A Typology for Visualizing Uncertainty

Conference Paper *in* Proceedings of SPIE - The International Society for Optical Engineering · March 2005

DOI: 10.1117/12.587254

in press (following revisions) for Conference on Visualization and Data Analysis 2005 (part of the IS&T/SPIE Symposium on Electronic Imaging 2005), 16-20 January 2005, San Jose, CA USA

A Typology for Visualizing Uncertainty

Judi Thomson[a] and Beth Hetzler[a] and Alan MacEachren[b] and Mark Gahegan[b] and Misha Pavel[c*]

[a] Battelle Memorial Institute, PNWD, Richland, WA, USA;
[b] Penn State University, State College, PA, USA;
[c] Oregon Health and Sciences University, Portland, OR, USA;

ABSTRACT

Information analysts, especially those working in the field of intelligence analysis, must assess not only the information presented to them, but the confidence they have in that information. Visual representations of information are challenged to incorporate a notion of confidence or certainty because the factors that influence the certainty or uncertainty of information vary with the type of information and the type of decisions being made. Visualization researchers have no model or framework for describing uncertainty as it relates to intelligence analysis, thus no consistent basis for constructing visualizations of uncertainty. This paper presents a typology describing the aspects of uncertainty related to intelligence analysis, drawing on frameworks for uncertainty representation in scientific computing.

Keywords: uncertainty, framework, geospatial information

1. INTRODUCTION

Analysts make allowances for uncertainty in data on a daily basis. They understand that the information upon which decisions are based is rarely absolutely true, and that the resulting decisions are based on best guesses and assumptions. In most situations, uncertainty is not a life threatening issue- people learn to cope with potential variants in the information they are given and make their decisions accordingly. However, where the information is used in high risk decisions affecting large numbers of people, a comprehensive understanding of the uncertainty in the data becomes far more important.

The thesis of this paper is that those analysts who have consistent, comprehensive representations for the multiple uncertainties associated with data, and understanding of the impact of those uncertainties on decisions, make better decisions. Implied in this thesis is the notion that the relevance of particular types of uncertainty changes depending on the task or decision facing the analyst. For instance, suppose an analyst is trying to determine the immediate impacts of a chemical explosion in a building. The analyst would be quite concerned about accurately knowing the number of people in the building, but possibly less concerned with accurate information about the wind speed and weather patterns for the area around the building. On the other hand, an analyst tasked with determining longer term impacts of the same explosion might be more concerned with accurate information about wind and weather patterns. Uncertainty in the same data affects the different analysts in different ways. Mechanisms for flexible representations of the uncertainties important to the analytical process are needed and to provide those representations, an understanding of the different facets of uncertainty and how they affect analysis is required.

Significant work has been done to understand and approach inaccuracies in measures such as location, time of day, etc., which allows rudimentary understandings of uncertainties, but are insufficient for representing the myriad uncertainties encountered by analysts. In addition to uncertain measures, analysts are concerned with abstract

[*] Further author information: (Send correspondence to J.T.)
J.T.: E-mail: judi@jthomson.net, Telephone: 1 509 375 4438
B.H.: E-mail:beth.hetzler@pnl.gov, Telephone: 1 509 375 6690

uncertainties such as the credibility of a particular source, or the completeness of a set of information. As the uncertainty becomes more abstract, it is more difficult to quantify, represent and understand. Without this understanding, a visual representation of the uncertainty will be ad hoc and ineffective.

The uncertainty typology presented here bridges the gap between uncertainty as it affects analysts and uncertainty as it is typically measured and understood in a scientific study field (geographic information systems). The typology describes generic types of uncertainty and then applies those types to the field of intelligence analysis with specific examples. We begin with a discussion of the analytical understanding of uncertainty followed by a survey of the common approaches for representing geospatial uncertainty. We then show how these two different approaches to uncertainty representation can contribute to an effective framework for representing uncertainty to analysts who work with geospatially-referenced information. We conclude with a discussion of how the typology could be used to create computational representations of uncertainty for use in visualization.

2. ANALYTIC UNCERTAINTY

The goals, process, and even outcomes of intelligence analysis are intrinsically tied to an understanding of uncertainty. Jack Davis writes that "the central task of intelligence analysis is to help US officials…deal more effectively with substantive uncertainty" [1]. While uncertainty plays a role in all stages of intelligence production and use, the uncertainties associated with the information and data greatly affect the production of analytic products which in turn affect the decisions made by officials.

Information collected to support creation of analytic products has several aspects that affect the uncertainty associated with the information. For instance, even the granularity of the information can affect the uncertainty surrounding it. A report identifying the location of an individual only by city has much less certainty, with respect to the location than one identifying the location by street address. Uncertainties can also be introduced during the analytic process. Collection, processing, translation and dissemination of information can all add uncertainty to the quality/accuracy of the information. An understanding of how each of these different aspects of uncertainty affects the particular analytic task is important to researchers trying to depict the uncertainty in a meaningful way for the analyst.

For example, the particular collection capability used to gather the information (e.g., satellite, sensor) may have limited accuracy or precision which would affect the reliability of the gathered information. Or, a collection method might be vulnerable to deception or disinformation, meaning that the information collected is reliable most of the time, but quite unreliable if it has been compromised. Perhaps a collection method is only reliable for a certain period of time (such as a temperature sensor that is unreliable in full sun). Analysts using collected data must be aware of the uncertainties associated with the collection mechanism if they are to make reasonable conclusions from the collected information.

The lineage or provenance of the information can be traced by keeping careful records of each process or transformation the information undergoes. Each processing step can introduce additional elements of uncertainty. For example, translation, decryption or numerical transformations, could introduce errors or reductions in precision. The interpretation of images, usually done by human experts, can introduce uncertainties caused by subjectivity biases. Thus it is also important that analysts be aware of information provenance and its effects on the uncertainty associated with the data.

A key element of uncertainty is associated with the source of the information, especially when the source is human reporting or personal communication. The reliability of human sources is evaluated based on past performance but that judgment is difficult to quantify and represent. Individual pieces of information may contain indicative wording, such as "believe" or "guess" to show the source's own evaluation of information reliability, but these do nothing to help evaluate the reliability of the source itself. Proximity and appropriateness are other measures that factor into an understanding of source reliability. A direct observation of an event is more proximate, thus may have less uncertainty than a report related second or third hand. A source is appropriate if it is known to have knowledge about the subject matter in question; the source may be an expert on some matters and wholly inexperienced in others. [2]. An appropriate source has less uncertainty associated than an inappropriate source. Motivation on the part of the information supplier also affects the quantification of uncertainties associated with the data source. A source may embellish their information to make it more appealing or more important. Volunteered information may be a ploy by the opponent to pass on disinformation [3] and could then be considered to have a higher degree of uncertainty associated. The possibility of a source intentionally providing deceptive information raises an even wider variety of uncertainties including the likelihood of incomplete information, conflicting information, or even information with intentional errors.

Uncertainty can be exacerbated by temporal delays between occurrence and information acquisition. The actual occurrence of events, the observation of those events, the reporting of the events and the time at which the analyst becomes aware of the report is usually distinctly different. . When there is a gap between the learning of information and the reporting of it, uncertainty about the accuracy of the information due to faulty memory or a change in information increases [4]. Temporal differences can be interpreted in multiple ways. For example, a report about a new combat unit may indicate that the unit is actually new, or it might mean that the unit has been present for some time but only just detected [4]. Also, if multiple versions of information exist, the uncertainties multiply because analysts must also wonder if they have the most current version.

Information available to analysts is never complete. Even with excellent collection capabilities, observed evidence is likely to only be a small portion of the whole [5] and analysts must not close their minds entirely to alternate explanations, even those that seem only remotely possible. The Rumsfeld Report focuses strongly on opponent's denial and deception capabilities [6] and advocates that analysts explicitly identify gaps in their knowledge. Even individual pieces of information contain gaps, such as missing parts of intercepted conversations, partially occluded images, or sensor readings that are incomplete due to adverse conditions. An absence of information also presents uncertainties to the information analyst. If an analyst is unable to find evidence of a particular event, the event may not have happened, or there may just be no information available about it Perhaps the information was not collected; perhaps one of the collection vulnerabilities came into play and the opponents were able to hide the event; perhaps evidence does exist, but the analyst's search methods have just not found it. In many instances, a lack of information must be treated as incomplete information, where several types of uncertainty come into play.

The goal of the information analyst is to minimize the effects of the uncertainties on decisions and conclusions that arise from the available information. This can be done by identifying and accounting for the uncertainty and ensuring that the analyst understands the impacts of the uncertainty. One mechanism for reducing uncertainty is to find corroborating information from multiple sources with different types of uncertainty. Then, especially when the uncertainties are known, the different sources can be used to create a more rounded impression of the topic under consideration. This technique introduces its own uncertainties because there is rarely a reliable way to assess the independence of the sources. For example, even if two different news articles report the same information, one article may actually be based on the other, or both may derive from a common source. Analysts may only know that a given piece of information comes from "a source with access who has reported reliably in the past" [4] which provides no clues about source independence.

Another common technique is to make use of stated assumptions and models of the situation. While the models may help the analyst to identify holes and missing information, they introduce more kinds of uncertainty, such as uncertainty about the validity of the assumptions, uncertainty about the model used, and uncertainty due to mismatches between the data driving the model and reality. Nye relates an example from the 1980s, where analysts estimated Iraq's ability to build a nuclear weapon, assuming that they would use modern techniques. The assumption was inaccurate; part of Iraq's program used an outdated technique long abandoned by the U.S. [7]. No matter the task, the analyst must be cognizant of the uncertainties arising from the different sources, those that arise from interactions between the sources, and those that are present due to missing or incomplete information or models.

Given the complexity of both the underlying data and the associated uncertainties, intelligence analysts risk a number of cognitive biases, such as those described by [8]. We believe that a systematic method for visualizing uncertainties in the data will help counter such biases and may be a key addition to current analytic methods. However, due to the wide variety of these uncertainties, a first necessary step is to organize the uncertainties into a logical framework or typology that allows analysts and information scientists to describe and examine the characteristics of the different kinds of uncertainties. The next section of this paper explores frameworks for uncertainty that have been developed for representation within the geosciences and scientific visualization community. The subsequent section presents a typology for representing the uncertainty in geospatially referenced information, as well as a derivative framework specifically focused on intelligence analysis in a geospatial domain.

3. UNCERTAINTY TYPOLOGIES AND REPRESENTATIONS

Uncertainty has been defined as the degree to which the lack of knowledge about the amount of error is responsible for hesitancy in accepting results and observations without caution [9]. In other definitions, the term uncertainty represents a broader range of doubt or inconsistency than is implied by the term error and error is just one component of uncertainty [10]. Other definitions, including accuracy, reliability, ignorance, precision, clearness, distinctiveness, are also used in the literature [11, 12]. More generally, definitions of uncertainty imply that there is imperfection in the users' knowledge about a dataset, process or result.

Given the range of definitions for uncertainty, it is not surprising that significant effort has gone into developing classifications for different types of uncertainty. There have been several efforts to delineate the components of information uncertainty and the relation of those components to visualization of the information and its uncertainty. These efforts have proceeded somewhat in parallel (with limited sharing of ideas) in the Geographic Visualization/Geographic Information Science (geovisualization/GIScience) and the Scientific Visualization (SciVis)/Information Visualization (InfoVis) communities.

3.1. Uncertainty Typologies in Geovisualization/GIScience

Most of the efforts to formalize an approach to uncertainty visualization within geovisualization (and GIScience more generally) derive from long term efforts to develop spatial data transfer standards (SDTS) [13-15]. The focus of the initial efforts here was on specifying categories of "data quality" that should be encoded as part of the metadata for cartographic data sets. The categories defined as part of the SDTS are listed below (all quotes from [16]).

- **lineage**: source material and the methods of derivation, including all transformations involved
- **positional accuracy**: degree of compliance to the spatial registration standard
- **attribute accuracy**: both measurement accuracy (for features measured on a continuous scale) and class assignment accuracy (for categorical features) are included here.
- **logical consistency**: fidelity of relationships encoded in the data structure
- **completeness**: relationship between the objects represented and the abstract universe of all such objects. Issues such as selection criteria (e.g., size thresholds for spatial features, frequency counts for attributes), definitions used, and other mapping/abstraction rules. A particular focus is placed on "exhaustiveness" of a set of features.

These metadata about data quality can be used to represent some aspects of the uncertainty about data, but offer no suggestions about how the metadata can be visually represented. Buttenfield [17] was among the first to focus on a framework for categorizing components of “data quality” with a specific focus on their cartographic representation (see Figure 1). Her approach matched the 5 categories of data quality with three data types: discrete (point and line features), categorical (area features assigned to categories or attributes assigned to classes), and continuous (surfaces and volumes). For each cell in the resulting matrix, she focused on which “visual variables” [18] were most appropriate to depict the category. MacEachren [19, 20] added specific attention to information precision (distinguished from

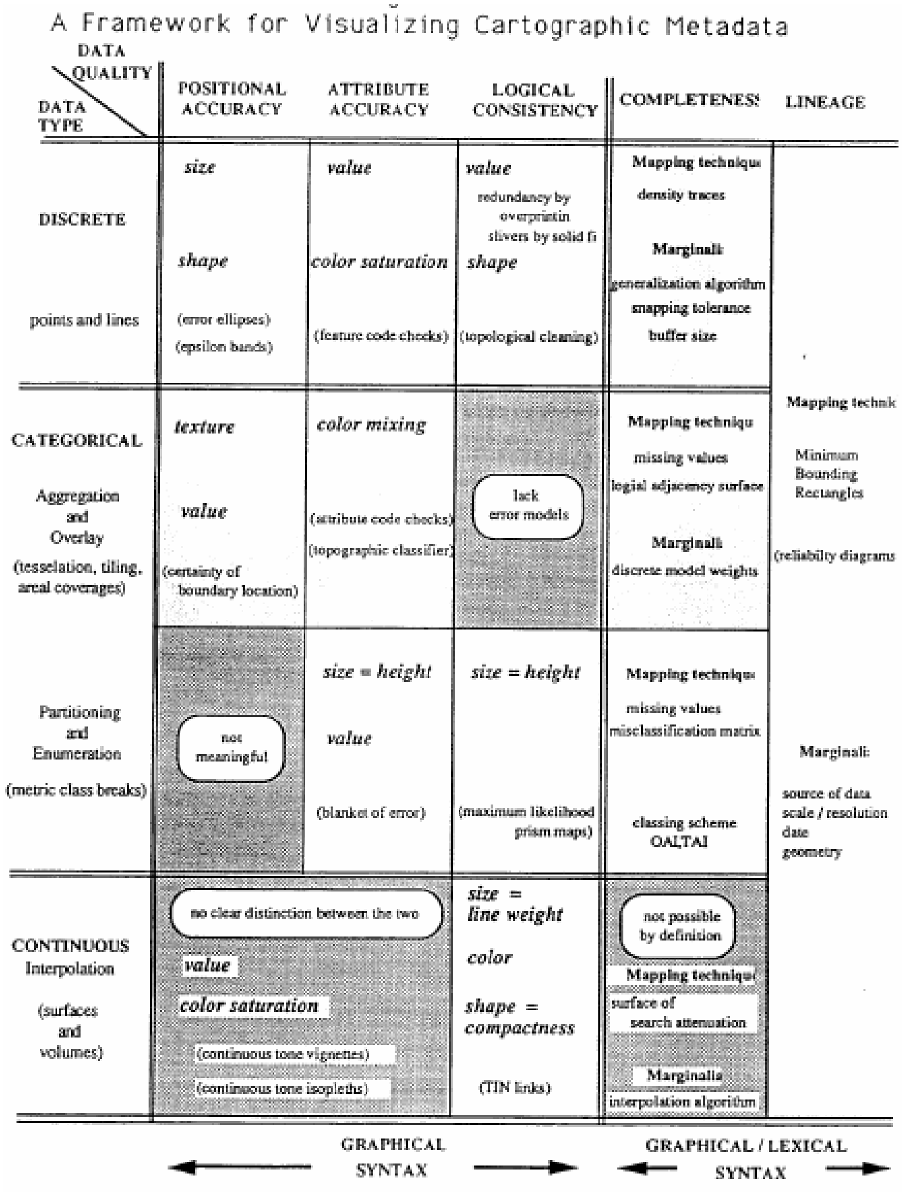

Figure 1. Buttenfield’s framework

accuracy) and focused on matching kinds of uncertainty to an early proposed [21] distinction among location, attribute, and time components of data. Buttenfield and Beard [22] also adapted Buttenfield’s initial framework to include location, attribute, and time components, and added resolution.

In related work, Gahegan and Ehlers [10] focus on modeling uncertainty within the context of fusing activities between GIS and remote sensing. Their approach matches 5 types of uncertainty (data/value error/precision, space error/precision, time error/precision, consistency, and completeness) against four models of geographic space (field, image, thematic, and object). Thus, they merge the location, attribute, time distinction discussed above with kinds of uncertainty, while focusing on the implications of the different approaches to modeling geographic space. A particular emphasis in their work is on error (and uncertainty) propagation.

3.2 Scientific Visualization/Information Visualization

Visualization of uncertainty (or reliability) has been a topic of attention for several researchers within the SciVis and InfoVis communities (e.g.,[23-28]). In spite of the wide array of visualization methods proposed, there seems to have been less attention to formalizing approaches to uncertainty visualization within the SciVis and InfoVis communities than in the geovisualization/GIScience communities. There are, however, two important contributions to formalizing an approach to uncertainty visualization, both discussed below.

Pang and colleagues [29] produce a classification of methods for uncertainty visualization that matches data type (scalar, multivariate, vector, and tensor) to visualization extent (discrete and continuous). For each of the 8 cells in the resulting matrix, they propose some logical representation methods (including both static and dynamic representation forms).

Gershon [30] takes a different approach, focusing on kinds of "imperfection" in information that might be provided to an analyst or decision maker. He argues that imperfect information is more complex than typically considered from the viewpoint of uncertainty. Figure 2 depicts Gershon's "high-level taxonomy of causes for imperfect knowledge of the information state." Two important points are that (1) uncertainty is considered to be just one of 6 inputs and that (2) the quality of the presentation is a critical factor. It is also interesting that the elements of this taxonomy align well with the

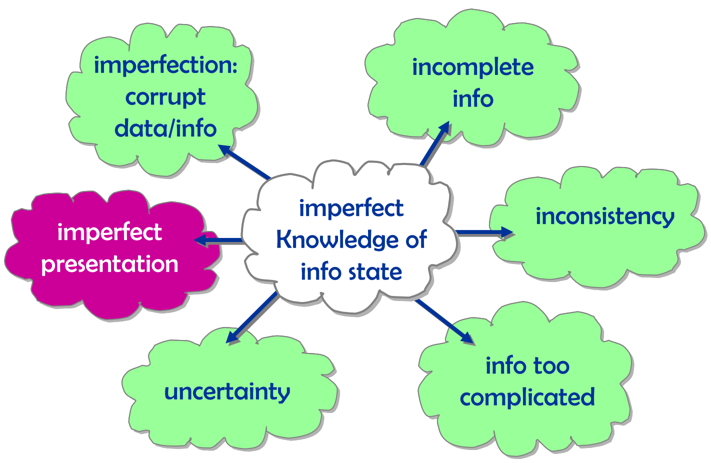

Figure 2. Causes for Imperfect Knowledge (Gershon, 1998)

somewhat abstract causes for uncertainty faced by information analysts.

Most typologies for uncertainty and its representation include one or more levels of abstraction for uncertainty representation. For example, positional and temporal error describe uncertainty in a metric sense within a space-time framework, whereas completeness and consistency represent more abstract concepts describing coverage and reliability. The abstract notions of uncertainty are more problematic to describe and quantify, consequently, most example typologies tend to focus heavily on the types of uncertainty that can be represented using measurements of error and precision. To make a typology useful for the abstract kinds of uncertainty faced by intelligence analysts, the descriptions and recommendations must include types of uncertainty that are not easily measured., Fortunately, we believe that these types of uncertainty can be relatively easily identified within the context of specific tasks or domains, and that their effects on these tasks can be modeled. The remainder of this paper presents a general typology framework for uncertainty, along with a specific instance of the typology applied to the task of producing an analytic product.

4. A TYPOLOGY FOR GEOSPATIALLY REFERENCED INFORMATION

Ideally, a typology for uncertainty visualization would give the user some guidance about both visual and computational representations for the different types of uncertainty, including those abstract types encountered by information analysts. It seems reasonable that some categories of representations techniques, both computational and visual, should work particularly effectively to represent specific types of uncertainty and less well for others. The finished uncertainty typology can be used to guide the visualization designer to select appropriate visual strategies for both the information and the associated uncertainty.

The first step to creating such a typology is to identify the broad categories for uncertainty and determine how to identify and model the specific instances of those types of uncertainty in a particular domain. The tables below illustrate the broad categories in a typology that draws from those discussed earlier but expands the categories to include abstract types of uncertainty that are often excluded. Table 1 gives a term and general definitions for each uncertainty category. Table 2 gives examples of metadata that could be used to represent each abstract category of uncertainty within a specific set of tasks or domain. It is easy to see that some of the metadata would be easy to model or collect in real time (such as coordinates and counts) while other metadata would be far more difficult to obtain or accurately model (such as the expertise level of the informant or the accuracy of the simulated model). This typology also contains suggestions for the quantitative representation for uncertainty in each of the general categories. We hope that a quantitative representation will ease the task of combining and propagating the various uncertainties for visual presentation to the information analyst. These representation suggestions are discussed further in section 5.

Table 1. Typology Categories

Category	Definition
Accuracy/error	difference between observation and reality
Precision	exactness of measurement
Completeness	extent to which info is comprehensive
Consistency	extent to which info components agree
Lineage	conduit through which info passed
Currency/timing	temporal gaps between occurrence, info collection & use
Credibility	reliability of info source
Subjectivity	amount of interpretation or judgment included
Interrelatedness	source independence from other information

Table 2. Domain Examples for Typology

Category	Attribute Examples	Location Examples	Time Examples
Accuracy/error	counts, magnitudes	coords., buildings	+/- 1 day
Precision	nearest 1000	1 degree	once per day
Completeness	75% reporting	20% cloud cover	5 samples for 100
Consistency	multiple classifiers	from / for a place	5 say M; 2 say T
Lineage	transformations	# of input sources	# of steps
Currency/timing	census data	age of maps	C = Tp - Ti
Credibility	U.S. analyst; informan	knowledge of place	reliability of model
Subjectivity	fact <…> guess	local <…> outsider	expert <…> trainee
Interrelatedness	same author	source proximity	time proximity

A key difference between the typology for geospatially referenced information and past efforts is the focus on the task of the information analyst. This typology has been developed to specifically address the types of uncertainty intelligence analysts face. The categories in this typology represent the types of uncertainties previously identified as having significant influence on analytical decisions. Indeed, analysts routinely keep track of these types of uncertainties manually, in an attempt to incorporate that knowledge into any product they make. We believe that the finished typology will provide a guide for the methodical development of visual representations that incorporate and identify the uncertainties most important to the analyst for the task at hand.

This is a second key difference between the typology presented here and other previous efforts. The typology for geospatially referenced information is presented as a general typology that must be instantiated to define the specific uncertainties relevant to a particular type of task. As illustration, Table 3 gives an instantiation of the uncertainty typology for the task of intelligence analysis with the goal of creating analytic product. The table shows the subcategories of each of the main categories that are particularly relevant to the selected task along with examples of when that type of uncertainty might occur.

Table 3. Analytic Uncertainty Typology

Category	Subcategories	Examples
Accuracy/error	• Collection Accuracy • Processing errors • Deception	• Documents that are translated into English may contain translation errors. • A report may note that 50 tanks were observed although the tanks may in fact be dummy placements.
Precision	• Precision of collection capability	• resolution of satellite imagery
Completeness	• Composite completeness • Information completeness • Incomplete sequence	• images of a site may not be available on a particular day, due to adverse weather conditions. • an intercepted conversation may have words that were not clear • the lack of confirming information might signal incompleteness
Consistency	• Multi-INT Conflict • Model/observation Consistency	• multiple sources may actually conflict • models of events may differ from observations
Lineage	• Translation • Transformation • Interpretation	• Machine translation is more uncertain than human linguist translation • Measurements or signals may have been transformed • Information that comes directly from the collection capabilities has a different lineage than an interpretive report produced by an analyst
Currency/timing	• Temporal gaps • Versioning	• Images that show new objects do not show when the object first appeared • Time between when events occurred, when they were reported, and when the information is available to analysts • Reports may have multiple versions, sometimes with major changes.
Credibility	• Reliability • Proximity • Appropriateness • Motivation (of the source)	• Possibility of deliberate disinformation • Source may not have expertise on this subject • Information may be second hand
Subjectivity	• Analytic judgment	• Amount of interpretation added rather than pure facts
Interrelatedness	• Source independence	• Likelihood that the source derives from other reported information (such as repeated news stories)

The advantage of the tiered approach to creating typologies and in working from a generalized description of categories to create a typology specific to a task is that the specific typologies will be interoperable. A typology created for analytic uncertainty will have the same high level categories as a typology created for uncertainty in scientific data collection. This means that information workers, and automated tools, have a common frame of reference when describing the uncertainties. Tools designed for one task can more easily be extended to represent uncertainties for a second task when the underlying model of uncertainty has a common basis. One feature of this typology that promotes interoperability between specific instantiations is the definition of quantitative representation mechanisms for each of the general categories of uncertainty. The key assumption in this approach is that a quantitative representation of uncertainty will be more useful and effective than a qualitative approach alone because it provides principled ways of combining different types of uncertainties.

5. ENABLING UNCERTAINTY COMBINATION AND PROPOGATION

Many different representations are both possible and appropriate for the different types of uncertainty. As illustration of how representation definition can enhance the use of the typology, we present an example based on probabilistic representation. In fact, the example is not too restrictive because many of the typology-related concepts can be recast in different frameworks for uncertainty. A thorough attempt to integrate different frameworks for reasoning under uncertainty and to elucidate various analogies and connections among the different types of representations is found in [31]. A principled way of combining different types of uncertainty will permit the analysts to compare his or her conclusion with that of the system and this challenge has been found useful in much realistic decision analysis.

Even with this constraint of a probabilistic representation, there are a large number of ways to map the various types of uncertainties and their interactions into a quantitative representation. We discuss the following example to demonstrate one simple way of representing uncertainty. We assume that the actual state of the world e.g., location of a stockpile, can only be observed through obtaining information from a variety of sources. Each source yields an output reading, which can comprise measurements, locations, categorical statements, set of propositions, etc. The uncertainty associated with the output from each source is defined as the distribution of the output given the actual state of the world. This definition, interpreted in the most general terms, would then be used to construct Bayesian networks with specific mappings between the uncertainty types and probabilistic concepts as shown in Table 4.

The first column of Table 4 lists the basic types of uncertainty from the typology presented earlier. The second column of this table shows one mechanism for representing each type of uncertainty in probabilistic terms. The third column provides examples for simple, parametric interpretation of uncertainty associated with each of the concepts. For example, the variance is one representation of the width of the distribution of output observations, which gives some quantifiable measure of the uncertainty of the observations. The final column gives specific parameters that could be used with a linear model for combining and propagating the uncertainties. A creator of visual representations can use the probabilistic networks to determine quantitative values for the composite uncertainties associated with specific information. The resulting visualization could display a single uncertainty value for the analyst, while retaining the underlying model and metadata that allows the analyst to decompose the uncertainty visualization to inspect the underlying components.

While recognizing that linear models may not be adequate for representing all aspects of uncertainty, we believe that a linear model will provide a convenient starting point for testing ways of eliciting and communicating uncertainties. We note, that despite the fact that most phenomena are not linear, statistical tests, such as analysis of variance, that are based on linearity assumptions have been shown to be robust with respect to the violation of these assumptions.

Table 4. Quantitative Representation of Uncertainty

Category	Probabilistic Representation	Parameter Examples	Model Examples
Accuracy/error	Distribution of measurement error	Source variance	$\sigma_x^2 = \sigma_s^2$
Precision	Measurement device limitations	Root mean square deviation	$\sigma_x^2 = E\left\{\left(x_s - \overline{x}_s\right)^2\right\}$

Completeness	Sampling error	Variance and bias due to sampling error	$\sigma_x^2 = \frac{\sigma_s^2}{\gamma}$
Consistency	Repeatability across multiple sources	Inter-source variance	$\sigma_x^2 = \sigma_k^2 + \sigma_s^2$
Lineage	Change due to transformation stages	Additional variance and/or bias due to channel noise	$\sigma_x^2 = \sigma_s^2 + \sigma_c^2$
Currency/timing	Dependence of accuracy or time of collection	Dependence on collection time $\Pr\{t_a, t_c\} \neq \Pr\{t_a\}$	$\sigma_x^2(t, t_s) = \sigma_s^2 e^{\frac{(t-t_s)}{\tau}}$
Credibility	Dependence of accuracy or source	Additional variance due to source lack of knowledge	$\sigma_x^2 = \frac{\sigma_s^2}{\alpha}$
Subjectivity	Distribution of data depends on the analyst	Variance of the analys	$\sigma_x^2 = \sigma_a^2 + \sigma_s^2$
Interrelatedness	Lack of independence	Source Correlation	$r_{1,2}^2 = E\{\tilde{x}_1 \tilde{x}_2\} \geq 0$

In the linear model interpretation, σ_s is the source variance, σ_k is the variance due to different sources of information, σ_c is the variance added due to error introduced in communication. The parameters $\gamma \in [0,1]$ and $\alpha \in [0,1]$ represent completeness and credibility. Finally the time constant τ represents the rate at which the information value decays with time since the data collection. We also note that the parameterization in terms of variances is consistent with the notion of Fisher information [32]. Many different models are possible, in fact we believe that specific models will prove to be effective for specific types of tasks and goals. Our current efforts are focused on defining a single model and evaluating its utility.

6. SUMMARY: USING THE TYPOLOGY

We have presented a typology that provides a general categorization for the types of uncertainty present in geospatially referenced information along with a specification of the typology describing the specific uncertainties faced by intelligence analysts when tasked with creating an analytic product. We are currently evaluating the applicability of the typology to practical visualizations aimed at increasing the accuracy of decisions that are based on visual representations of geospatially referenced information. We believe that the typology will be a valuable framework for use by persons who construct visualizations for analysts. The graphical components of the visualization that are related to information uncertainty can be linked to specific aspects of the typology. This allows the visualization to use the model underlying the typology to combine and propagate uncertainties as well as information about the types of visual components that are most suited for representing specific types of uncertainty. The next step in this work is to identify visual metaphors that are particularly effective for each of the different types of uncertainty, as well as for uncertainty that is a composite of two or more types. We have begun evaluations to identify suitable glyphs for annotating geospatial visualizations, and are exploring the use of manipulative icons and glyphs for representing composites

Our belief is that a visualization that maps each type of uncertainty into a different visual dimension will permit the analyst to express the implications of each dimension separately and to perform cognitively the fusion of the different type of uncertainties. At the same time the system will estimate the parameters of the underlying model and be ready to present the fusion that results from applying the model to the known uncertainties. In that way the analyst can make comparisons between their own conclusions and the systems, which provides an appropriate check on the consistency of the analysis.

.

ACKNOWLEDGMENTS

This study was supported and monitored by the Advanced Research and Development Activity (ARDA) and the National Geospatial-Intelligence Agency (NGA) under Contract Number HM158204-C-0021."

The views, opinions, and findings contained in this report are those of the author(s) and should not be construed as an official Department of Defense position, policy, or decision, unless so designated by other official documentation.

REFERENCES

1. Davis, J., *Improving CIA Analytic Performance: Strategic Warning*. Sherman Kent Center for Intelligence Analysis ed. 2002.
2. Krizan, L., *Intelligence Essentials for Everyone*. 1999: Joint Military Intelligence College.
3. Shulsky, A.N. and G. Schmitt, *Silent Warfare*. 3rd (May 15, 2002) ed. 2002: Brasseys, Inc. 320.
4. Berkowitz, B. and A.E. Goodman, *Strategic Intelligence for American National Security*. 1989: Princeton University Press. 264.
5. Grabo, C., *Anticipating Surprise: Analysis for Strategic Warning*. 2003: Joint Military Intelligence College.
6. Rumsfeld, D.H., *Executive Summary of the Report of the Commission to assess the Ballistic Missile Threat to the United States*. 1998, U. S. Congress: Washington, D.C.
7. Nye, J.S.J., *Peering into the Future*. 1994.
8. Heuer, R., *Psychology of Intelligence Analysis*. 1999: United States Government Printing Office. 210.
9. Hunter, G.J. and M.F. Goodchild, *Managing uncertainty in spatial databases: Putting theory into practice.* Journal of Urban and Regional Information Systems Association, 1993. 5(2): p. 55-62.
10. Gahegan, M. and M. Ehlers, *A framework for the modelling of uncertainty between remote sensing and geographic information systems.* ISPRS Journal of Photogrammetry and Remote Sensing, 2000. 55: p. 176-188.
11. Crosetto, M., J.A.M. Ruiz, and B. Crippa, *Uncertainty propagation in models driven by remotely sensed data.* Remote Sensing of Environment, 2001. 76: p. 373-385.
12. Duckham, M., et al., *A formal approach to imperfection in geographic information.* Computers, Environment and Urban Systems, 2001. 25: p. 89-103.
13. Morrison, J., ed. *NCDCDS (National Committee for Digital Cartographic Data Standards), The proposed Standard for Digital Cartographic Data*. 1988, ACSM: Washington, D.C.
14. Fegeas, R.G., J.L. Cascio, and R.A. Lazar, *An overview of FIPS 173, The spatial data transfer standard.* Cartography and Geographic Information Systems, 1992. 19(5): p. 278-293.
15. Moellering, H., *Continuing research needs resulting from the SDTS.* Cartograph and Geographic Information Systems, 1994. 21(3): p. 180-189.
16. USGS, *Spatial Data Transfer Standard (SDTS): Logical Specifications*. 1977.
17. Buttenfield, B.P. and J.H. Ganter. *Visualization and GIS: What should we see? What might we miss?* in *Proceedings of the 4th International Symposium on Spatial Data Handling (pre-publication copy)*. 1990.
18. Bertin, J., *Semiology of Graphics: Diagrams, Networks, Maps*. 1983.
19. MacEachren, A.M., *Visualizing uncertain information.* Cartographic Perspectives, 1992(13): p. 10-19.
20. MacEachren, A.M., *SOME Truth with Maps: A Primer on Design and Symbolization*. AAG Resource Publications in Geography, ed. E.C.a.R. Cromely. 1994, Washington, D. C.: Assocation of American Geographers.
21. Sinton, D., *The inherent structure of information as a constraint to analysis: mapped thematic data as a case study*, in *Harvard Papers on Geographic Information Systems*, G. Dutton, Editor. 1978, Addison-Wesley: Reading MA, USA.
22. Buttenfield, B.P. and M.K. Beard, *Graphical and geographical components of data quality*, in *Visualization in Geographic Information Systems*, D. Unwin and H. Hearnshaw, Editors. 1994, John Wiley & Sons: London. p. 150-157.
23. Gershon, N.D. *Visualization of fuzzy data using generalized animation*. in *Proceedings, Visualization '92*. 1992. October 19-23, Boston, Mass.: IEEE Computer Society Technical Committee on Computer Graphics.
24. Wittenbrink, C.M., et al., *Glyphs for visualizing uncertainty in environmental vector fields.* IEEE Transactions of Visualization and Computer Graphics, 1995. 2(3): p. 266-279.
25. Lodha, S.K., et al. *UFLOW: Visualizing uncertainty in fluid flow*. in *Visualization '96*. 1996. San Francisco, CA: IEEE.

26. Cedilnik, A. and P. Rheingans. *Procedural annotation of uncertain information*. in *Visualization 2000*. 2000.
27. Dungan, J.L., D.L. Kao, and A. Pang. *Modeling and Visualizing Uncertainty in Continuous Variables Predicted Using Remotely Sensed Data*. in *Internation Geoscience and Remote Sensing Symposium*. 2003.
28. Johnson, C. and A. Sanderson, *A next step: Visualizing errors and uncertainty.* IEEE Computer Graphics and Applications, 2003. 23(5): p. 6-10.
29. Pang, A.T., C.M. Wittenbrink, and S.K. Lodha, *Approaches to uncertainty visualization.* Visual Computer, 1997. 13: p. 370-390.
30. Gershon, N.D., *Visualization of an imperfect world.* Computer Graphics and Applications, 1998. 18(4): p. 43-45.
31. Halpern, Y., *Reasoning about Uncertainty*. 2003: MIT Press.
32. Cover, T.M. and J.A. Thomas, *Elements of Information Theory*. 1991: Joh Wiley and Sons.