on each simulated image, and then recording/analyzing their outputs, e.g., boundary lengths of resulting regions, shape characteristics of resulting spatial patterns/objects. The joint distribution of all such outcomes provides a quantitative evaluation of the uncertainty in vision-related tasks due to the uncertainty input maps.

V) Impact of spatial uncertainty to decision-making

The impact of uncertain maps on decision-making can be addressed within a decision-theoretical framework, e.g., Pratt et al. (1995), given a link between a suite of vision-related operations and a decision with a consequent loss or gain. The availability of such alternative actions and associated losses or gains, allows a quantitative evaluation of the expected value of ignoring uncertainty, which is formally defined as the expected difference between the consequences of an optimal (with respect to a given loss function) decision made after and before incorporating uncertainty in the analysis (Morgan and Henrion, 1990).

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. BCS 0322349. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

References

Beard, M.K., and Buttenfield, B.P. (1991): NCGIA Research Initiative 7: Visualization of spatial data quality, *NCGIA Technical Paper 91-26*, National Center for Geographical Information and Analysis, University of California, Santa Barbara

Chilès, J.-P., and Delfiner, P. (1999): *Geostatistics: Modeling Spatial Uncertainty*, John Wiley & Songs, New York

Davis, T.J., and Keller, P.C. (1997): Modelling and visualizing multiple spatial uncertainties, *Computers & Geosciences*, 23, (4), 397-408

Dayan, P., and Abbott, L.F. (2001): *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, MIT Press, Cambridge, MA

DiBiase, D., MacEachren, A.M., Krygier, J.B., and Reeves, C. (1992): Animation and the role of map design in scientific visualization, *Cartography and Geographic Information Systems*, 19(4), 201-214, 265-266

Ehlschlaeger, C.R., Shortridge, A.M., and M.F. Goodchild (1997): Visualizing spatial data uncertainty using animation, *Computers & Geosciences*, 23(4), 387-395

Fisher, P.F. (1993): Visualizing uncertainty in soil maps by animation, *Cartographica*, 30(2&3), 20-27

Fisher, P.F. (1996): Animation of reliability in computer-generated dot maps and elevation models, *Cartography and Geographic Information Systems*, 23(4), 196-205

Forsyth, D.A., and Ponce, M. (2002): *Computer Vision: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ

Gonzales, R.C., and Woods, R.E. (2002): *Digital Image Processing*, 2nd Edition, Prentice Hall, Upper Saddle River, NJ

Kao, D., Dungan, J.L., and Pang, A. (2001): Visualizing 2D probability distributions from EOS satellite image-derived data sets: A case study, in: *Proceedings of the IEEE Symposium on Visualization 2001*, San Diego, October 21-23

Kyriakidis, P.C., Shortridge, A.M., and Goodchild, M.F. (1999): Geostatistics for conflation and accuracy assessment of digital elevation models, *International Journal of Geographical Information Science*, 13(7), 677-707

MacEachren, A.M. (1992): Visualizing uncertain information, *Cartographic Perspectives*, 13, 10-19

Morgan, M.G., and M. Henrion (1990): *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, Cambridge

Pang, A.T., Wittenbrink, C.M., and Lodha, S.K. (1997): Approaches to uncertainty visualization, *Visual Computer*, 13(8), 370-390

Pratt, J.W., Raiffa, H., and Schlaifer, R. (1995): *Introduction to Statistical Decision Theory*, MIT Press, Cambridge, MA

Shapiro, L.G., and Stockman, G.C. (2001): *Computer Vision*, Prentice Hall, Upper Saddle River, NJ

University Consortium for Geographic Information Science [UCGIS] (1996): Research priorities for geographic information science, *Cartography and Geographic Information Systems*, 23(3), 115-127

Wilson, H. (1999): *Spikes, Decisions and Actions: The Dynamical Foundations of Neuroscience*, Oxford University Press, New York

Zhang, J., and Goodchild, M.F. (2002): *Uncertainty in Geographical Information*, Taylor and Francis, London