

# The Nature of Geospatial Information Uncertainty

Robert E. Roth | [reroth@psu.edu](mailto:reroth@psu.edu)

## Synopsis

The following entry provides a primer on geospatial information uncertainty, important background information for proper representation of and interaction with uncertain geospatial information. This entry begins with an introduction to the origins of geospatial information uncertainty, arguing that uncertainty is present in all geospatial information and therefore should be made usable through cartographic representation. I then define what I mean by ‘uncertainty’ and discuss two important subdivisions within uncertainty that are necessary for understanding its nature: uncertainty filters and uncertainty categories.

Folks who have inspired much of my thinking on this: [Barbara Buttenfield](#), [Helen Couclelis](#), [Rob Edsall](#), [Michael Goodchild](#), [Gary Hunter](#), [Alan MacEachren](#), [A-Xing Zhu](#)

## Keywords

uncertainty, abstraction, quality, fitness-of-use, uncertainty filter, uncertainty category, accuracy, completeness, consistency, credibility, currency, errors, interrelatedness, lineage, precision, resolution, subjectivity

## Maps are Models

*“We actually made a map of the country on the scale of a mile to a mile!” [said Mein Herr]*

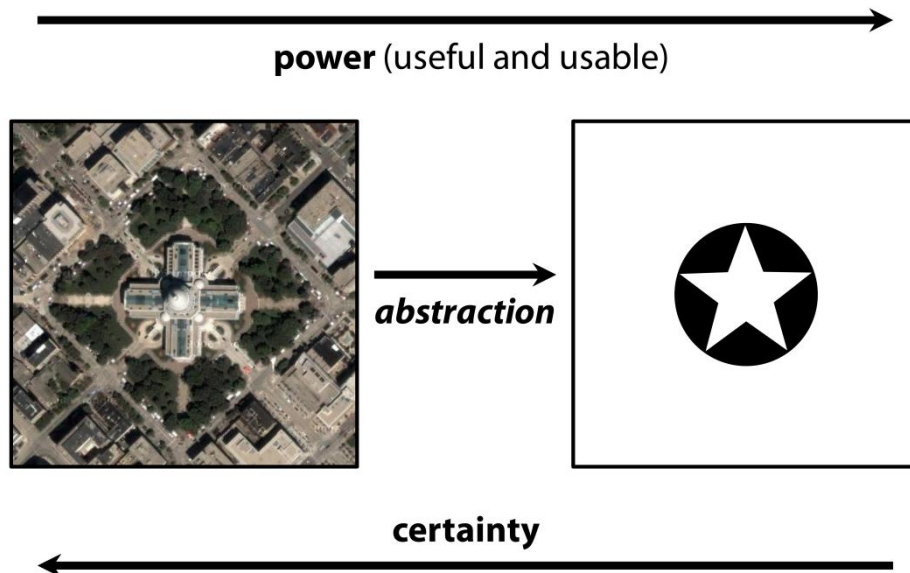
*“Have you used it much?” I enquired.*

*“It has never been spread out, yet,” said Mein Herr: “the farmers objected: they said it would cover the whole country and shut out the sunlight! So we now use the country itself, as its own map, and I assure you it does nearly as well.” (Carroll 1893\1982, 726.)*

The above Lewis Carroll passage has been used *ad nauseum* as a lighthearted way to introduce the discipline of Cartography. At the expense of being a product of this promulgation, I tend to agree with my predecessors that there is no singular passage that better demonstrates both the central need for Cartography and the fundamental dilemma presented to cartographers.

Maps work because they are models of reality, not reality itself. A map must necessarily overlook nearly every aspect of reality in order to be understandable, usable, and useful. As noted in the Carroll passage, a map that does not abstract reality is no more useful than interacting with the world itself. This is the reason that hyper-realistic, 3D views [\[update link\]](#) hinder communication without provision of extensive interactive tools. Thus, Cartography and cartographers are needed in order to perform this conceptual and graphical generalization [\[update link\]](#). It is abstraction that gives the map its power.

Unfortunately (and what your financial planners may not have told you...) there is an obvious downside to all models: they are not reality. The double-edged sword of abstraction removes geospatial information from the display that may be requisite for a clear and comprehensive understanding of geographic phenomena and processes. Tasks employing the representation or decisions informed by the representation can no longer be accomplished with complete certainty. I call this (in a much more academic-y way than Carroll) the *cartographic problematic* (Figure 1): in order to create an abstraction of reality that makes complex geospatial information understandable and usable, uncertainty is introduced into the representation (and into the knowledge constructed from the representation) as a necessary compromise. Uncertainty is therefore inherent to all geospatial information and thus all representations thereof. No map is perfect.



**Figure 1: The Cartographic Problematic.** In order to create an abstraction of reality that makes complex geospatial information understandable and usable, uncertainty is introduced into the representation (and into the knowledge constructed from the representation) as a necessary compromise.

Cartographers have been grappling with this issue through the history of Cartography. Usually, their solution is to ignore the uncertainty they are introducing during abstraction, assuming that the map user will immediately understand the limitations imposed by abstraction. The resulting maps appear to be pillars of objectivity and veracity, invoking trust from all of their users. For many map uses, large amounts of hidden uncertainty are acceptable and, again, necessary for effective communication. But for a special subset of instances – particularly those when the map is being used in support of a decision with grave consequences – it is essential that the map user has an idea of the map’s certainty. In these situations, I agree with [Stephanie Deitrick and Rob Edsall \(2008: 278\)](#) that “uncertainty in geographic data should be made *usable* through innovative visualization techniques.” Provision of uncertainty representations, and interfaces to manipulate these representations, provides users a better informed view of their geospatial information. This ultimately leads to better informed decisions.

## What is uncertainty?

Cartographers have been studying issues of quality for some seventy-five years. *Quality* is often described as the fitness-of-use of data or information for a given purpose. This focus is important from a practical standpoint, but problematic from a theoretical perspective in three ways. First, it treats imperfection as the enemy. It is the responsibility of the cartographer to purge all imperfection during information assembly. This perspective assumes that all uncertainty can and should be removed; the above discussion demonstrates why this is not possible. Second, this focus treats measures of imperfection solely as a quality check for use by the cartographer prior to representation, not as useful metadata for interpreting the map during use. Such an approach is like a car manufacturer installing a speedometer only to ensure that the car can reach the advertised maximum speed, then removing the speedometer before selling the car! Finally, quality is dependent upon the end use of the map. The same information can be fit for use in one scenario (e.g., walking through an amusement park) and not fit for use in another (e.g., evacuating the amusement park during an emergency).

Drawing from literature in risk assessment, [Alan MacEachren \(1992\)](#) recommended the term ‘uncertainty’ as an alternative. *Uncertainty* is defined as the difference between a real geographic phenomenon and the user’s understanding of the geographic phenomenon ([Longley et al. 2005](#)). I particularly like this definition because it appropriately places the user (not the dataset or the cartographer) in the center [\[update link\]](#). Also, this is a purposefully broad or catch-all approach to geospatial information uncertainty. Much work has been conducted to

further divide the concept of uncertainty or distinguish it from related concepts such as *ambiguity* (differing perceptions about or imperfect indicators of a geographic phenomenon), *fuzziness* (the degree to which a geographic phenomenon belongs to a particular definition, class, or other attribute), *vagueness* (inherently inexact definition of a geographic phenomenon or its attributes) and quality (as described above). In the following, I describe two such subdivisions I find essential for a complete understanding of uncertainty: (1) how uncertainty enters into the user's understanding of a geographic phenomenon (through what are termed uncertainty filters) and (2) the forms that uncertainty takes in geospatial information (termed uncertainty categories).

## Uncertainty Filters: How uncertainty enters into the user's understanding

A first important subdivision of uncertainty is the different stages at which uncertainty can enter into analysis, all of which contribute to a mismatch between the user's understanding of reality and reality itself. To help explain the nature of uncertainty, Longley and colleagues (2005) introduce the concept of uncertainty filters. The authors typify the geographic analysis in five iterative steps: (1) the real world, (2) conception, (3) measurement and representation, (4) analysis, and (5) interpretation, validation, and exploration. Between each step, there is an *uncertainty filter* that acts to remove direct correspondence between phenomenon and the user's understanding of the phenomenon.

When explaining the concept of uncertainty filters, I often make slight changes to how the steps are phrased in the Longley et al. (2005) graphic. This adaptation (Figure 2) better mirrors broader discussions in Cartography, but still has the same general effect as the original offering. The logical starting point is still with the real world. This step represents the former portion of the above definition of uncertainty: the real geographic phenomenon. The next step is to identify and define the geographic phenomenon of interest within this reality. Unfortunately, many phenomena of interest are inherently ambiguous or vague, making them difficult to define. This is particularly true with human/social phenomena (e.g., how do you define poverty?), but is the case for physical/environmental phenomena as well. Once a workable variable definition is in place, data is then collected. There are a variety of concerns when actually collecting data, including collection errors, the inability to collect data at every point in space and time, and sensor reliability. Following collection, data is converted to information through some sort of systematic interpretation. In the context of Cartography, the raw data is converted into a state that is mappable so that it can be consumed by the human-sensory system. Finally, the information is explored and analyzed to arrive at knowledge. Knowledge is the meaning generated about the geographic phenomenon in question. This step represents the latter portion of the above definition of uncertainty: the user's understanding of the geographic phenomenon.



**Figure 2: Uncertainty Filters.** Between each step in the analytical process, there is an uncertainty filter that acts to remove direct correspondence between phenomenon and the user's understanding of the phenomenon.

This process is one of abstraction, moving from the geographic phenomenon itself to an understanding of the phenomenon that is simple enough to be held in the mind of the analyst. Following the knowledge stage of analysis, this new understanding is often used to make a decision or policy. If the same person completes all steps in this analytical process, it can then be assumed that an informed decision was made that takes into account the uncertainties introduced at each step. This scenario is rarely the case, however, as a different person or group of people often completes each stage with poor communication across groups. Therefore, it is essential that cartographers – focusing primarily upon the information stage – both obtain the uncertainty recordings that entered in previous steps and communicate these recordings to those constructing knowledge from the information by directly symbolizing this uncertainty on the map.

## Uncertainty Categories: The forms that uncertainty takes in geospatial information

The second important subdivision is the set of forms that uncertainty takes, or *uncertainty categories*. In the following, I focus on uncertainties present in geospatial information. This parallels the broader focus in Cartography on the representation of information. Because of this, only descriptions of each uncertainty category are discussed rather than details of methods for collecting these metrics.

Numerous taxonomies of geospatial information uncertainty have been offered in the literature after Sinton's (1978) recognition that uncertainty has multiple categories that need to be treated differently during measurement and representation. Scholars approaching uncertainty from a statistical perspective typically focus on *accuracy*, defined as the difference between observation and reality. Closely related to the concept of accuracy is *error*, defined as the difference among observers or among measuring instruments. In this regard, error is the degree of repeatability in the measurement, either among observers or among measuring devices, and not related to the ability to correctly capture the underlying reality. Because the accuracy of a data value cannot be measured (if the actual reality could be captured independently of the situatedness of an observer, there would be no need to find a measure of the difference between observed and reality), an estimation of error, such as a confidence interval, commonly is used as a proxy.

Most scholars also generally agree that the precision of the information is an important category of uncertainty. *Precision* is the exactness of the measurement during data collection. In this regard, precision is equated to the number of digits used to report a measurement. However, broadening the definition of precision to the degree of detail that can be recorded acknowledges the connection of precision to *resolution*, defined as the level of detail. Because the degree of detail that can be recorded is reliant upon the precision of the recording device, resolution is often considered as a subcomponent of precision. Precision and resolution both deal with a central concern of geospatial data collection: it is impossible to collect data at an infinitely minute granularity of space and time. This problem forces the collection of data over subdivisions of space, such as the pixels in a raster data model and discrete point, line, or polygon entities in a vector data model. The measurement of data at aggregate spatial units removes detail from the representation due to increased abstraction, introducing uncertainty attributed to both precision and resolution.

Accuracy and precision are the two most commonly reported measures of uncertainty. This is perhaps due to the relative ease of generating quantitative metrics for these categories. Figure 3 illustrates the difference between the two categories of uncertainty using the popular dart board example.

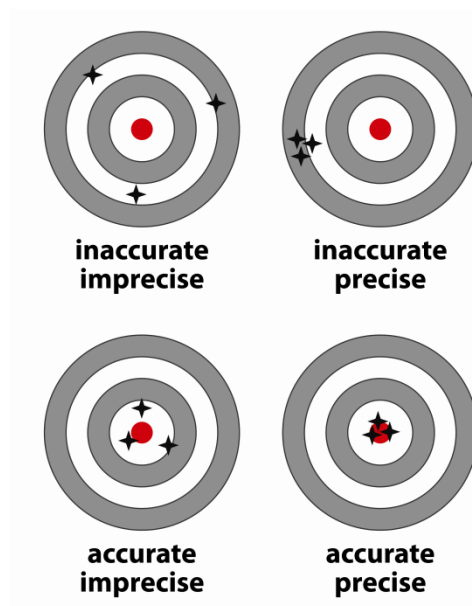


Figure 3: Accurate versus Precision.

Finally, many uncertainties are derived from the general *trustworthiness* of the information. Issues of information trustworthiness are particularly a problem in such application domains as intelligence analysis and crisis management, where misinformation may be purposefully circulated to mislead analysts and first responders. Categories related to information trustworthiness include the completeness, consistency, currency, lineage, and subjectivity of the information itself, the credibility of the information source, and the interrelatedness of the information source with other sources. Because of their inherent subjectivity, these categories of uncertainty are often more difficult to quantify.

An excellent review of each category is provided by [Alan MacEachren, Anthony Robinson, and colleagues \(2005\)](#). Table 1 lists the definitions provided by MacEachren et al. (2005) and provides an example of each category in the context of floodplain map as offered in [Roth \(2009\)](#).

<i>Uncertainty Category</i>	<i>Definitions</i>	<i>Example of Uncertainty in Floodplain Mapping</i>
<b>Accuracy/Error</b>	The difference between observation and reality.	The information has an error margin of +/- 100m for the floodplain delineation.
<b>Completeness</b>	The extent to which information is comprehensive.	The information only covers 70% of the study region.
<b>Consistency</b>	The extent to which information components agree.	The information is a composite of three different commissions with little overlap, completed by different firms.
<b>Credibility</b>	The reliability of information source.	The information was commissioned by the realty firm trying to sell you property.
<b>Currency</b>	The time span from occurrence through information collection/processing to use.	The information was compiled twenty years ago.
<b>Interrelatedness</b>	The source independence from other information.	Two different information sources show major disagreement in the position of the floodplain.
<b>Lineage</b>	The conduit through which information has passed.	It is information how the dataset was created.
<b>Precision/Resolution</b>	The exactness of measurement/estimate.	The information was generated using a DEM with a spatial resolution of 1km.
<b>Subjectivity</b>	The extent to which human interpretation or judgment is involved in information construction.	The information was created by a person tracing a scanned contour map by hand.

**Table 1: Uncertainty Categories.** The [MacEachren et al. \(2005\)](#) typology includes nine categories: (1) accuracy/error, (2) completeness, (3) consistency, (4) credibility, (5) currency, (6) interrelatedness, (7) lineage, (8) precision/resolution, and (9) subjectivity. Examples in the domain of floodplain mapping are given, as offered in [Roth \(2009\)](#).

## **Glossary of Terms**

*accuracy* - the difference between observation and reality

*ambiguity* - differing perceptions about or imperfect indicators of a geographic phenomenon

*cartographic problematic* - in order to create an abstraction of reality that makes complex geospatial information understandable and usable, uncertainty is introduced into the representation (and into the knowledge constructed from the representation) as a necessary compromise

*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

*error* - the difference among observers or among measuring instruments

*fuzziness* - the degree to which a geographic phenomenon belongs to a particular definition, class, or other attribute

*interrelatedness* - the source independence from other information

*lineage* - the conduit through which information has passed.

*quality* - the fitness-of-use of data or information for a given purpose

*precision* - the exactness of the measurement during data collection

*resolution* - the level of detail in the information

*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

*uncertainty categories* - the forms that uncertainty takes in geospatial information

*uncertainty filters* - how uncertainty enters into the user's understanding of a geographic phenomenon

*vagueness* - inherently inexact definition of a geographic phenomenon or its attributes