[CV-04-018] Representing Uncertainty

Abstract

Using geospatial data involves numerous uncertainties stemming from various sources such as inaccurate or erroneous measurements, inherent ambiguity of the described phenomena, or subjectivity of human interpretation. If the uncertain nature of the data is not represented, ill-informed interpretations and decisions can be the consequence. Accordingly, there has been significant research activity describing and visualizing uncertainty in data rather than ignoring it. Multiple typologies have been proposed to identify and quantify relevant types of uncertainty and a multitude of techniques to visualize uncertainty have been developed. However, the use of such techniques in practice is still rare because standardized methods and guidelines are few and largely untested. This contribution provides an introduction to the conceptualization and representation of uncertainty in geospatial data, focusing on strategies for the selection of suitable representation and visualization techniques.

Keywords: geospatial data uncertainty, imperfections in geographic information, map design techniques, mapping uncertainty, principles of map design, Uncertainty, uncertainty visualization

Author & citation

Kinkeldey, C., & Senaratne, H. (2018). Representing Uncertainty. The Geographic Information Science & Technology Body of Knowledge (2nd Quarter 2018 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2018.2.3

This Topic is also available in the following editions:

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Visualization of uncertainty. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

Explanation

1. Definitions
2. Introducing the Concept of Uncertainty
3. Representing Uncertainty

1. Definitions

accuracy: the degree of closeness between observation and reality

adjacent view: a representation strategy that uses separate views for data and uncertainty, typically a side-by-side (map pair) view (opposed to "coincident view")

coincident view: a representation strategy that integrates the representation of uncertainty into the same view as the data (opposed to "adjacent view")

completeness: the extent to which information is comprehensive

consistency: the extent to which information components agree

credibility: the reliability of information source

currency: the time span from occurrence through information collection/processing to use

dynamic view: a representation strategy that uses animation and/or interaction to depict uncertainty (opposed to "static view")

error: the difference between observation and reality

extrinsic mapping: a representation strategy that maps uncertainty independently from the data, typically as symbols, glyphs, isolines or similar (opposed to "intrinsic mapping")

interrelatedness: the source independence from other information

intrinsic: a representation strategy that represents uncertainty by manipulating attributes of existing visual objects, typically their color or transparency (opposed to "extrinsic mapping")

lineage: the conduit through which information has passed

precision: the exactness of measurement/estimate

static view: a representation strategy that does not use animation and/or interaction to depict uncertainty (opposed to "dynamic view")

subjectivity: the extent to which human interpretation or judgment is involved in information construction

uncertainty: the difference between a real geographic phenomenon and the user's understanding of the geographic phenomenon

UVis3 (uncertainty visualization cube): categorization for uncertainty visualization defining the three axes "intrinsic/extrinsic," "coincident/adjacent," and "static/dynamic."

2. Introducing the Concept of Uncertainty

Uncertainty is a concept that emerged from research on geospatial data quality. In the early 1990s, the NCGIA (National Center for Geographic Information and Analysis) Research Initiative on "Visualization of Spatial Data Quality" addressed the need for exploration, development, and evaluation of visual techniques to communicate uncertainty (Beard et al, 1991). The representation of uncertainty has become a topic of interest in different visualization domains such as geographic visualization, information visualization, and scientific visualization, as well as in related domains such as Visual Analytics and decision science.

Several definitions of uncertainty exist within the GIS&T domain and the distinction is fuzzy

between related concepts such as data quality or reliability. Here, we adopt the definition by Longley et al. (2005): **uncertainty** is the difference between a real geographic phenomenon and the user's understanding of the geographic phenomenon. It is distinct from concepts such as **error** or **accuracy** because it acknowledges an unknowable component of all geospatial information. For instance, with an elevation measurement via GPS at a certain point in nature we can assume that there is always some difference between the measurement and the real value. If we know the real value, which is often the case, we can compute the error, which may be 10 m. If we do not know the real value we can apply the concept of uncertainty and estimate that the deviation lies within a certain range, for example, +/- 10 m.

There are many sources for uncertainty of geospatial information. For instance, we work with models of reality that are inherently uncertain (e.g., the unavoidable inaccuracy of interpolated values in an air temperature map) and most data we use come from measurements that are always subject to uncertainty (e.g., incompleteness of census data). In addition, the nature of the data sources can also play a role in causing uncertainty in the collected data, e.g., authoritative data collected by experts following standardized procedures and non-authoritative data collected by volunteers using methods that may not adhere to standardized protocols. In order to make such inherently uncertain data usable for different applications, these uncertainties need to be adequately identified, estimated, and represented. For further information on the concept of uncertainty, see **Modeling Uncertainty**, **Error-based Uncertainty**, and **Spatial Data Uncertainty**.

Various typologies for uncertainty have been defined to facilitate its use in practice. The MacEachren et al. (2005) typology, extending Thomas et al. (2005), has received particular attention in cartography and visualization for broad application to locational, temporal, and attribute uncertainties in geospatial data (after Sinton, 1978). It lists nine types of uncertainty (accuracy/error, precision, completeness, consistency, lineage, currency, credibility, subjectivity, and interrelatedness) providing examples for the locational, temporal, and attribute components of geospatial information listed above (**Table 1)**. Recently, Senaratne et al. (2017), after conducting an extensive review on uncertainty and its assessment methods, extended these uncertainty types to also include semantic accuracy, usage, trust, content quality, vagueness, local knowledge, experience, recognition, and reputation. This shows that evolving data and user requirements may call for further types of uncertainty in the future (see **Section 3.3**).

Table 1. Typology of uncertainty of geospatial information (reproduced with permission from MacEachren et al, 2005).

Category	Attribute Examples	Location Examples	Time Examples
Accuracy/error	counts, magnitudes	coordinates, buildings	+/- 1 day
Precision	nearest 1000	1 degree	once per day
Completeness	75% of people reporting	20% of photos flown	2004 daily / 12 missing
Consistency	multiple classifiers	from / for a place	5 say Monday; 2 say Tuesday
Lineage	transformations	#/quality of input sources	# of steps
Currency	census data	age of maps	C = Tpresent - Tinfo

Category	Attribute Examples	Location Examples	Time Examples
Credibility	U.S. analyst interpretation of financial records <...> informant report of financial transaction	direct observation of training camp <...> e-mail interception with reference to training camp	time series air photos indicating event time <...> anonymous call predicting event time
Subjectivity	fact <...> guess	local <...> outsider	expert <...> trainee
Interrelatedness	all info from same author	source proximity	time proximity

3. Representing Uncertainty

Since geospatial information uncertainty can have a strong impact on the outcome of analyses, interpretations, or decisions, representing uncertainty on maps can help users to reduce errors, derive more trustworthy results or acknowledge the insufficiency of data to avoid misinterpretation or ill-informed decisions. Yet communicating uncertainty to the user can be a challenging task since it adds dimensions to the underlying data and therefore increases complexity. **Uncertainty representation** can be a means to counter this challenge by using visual encodings to communicate uncertainty about information, presenting an often easier to understand reference for uncertainty than tables or numbers alone. However, creating representations of uncertainty can be a complex procedure due to reasons such as the heterogeneity of data, spatial and temporal variation of data, different types of uncertainty in data, and also the various abstract definitions of uncertainty measures that are adapted to suit the context of data usage (Gerharz et al., 2012). The remainder of this section provides an overview on techniques to represent uncertainty visually and discusses the intuitiveness of visual encodings as well as data and user requirements.

3.1 Categorization of uncertainty representation techniques

A wide range of approaches for visual representation of uncertainty exist and it can be challenging to select techniques suitable for specific applications and user tasks. One way to provide guidance with this procedure is to use typologies that help categorize approaches and techniques. Similar to typologies to organize different types of uncertainty (see **Section 2**) there are typologies for uncertainty visualization that suggest techniques for certain types of data (areas / points / lines, discrete / continuous etc.) and types of uncertainty (e.g., Buttenfield & Weibel, 1988; Pang et al., 1997; Sanyal et al., 2009). Kinkeldey et al. (2014) outline a framework called UVis3 (uncertainty visualization cube) that describes all uncertainty representation approaches based on three design axes: (1) intrinsic / extrinsic, (2) coincident / adjacent, and (3) static / dynamic (**Figure 1**). Each corner of the cube represents a combination of the three categories.

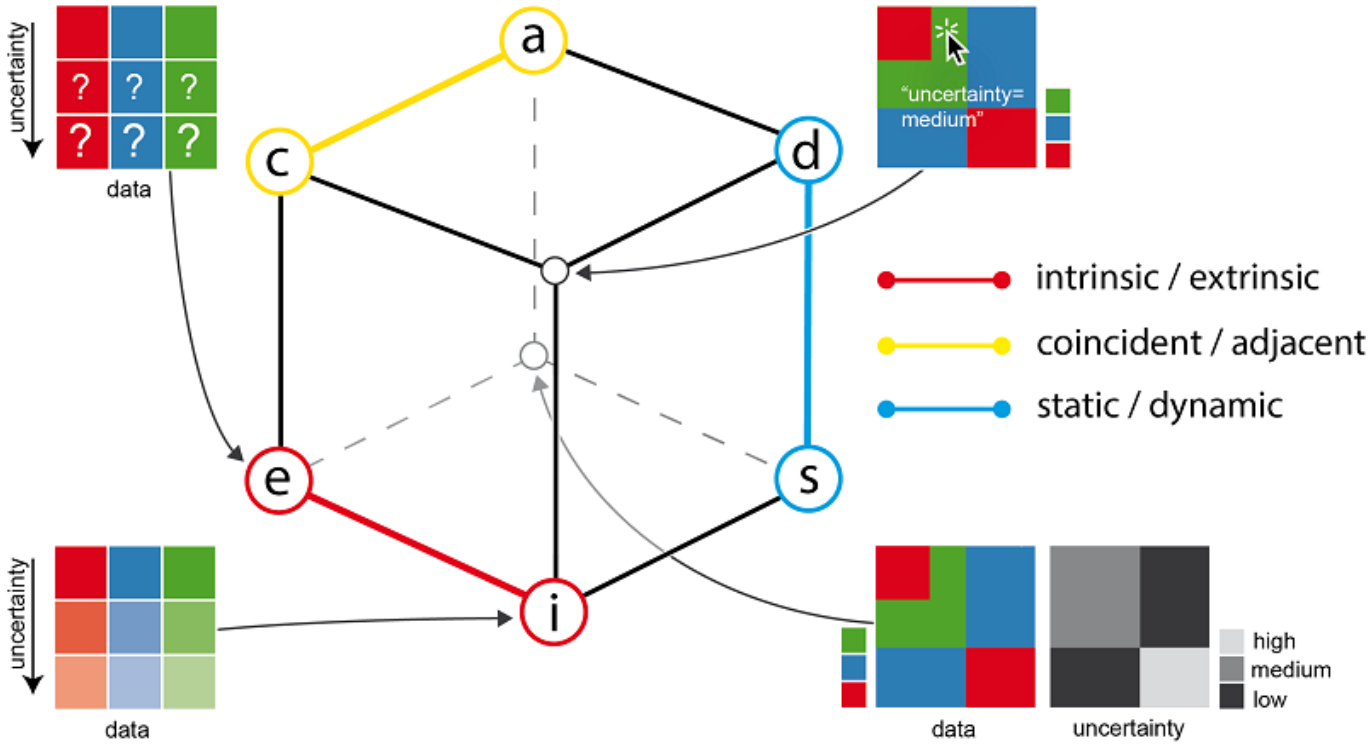


Figure 1: UVis3 (uncertainty visualization cube), extended from Kinkeldey et al. (2014), including four examples: representation of uncertainty by symbol size (top left), color value (bottom left), gray scale in a side-by-side view (bottom right), and by an invisible layer providing uncertainty retrieval through user interaction (top right).

- **Intrinsic / extrinsic** describes whether uncertainty is represented as an attribute of existing visual objects (intrinsic, e.g., by manipulating their color) or in the form of additional objects (extrinsic, e.g., symbols expressing the uncertainty). Common intrinsic techniques are color value / lightness / brightness (lighter color = higher level of uncertainty), transparency (more transparent = higher level of uncertainty) or saturation (less saturated = higher level of uncertainty) (see **Symbolization & the Visual Variables**). Apart from color, the use of fuzziness (blur) to represent uncertain areas (blurry fill pattern) or boundaries (blurry line) falls into the 'intrinsic' category, as does texture (dashed or 'sketchy' lines, hatched areas or similar). Extrinsic techniques include overlaid symbols (crosshairs, emoticons), glyphs (crosses, circles), or isolines. Figure 1 shows an example for an intrinsic technique in the bottom left (color value represents the level of uncertainty) and an extrinsic technique in the top left (size of question marks represents the level of uncertainty).
- **Coincident / adjacent** describes whether the information about uncertainty is mapped to the same view as the data (coincident) or to an additional view (adjacent). An example for coincident views are bivariate maps that show the relationship between data and uncertainty (see **Bivariate & Multivariate Mapping**). Adjacent views are, for instance, side-by-side views or small multiples (see **Mapping Time**). The bottom right example in Figure 1 is a simplified adjacent view where the data view on the left remains unchanged and the related uncertainty is represented in the right view using a gray scale.
- **Static / dynamic** describes whether animations and / or interactive means are used to represent uncertainty. The most common are static visualizations but a number of approaches use animation (e.g., showing an uncertainty layer alternating with a data layer) or interaction (e.g., uncertainty can be retrieved by clicking on an area in a

map, see the top right example in Figure 1) (see **UI/UX Design**).

The categories of the UVis3 help to organize existing approaches and can serve as a starting point when choosing suitable uncertainty representation techniques in a design process (see **Section 3.3**). Clearly, the choice depends on the data and user requirements. In general, the simplest approach is preferable and that is why many visual uncertainty representations use intrinsic, coincident, and static displays (i.e., non-animated, non-interactive visualizations of uncertainty integrated into the existing display). But, as mentioned above, there are alternatives:

- Extrinsic approaches such as symbols representing uncertainty can work very well with clearly defined areas as in choropleth maps. But they tend to focus the eye to the uncertain areas, more than most intrinsic approaches.
- Adjacent views can be a viable alternative using the advantages of multiple displays to prevent clutter when page or screen real estate is not a concern. At the same time, they take up more space and require more eye movement (Kinkeldey et al., 2014).
- Static approaches can be printed and used on paper but dynamic approaches are often necessary due to the visual complexity of many maps that makes static representations of uncertainty difficult. The examples in Figure 1 can help with the choice of a suitable approach.

3.2 Intuitive visual encodings

As mentioned above, a wide range of techniques have been proposed to represent uncertainty in maps and generally, any kind of visual variable can be used (e.g., color hue / value / saturation, location, shape, etc.). However, some visual variables more intuitively connote certainty versus uncertainty than others. To this end, several scholars have proposed selecting visual variables that evoke an uncertainty metaphor (e.g., MacEachren, 1992; McGranaghan, 1993), such as

- color value (colors “fade out” with higher levels of uncertainty),
- color saturation / purity (“gray out” effect imitates fog covering uncertain objects),
- fuzziness / blur (uncertain objects appear “out of focus” as with a camera lens),
- grain / resolution (uncertain objects appear coarser than others, similar effect to "out of focus"),
- texture (noisy or sketchy texture signifies uncertainty), or
- location (non-matching locations of objects represent uncertainty).

In an extensive user study by MacEachren et al. (2012) on intuitiveness of visual encodings for uncertainty visualization, most of the visual variables that participants found intuitive were linked to a visual metaphor (with fuzziness, location, and color value ranked highest, see **Figure 2**). Color saturation, however, although often used in uncertainty visualizations due to its "gray out" effect, is not among the most intuitive variables. Apart from this, drawbacks of color saturation include the lack of readability with less saturated colors (that become hard to distinguish), as well as a lack of intuitiveness because users are not always sure if unsaturated colors depict high or low levels of uncertainty (Kinkeldey et al, 2014). This shows that, even if not all visual encodings from the above list are necessarily the best choices with respect to readability, they should be taken into account since an intuitive representation of uncertainty should be the goal.

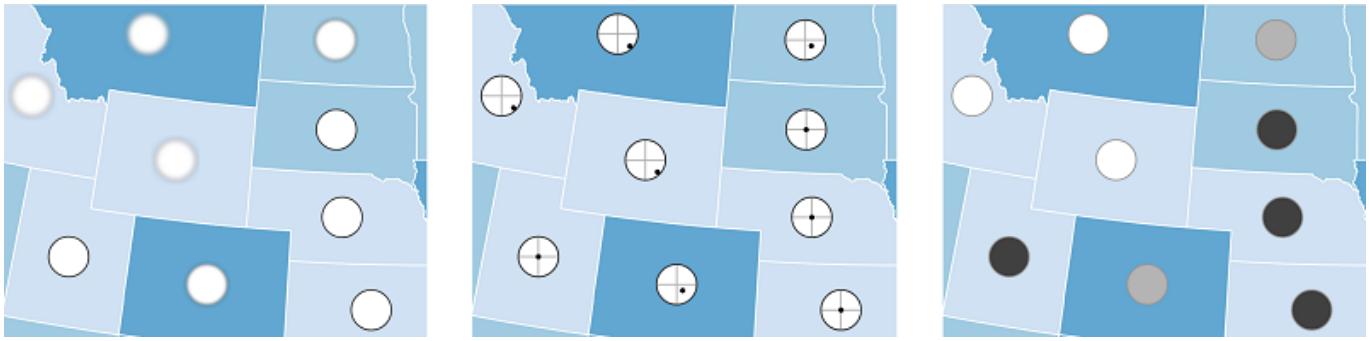


Figure 2: Examples showing the use of visual variables rated as highly intuitive for representation of uncertainty (MacEachren et al., 2012): fuzziness / blur (left), location (center), and color value (right).

3.3 Data and user requirements

It seems obvious that for the choice of uncertainty visualization techniques, the dimensionality of the data plays an important role: point data require different techniques than line or polygon data, and with discrete data, such as the population in each state of the US, other approaches may be more successful than with continuous data such as the population density over the whole country.

Other crucial requirements that are often neglected are those pertaining to the user (see **Usability Engineering & Evaluation**). Different types of users can have different tasks in mind, e.g, an expert user with domain knowledge may want to freely explore the data whereas other users may have predefined questions they would like to answer. For a low-level task such as the search for extreme values in an area (e.g., the maximum level of particulate matter air pollution in a city), the uncertainty caused by a lack of data accuracy may be more relevant than for high-level tasks (the decision if and where measures should be taken to reduce air pollution), where the lack of trustworthiness of the data may play a more important role. Data and user requirements are intertwined (the data must fit the user requirements in the first place) and should be taken into account before choosing an uncertainty representation technique.

Since there are no standardized procedures for leveraging uncertainty when working with geospatial information, it is essential to include potential target users in the creation process. This can help to ensure that visual representations are understandable and that it is clear to the user what type(s) of uncertainty is/are described. These are crucial requirements for designing representations that successfully support users in uncertainty-aware data analysis and decision making.

References

Beard, M., Buttenfield, B., & Clapham, S. (1991). NCGIA Research Initiative 7: Visualization of Spatial Data Quality, Scientific Report for the Specialists Meeting, NCGIA Technical Paper, 91-26, Santa Barbara, California.

Buttenfield, B. P., & Weibel, R. (1988). Visualizing the Quality of Cartographic Data. In Third International Geographic Information Systems Symposium (GIS/LIS '88), San Antonio, TX.

Gerharz, L. E., Autermann, C., Hopmann, H., Stasch, C., & Pebesma, E. (2012). Uncertainty visualisation in the model web. In EGU General Assembly Conference Abstracts, 14, 4421.

Kinkeldey, C., MacEachren, A. M., & Schiewe, J. (2014). How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies. The Cartographic Journal 51(4), 372-386.

Kinkeldey, C., MacEachren, A. M., Riveiro, M., & Schiewe, J. (2015). Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. Cartography and Geographic Information Science 44(1), 1-21.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). Geographic Information Systems and Science (2nd edition). Chichester: Wiley.

MacEachren, A. M. (1992). Visualizing Uncertain Information. Cartographic Perspectives, 13, 10-19.

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

MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual Semiotics & Uncertainty Visualization: An Empirical Study. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2496-2505.

McGranaghan, M. (1993). A Cartographic View of Spatial Data Quality. Cartographica, 30, 8-19.

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

Sanyal, J., Zhang, S., Bhattacharya, G., Amburn, P. & Moorhead, R. (2009). A User Study to Compare Four Uncertainty Visualization Methods for 1D and 2D Datasets. IEEE Transactions on Visualization and Computer Graphics 15(6), 1209-1218.

Senaratne, H., & Gerharz, L. (2011). An Assessment and Categorisation of Quantitative Uncertainty Visualisation Methods for Geospatial Data. In 14th AGILE International Conference on Geographic Information Science - Advancing Geoinformation Science for a Changing World.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. International Journal of Geographical Information Science 31(1), 139-167.

Sinton, D. F. (1978). The inherent structure of information as a constraint to analysis:

mapped thematic data as a case study. In Harvard Papers on Geographic Information Systems, Edited by: Dutton, G. Reading, M. A.: Addison-Wesley. Volume 6

Thomson, J., Hetzler, E., MacEachren, A., Gahegan, M., & Pavel, M. (2005). A typology for visualizing uncertainty, In Proc. SPIE 5669, Visualization and Data Analysis 2005, 145-157.