Envisioning Uncertainty in Geospatial Information

Kathryn B. Laskey
George Mason University
MS 4A6, 4400 University Drive
Fairfax, VA 22030-4400
klaskey@gmu.edu

Edward J. Wright
Information Extraction and Transport, Inc.
1911 N. Ft Myer Dr., Suite 600
Arlington, VA 22209
ewright@iet.com

Paulo C. G. da Costa
George Mason University
MS 4B5, 4400 University Drive
Fairfax, VA 22030-4400
pcosta@gmu.edu

Abstract

Geospatial Reasoning has been an essential aspect of military planning since the invention of cartography. Although maps have always been a focal point for developing situational awareness, the dawning era of Network Centric Operations brings the promise of unprecedented battlefield advantage due to improved geospatial situational awareness. Geographic information systems (GIS) and GIS-based decision support systems are ubiquitous within current military forces, as well as civil and humanitarian organizations. Understanding the quality of geospatial data is fundamental to using it intelligently. A systematic approach to data quality requires: estimating and describing the quality of data as it is collected; recording the data quality as meta data; propagating uncertainty through models for data processing; exploiting uncertainty appropriately in decision support tools; and communicating to the user the uncertainty in the final product. Bayesian reasoning provides a principled and coherent approach to representing and drawing inferences about data quality. This paper describes our research on data quality for military applications of geospatial reasoning, and describes model views appropriate for model builders, analysts, and end users.

1 INTRODUCTION

The focal point of the battlefield command post is the map. Through interactions with the map, the commander and staff collaborate to build a common operating picture. This common operating picture displays the area of operations, the militarily significant features of the terrain, the locations of adversary and friendly forces, and the evolving plan. A generation ago, planning centered on a paper map, its overlays of acetate covered with marks of grease pencils wielded by the staff members congregated around it. Today the paper map has been replaced in brigade and larger headquarters with a digitized map projected onto a large-screen display. The grease pencil has become an input device for drawing objects or selecting pre-computed overlays from a menu of options. The map and overlays are stored in the computer as data structures, are processed by algorithms that can generate in seconds products it would take soldiers many hours of tedious effort to duplicate, and can be sent instantly to relevant consumers anywhere on the Global Information Grid (GIG), the information processing infrastructure of the United States Department of Defense (DoD). The GIG is the physical infrastructure to enable Network-Centric Operations, the DoD's new doctrine for warfare in the 21st Century.

Advanced automated geospatial tools (AAGTs) transform commercial geographic information systems (GIS) into useful military services for Network Centric Operations. Because of their basis in commercial GIS, they also have widespread applicability to fire, police, disaster relief, and other problems characterized by a command hierarchy. The advanced situation awareness provided by AAGTs can do much more than simply speed up calculations. They are changing the way military operations are conducted. The development of tools is shaped by military necessity, but as the new century dawns, the decision making process itself is being shaped by the automated tools that provide warfighters with more robust situational awareness.

Widespread enthusiasm for AAGTs has created a demand for geospatial data that exceeds the capacity of agencies that produce data. As a result, geospatial data from a wide variety of sources is being used, often with little regard for quality. A concern is the influence of errors or uncertainty in geospatial data on the quality of military decisions made based on displays of geospatial data.

Quality of geospatial data is an issue that has received considerable interest in the academic GIS community (Goodchild, 1992). Studies have shown that, while all geospatial data contain errors, errors in geospatial data are not well documented, not well understood, and are commonly underestimated by users. A particular problem is the tendency of users to implicitly trust high resolution graphic computer displays of geographic data. The quality of the display masks the underlying uncertainty in the data (Lunetta & Congalton, 1991).

Scientifically based methodologies are required to assess data quality, to represent quality as metadata associated with GIS systems, to propagate it correctly through models for data fusion, data processing and decision support, and to provide end users with an assessment of the implications of uncertainty in the data on decision making. Statisticians have developed a wide variety of methods for analyzing and reasoning with spatial data (e.g., Cressie, 1993), and these methods are widely used in generating and analyzing geospatial data. A number of authors have applied Bayesian networks to reason about uncertainty in geographic information systems (e.g., Walker, et al., 2005). A Bayesian analysis plugin, based on the open source GeNIe/SMILE[1] open source Bayesian network system, has recently been released for the open source MapWindow[2] GIS system. Applications of Bayesian networks to geospatial reasoning include avalanche risk assessment (Grêt-Regamey and Straub, 2006), locust hazard modeling (Jianwen and Qin, 2005), and watershed management (Ames, 2002), and military decision support (Wright, 1998; 2002).

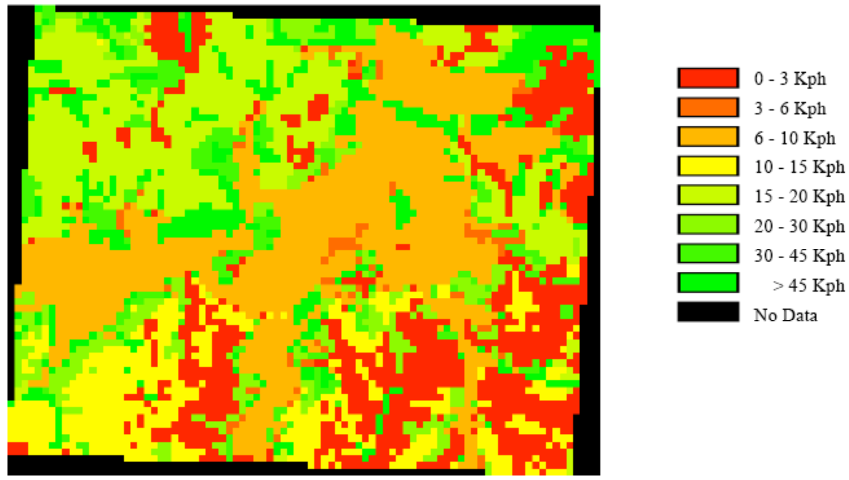

Figure 1: Traditional CCM Product
(M1 Tank, DMA Mobility Model, ITD Data, Korea)

In his dissertation on the application of Bayesian networks to tactical military decision aids, Wright (2002) considered all phases of the life cycle of geospatial data, including data generation, data management, analysis, display, and decision support. In this paper, we focus on improving decisions by representing, propagating through models, and reporting to users the uncertainties in geospatial data. We describe how model views can be applied to conveying the uncertainty in geospatial information to decision makers.

2 CASE STUDY: CROSS COUNTRY MOBILITY

As a case study to illustrate the challenges and opportunities of uncertainty management in geospatial information systems, we focus on Cross Country Mobility (CCM) analysis. CCM analysis is performed to evaluate the feasibility and desirability of enemy and friendly courses of action. The CCM Tactical Decision Aid (TDA) predicts the speed that a specific military vehicle or unit can move across country (off roads) based on the terrain. The terrain factors that influence CCM speed are slope, soil type, soil wetness, vegetation and vegetation attributes, ground or surface roughness, and presence of obstacles.

There are several CCM analysis models commonly in use by military organizations in the U.S. and around the world. The CCM product of Figure 1 was produced using the DMA CCM algorithm (DMS, 1993). CCM products can be generated for specific vehicle types, for classes of vehicles, or for military unit types. The products can be used as inputs to algorithms for producing mobility corridors, or combined with other information to generate avenues of approach for friendly or enemy forces. Traditional CCM algorithms use point estimates of their input data and produce point estimates of predicted speeds. Traditional CCM displays show predicted speeds without any attempt to estimate or communicate the quality of the prediction based on the quality of the underlying data and the quality of the algorithm used to make the prediction.

There are many sources of uncertainty in CCM estimates. Input data on the factors that influence speed may contain errors. In many cases, the input parameters required by models may be unavailable, and must be estimated using a combination of auxiliary models and human judgment. Models for predicting speed from input parameters are imperfect. As shown below, uncertainty can have decision implications, and decision making can be improved by properly considering uncertainty in decision support algorithms.

3 MILITARY GIS DATA

A wide range of military digital mapping products (digital terrain data) are available from the DoD National

[1] http://genie.sis.pitt.edu/

[2] http://www.mapwindow.org/

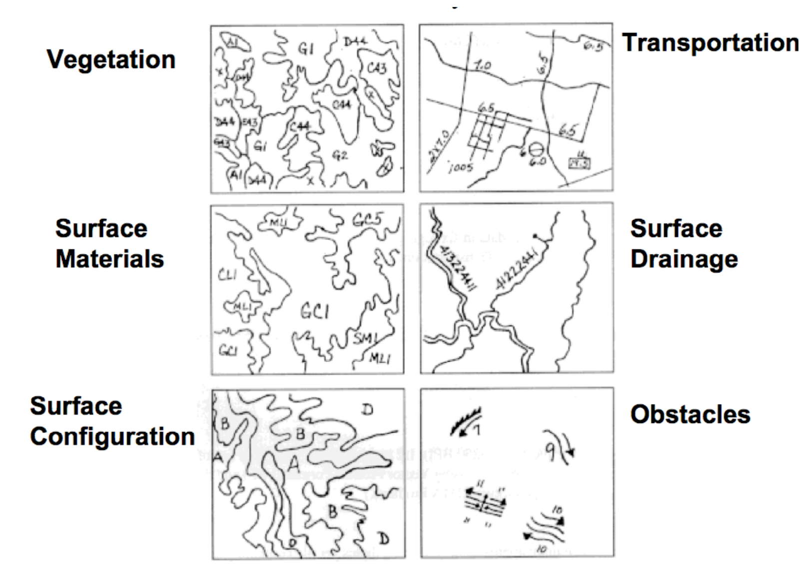

Figure 2: Information Content of Interim Terrain Data (ITD)

Geospatial Agency (NGA). Two commonly used products for military GIS analysis are Digital Terrain Elevation Data (DTED) and Interim Terrain Data (ITD).

DTED is an array of elevation values that represent the surface elevation of portion of the world. Elevation values are provided on a grid with a defined spacing in the North South and East West directions. NGA produces DTED level 1 data in cells covering an area of 1 degree by 1 degree, with a grid spacing of 3 arc seconds (approximately 100 meters at the equator). DTED level 2 is produced over smaller areas with a grid spacing of 1 arc second (approximately 30 meters at the equator). Higher resolution DTED at levels 3, 4, and 5 are available in limited areas. DTED is widely used for visualization and Line of Sight (LOS) applications.

ITD is the most widely available feature data in use by military GIS systems today. It was originally developed as an interim product, while users awaited a more detailed and robust digital terrain data product. ITD is available in two forms - ITD, and VITD (Vector Product Format (VPF) ITD) - which differ in format, although much of the information content is similar. ITD is digital vector data, where terrain features are represented as points, lines and polygons. Each terrain feature has a number of feature attributes defined for it. Figure 2 shows a graphic that illustrates the information content of ITD. Information is provided in six thematic layers. Vegetation polygons are defined for several types of wooded areas, orchards, and agricultural applications. Vegetation attributes include vegetation stem spacing, and stem diameter. The transportation layer contains features that represent roads, bridges, railroads, airfields, etc. Attributes define road widths, construction materials, bridge length, width, capacity, etc. The surface materials layer provides polygons of soil type and an attribute for surface roughness. The surface drainage layer contains information on rivers and streams, with attributes that define width, depth, bank height and slope. Surface configuration layer contains polygons for surface slope in defined categories. The obstacle layer contains information of other terrain features (like ledges, fences, pipelines, cuts and fills) that may be obstacles to military mobility.

ITD is used for a range of military GIS applications (Terrain Analysis) including mobility products such as CCM. Although ITD data is very valuable, it is expensive to produce, requiring a great deal of human-intensive feature extraction. NGA has recognized the inability to provide widespread coverage of ITD (or ITD-like data) in support of worldwide military operations. The NGA concept for future terrain data support envisions large area coverage of a subset of quickly produced data (Foundation Feature Data - FFD) to meet the military's immediate planning needs, and rapid production of more complete data (Mission Specific Data Sets - MSDS) to meet specific requirements in a crisis. This concept requires the ability to combine information from multiple sources to produce the needed products. The available inputs may be of varying quality and resolution. It is essential both to employ a sound methodology for propagating the uncertainties in the different inputs, and to communicate properly to end users the uncertainty in results.

4 PROPAGATING UNCERTAINTY

Figure 3 shows an example, taken from (Wright, 2002), of a Bayesian network (BN) for integrating data from different sources into an integrated vegetation cover map, an important input into a CCM tactical decision aid. Information was fused from ten sources, including digital elevation data; geology data; forest and vegetation maps; and various images from the years 1977, 1987, and 1988. This Bayesian network applies to a single pixel, and is replicated for each pixel in the data set. A more sophisticated model must be applied when errors at different pixels are not independent. For example, blurring of an image can introduce correlations between neighboring pixels, and registration errors can introduce a bias that affects all pixels in a given region. A graphical model for fusing elevation data described by Wright (2002) used undirected arcs to model spatial auto-correlation, and included random variables to represent vertical bias in elevation measurements. If spatial correlation and bias were considered serious sources of error, the model of Figure 3 could be extended in a similar manner.

To perform the kind of analysis described here, the fusion system must have the necessary information to characterize the quality of the input data sources.

Metadata that represents data quality information enables producer and consumer to communicate information about data quality needed for fusing that data with data from other sources.

The BN of Figure 3 also makes use of geology, topography, soils, and image data (or results from algorithms run on images). In order for this scheme to work, all data sources must publish relevant data quality information as metadata. Furthermore, all sources must describe appropriate structure (relationships between themes, and common image sources for products). That is, the metadata must include not just simple data quality attributes for results, but also the necessary structural information to enable a probabilistic reasoner to construct the appropriate Bayesian network for drawing inferences about vegetation cover. We have argued elsewhere (e.g., Costa, et al, 2007) that this information should be represented as a probabilistic ontology (PO).

An ontology specifies a controlled vocabulary for representing entities and relationships characterizing a domain. Ontologies facilitate interoperability by standardizing terminology, allow automated tools to use the stored data in a context-aware fashion, enable intelligent software agents to perform better knowledge management, and provide other benefits of formalized semantics. However, as described in (Costa, 2005), standard ontology formalisms do not provide a standardized means to convey both the structural and numerical information required to represent and reason with uncertainty in a principled way. POs, on the other hand, are designed for comprehensively describing knowledge about a domain and the uncertainty associated with that knowledge in a principled, structured and sharable way. Therefore, POs provide a coherent representation of statistical regularities and uncertain evidence, an ideal way of representing and propagating uncertainty in geospatial systems. Like a traditional ontology, a PO represents types of entities that can exist in a domain, the attributes of each type of entity, and the relationships that can occur between entities. In addition, a PO can represent probability distributions. This requires more than the simple ability to represent uncertainty about the attributes of entities of a given type. POs represent conditional dependencies on other attributes of the same or related entities, as well as uncertainty about the types of entities and the relationships themselves. PR-OWL (Costa, 2005) is an upper ontology, written in the OWL ontology language, that enables an OWL ontology to represent such relational uncertainty.

As an example, consider the problem of aggregating geospatial information from several databases. Suppose we consult three different databases, all three of which label a particular area as forested. Each report is tagged with a particular credibility. Because the three reports agree, standard statistical aggregation technologies would label the region as forested and assign a higher credibility than the three individual credibilities. However, if all

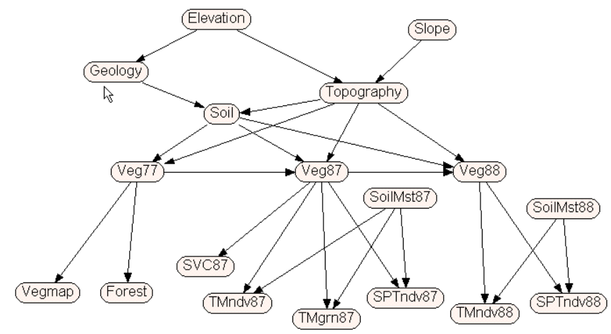

Figure 3 – Bayesian Network for Information Integration

three databases obtained their raw data for this area from the same satellite image, and all three applied similar algorithms for assigning a ground cover type label. In this situation, the credibility of the aggregate report is no greater than any of the individual input credibility values. In this case, we need to represent not just a single credibility number, but dependency information about how the credibility depends on the sensor and the data processing algorithm. If there is uncertainty about the source of the data in one of the databases, then the appropriate combination rule would be a probability weighted average, with weights equal to the posterior probability, given the observed data, of the different data sources. If the systems providing input give no data quality information, or supply insufficient information for a probabilistic reasoner to determine unambiguously the structure and/or probabilities for the Bayesian network, then the fusion system has an additional inference challenge – to determine the appropriate BN for fusing the diverse inputs.

A standard ontology annotated with probabilities could not represent these complex kinds of dependence relatinoships. A probabilistic ontology could, provided that it is based on a sufficiently expressive probabilistic logic. POs provide a flexible means to express complex statistical relationships, a crucial requirement for dealing with uncertainty in geospatial systems. Reasoners capable of handling general-purpose relational probabilistic models are not yet generally available. To compute the results shown in Figure 5, custom application was written to apply the Bayesian network of Figure 3 to each pixel in a geographic database, using an application programmer interface to a Bayesian network tool. Today, this example could be computed using the Bayesian plugin to MapWindow. More sophisticated models including spatial correlation and bias would still require custom software, although new theory and tools are emerging rapidly.

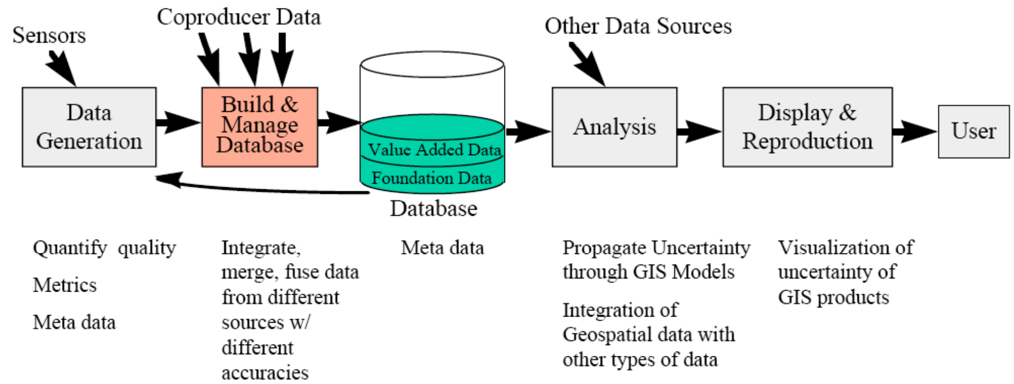

Figure 4: Conceptual Model for Mangement of Uncertainty in GIS Data

5 MANAGING UNCERTAINTY IN GIS DATA

There are errors, or uncertainties, in all geospatial data. Different kinds of uncertainties in geospatial data include uncertainties due to positional error, feature classification error, resolution, attribute error, data completeness, currency, and logical consistency (Kraak, & Ormeling, 1996). Unfortunately many of these types of uncertainty are difficult to quantify, and are often ignored in the production of GIS products - even for military applications.

Positional errors, absolute and relative errors in X,Y, and Z, are reasonably well understood and for most military geospatial data are fairly well defined. For many applications, like targeting and navigation, estimates of positional accuracy are sufficient to evaluate the suitability of the GIS data for use. Other GIS products, that depend on position are more complicated.

For example, the LOS product depends on the Z location (elevation) of the observer, a potential target, and multiple terrain points. LOS does not depend on absolute elevations, but on relative elevations of the various points. Unfortunately, acceptable relative elevations errors are not specified for DTED level 1 and 2 products, and are not used to estimate the uncertainty in LOS predictions.

Uncertainty due to feature classification errors and feature attribute errors are also commonly neglected in military GIS analysis. The product specification for ITD (and for related feature products) does not provide any standards for feature classification accuracy or feature attribute accuracy. The accuracy of the different thematic layers is in general unknown, although some studies have been done (Ryder, and Voyadgis, 1996) and results from civilian studies may be used as a guide. In general, estimation of terrain features like vegetation and soil type from imagery source materials - without extensive “ground truth” is very difficult. Results which achieve 80% accuracy are considered good. Terrain products, produced from terrain feature data, will be in error as a result of propagation of the uncertainties in the terrain data through the algorithm, or model, used to create the product. Today, military GIS systems typically do not attempt to estimate the uncertainty in GIS products, and have no way to incorporate uncertainty in algorithms or display it to users.

Other uncertainties due to data resolution, completeness, or consistency are also present in military GIS systems. Although users (terrain analysts) may be aware of these uncertainties, there is no systematic way to account for them or to communicate them to decision makers.

Figure 4, taken from Wright (2002), presents a model of the lifecycle of geospatial data, showing the management of uncertainty operations that are required at each stage. The first block, data generation, is the creation of geospatial data from source materials, often remote sensors. During this step, tools and techniques that measure the quality of data, as it is produced, are needed. Quality must be measured in appropriate quality metrics, and recorded as part of the meta data for the data. The next two steps, build and manage the database and the database itself, are important parts of the process that are often overlooked. Today we rarely generate all new data for a particular GIS project. In almost all cases existing data will be available, and there will be new data produced by other organizations. All this data must be integrated into a cohesive database. The data integration required to merge these different data layers is a critical and complex operation. In addition to merging the data, we need to merge the corresponding meta data as well, to derive meta data for the new integrated data.

The database, where available data is stored, is also explicitly shown as not “full.” Usually we will not have all the information we would like to have before we start to generate GIS products. Over time, as additional data is ingested, the database will contain more data - but usually our appetite for new data is also growing, so the data store will never be full. As the availability of data changes over time, the meta data must be updated to reflect the quality of currently available data.

The next block, analysis, is the application of GIS operations, according to some model, to produce a GIS product. Techniques for propagating the uncertainty in geospatial data through the GIS model into the GIS product are required. In the following block, the GIS product is displayed or presented to the user. In this step, it is important to present the user with a visualization of the uncertainty in the product. One of the challenges is to find good ways to present such information.

The final block in the geospatial life cycle is the user. This block also is an important step in managing uncertainty: ensuring that users are trained to ask for and use information about the quality of GIS products as part of their decision process.

6 VISUALIZING UNCERTAINTY

Visualization of uncertainty in GIS products is essential to communicate uncertainties to decision makers. This helps to prevent decision makers from being blinded by the quality of the display, and to make them aware of the underlying uncertainty of the product.

A few examples of uncertainty visualization ideas, taken from Wright (2002), are presented here. Figure 5 shows a fused vegetation map that displays the results of applying the Bayesian network model of Figure 3. The display shows color-coded highest-probability classifications, and provides the ability to drill down to view the uncertainty associated with the fused estimate. Figure 6 shows a visualization of CCM with associated uncertainty. The underlying computations for this display were performed by implementing a standard CCM algorithm as the Bayesian network shown in Figure 7 (Wright, 2002).

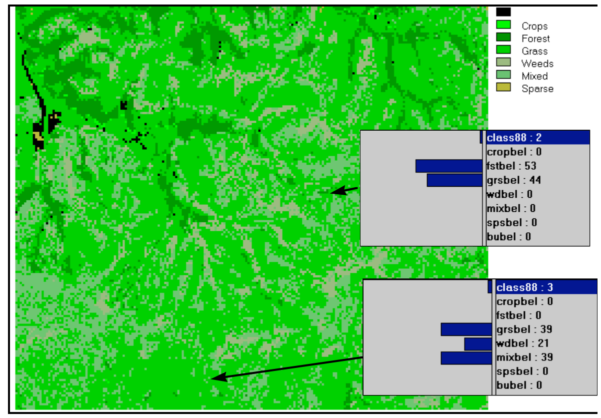

Figure 5 – Fused Vegetation Map for 1988

CCM uncertainty is shown two ways, in the legend and via interactive histograms that the user can control. The bi-variate legend uses color to represent the predicted CCM speed range. The quality of the color represents the quality of the prediction. There is enough information in the legend that it is difficult to interpret the product colors. This difficulty is exacerbated by the difficulty of matching colors from computer monitor to printed hardcopy. To offset the difficulty in interpretation, user

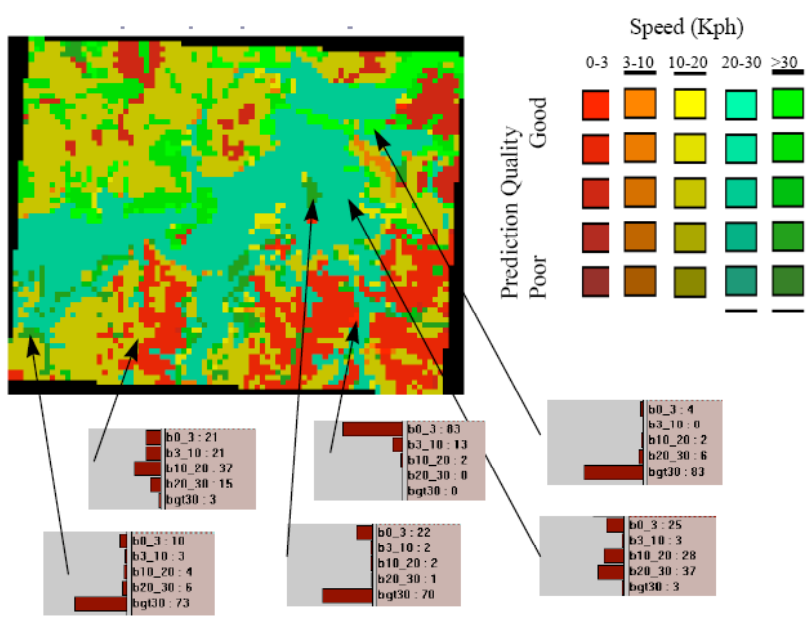

Figure 6: CCM Product with Visualization of Uncertainty in the CCM Prediction

controlled popup histograms were provided on the digital display. Several examples are shown in figure 7. The popup histograms are useful to illustrate how the legend works:

- For each pixel in the product display, a probability distribution for predicted CCM speed was generated (via Bayesian Network), based on uncertainties in the original feature data layers.
- The pixel color (legend column) was selected that corresponds with the highest probability speed bin.
- The prediction quality color (legend row) was selected based on the range of speed bins with probability equal or greater than 10%.
- For example, the top row, right histogram is for a bright green pixel, indicating that the predicted speed is reasonably fast, and there is little uncertainty. The bottom row, left histogram is also for a green pixel, indicating that the highest probability is for a fast CCM speed. However, there is also a 10% probability that the correct CCM speed range is the lowest speed bin, so the quality color of this pixel reflects that the actual CCM speed extends across the entire range of speeds.

This CCM display provides more information to decision makers about the quality of the prediction and (in the interactive versions) the popup histograms provide a means to query for more detailed predictions at specific points.

One type of query cannot be answered by the popup histograms of Figure 6. If the decision maker is interested in reducing the uncertainty in the CCM predictions - perhaps by allocating reconnaissance resources to collect additional terrain data, he would like to know the influence of individual terrain factors on the total uncertainty in the CCM prediction. The query is: “what terrain factor contributes the most to the uncertainty in the predicted CCM speed?” Figure 8 shows an additional visualization that makes it possible to answer this query.

The figure represents the uncertainty in the values of the terrain factors for one specific point on the terrain, as well as a graphical depiction of the impact of each of the individual factors. The visualization requires input of the probability distribution that describes the current estimate of the terrain parameters at a point. These probability distributions are used in a Monte Carlo technique to map variation in terrain inputs into variation in predicted CCM speed. The graphic output shows four small graphs that map each individual terrain parameter's effect on the CCM speed, assuming all other terrain parameters remain fixed (at the mean of their distribution). These small graphics each contain the curve of terrain value vs. CCM speed, and two histograms. The one on the bottom is the random variation of the terrain parameter, the one on the left is the resulting variation in the predicted CCM speed. Note that if the terrain parameter vs CCM speed curve is flat (or nearly flat) then there is very little variation in predicted CCM speed, even for large variations in terrain values. If the terrain parameter vs predicted CCM speed curve is steep, then there can be large variation in predicted CCM speed even if there is little uncertainty in the terrain values. The large histogram at the bottom shows the total distribution of predicted CCM speeds based on the combined variation of all the terrain inputs. The total distribution of predicted CCM speeds shows more variation in predicted speed than for any of the individual terrain parameters, because of the random combination of values and interaction between parameters.

In the visualization shown, - for this specific set of terrain inputs, and terrain uncertainties - the effects of errors in slope, stem spacing, and soil strength (Rating Cone Index - RCI) have only a small impact on the total uncertainty in predicted CCM uncertainty. The influence of stem diameter uncertainty, on the other hand, has a fairly large impact on the uncertainty in predicted CCM speed.

This kind of visualization could be used as an interactive guide during data collection: For a given area, and given the current best estimate of terrain values and terrain accuracies, it is possible to determine which terrain factor will provide the most improvement as a result of additional collection effort.

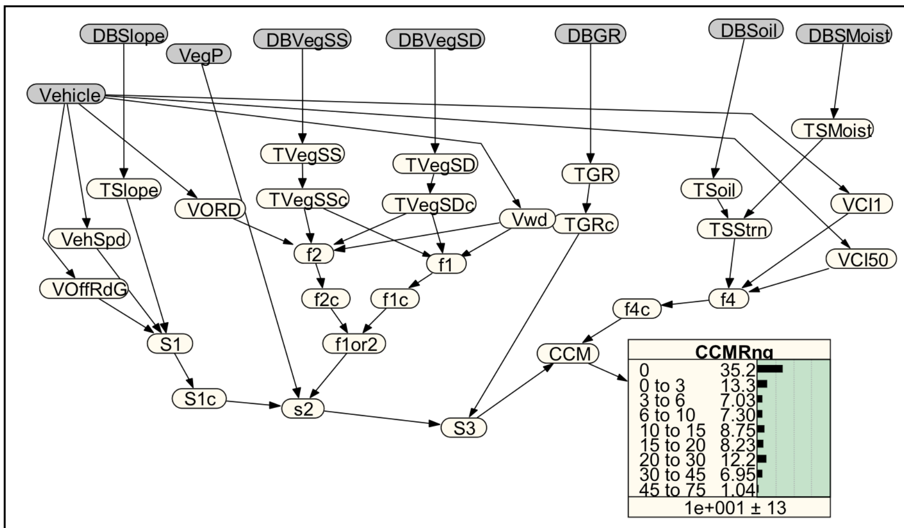

Figure 7: Bayesian Network Implementation of CCM Algorithm

The above ideas regarding possible visualizations of uncertain, incomplete data uncover another vital issue for a successful geospatial system – its ability to meet the distinct knowledge requirements of its distinct users. Note that we are not addressing cosmetic GUI customizations, but a much complex issue. The multitude and diversity of users relying upon a wide spectrum of possible features of a geospatial system suggests the need of a much richer approach for predicting the answers that have to be provided, the granularity of information sought by each type of user, or even the algorithms that need to be run to meet such requests. Merely listing types of users and crafting customized reports does not scale to geospatial systems intended to meet GIG-era requirements. A more flexible solution is required.

One possible approach to face the above challenge might be to employ an ontology conveying knowledge of patterns of system usage, which would trace characteristics related to each type of user to the particular aspects regarding the situation in which a given service is being requested. Depending on how rich this ontology is, the system would be able to predict parameters such as the user's decision level, precision, timelessness and expected granularity of information, most important factors for CCM predictions, etc, and then optimize its resources to provide the most adequate level of service to that specific situation (e.g. by selecting the most appropriate model views, fine-tuning plausible algorithms for CCM predictions, etc).

Finally, in order to meet the demands of a network-centric environment, a service-oriented architecture similar to the one in (Costa et al., 2007) is implied as a precondition for a ontology-driven, seamless interoperable employment of multiple, distributed information sources, repositories, and users of a geospatial system.

7 DECISION IMPLICATIONS

A simulation experiment reported by Wright (2002) demonstrates the importance of properly accounting for uncertainty in CCM calculations. Two versions of a CCM product were generated from the operational terrain database (original terrain data). Both used the same CCM algorithm, but the first used the algorithm directly and did not estimate the uncertainty of the CCM product, whereas the second used the Bayesian Network implementation shown in Figure 7. The data quality information used to generate the uncertain CCM product was the same as that used to generate the simulated terrain databases, with the exception that the Bayesian Network CCM process does not account for spatial accuracy.

Simulated agents without access to uncertainty information used a standard A* search algorithm to find the fastest route from start to finish, and applied a "padding factor" (a parameter varied in the experiment) to determine a start time that predicted them to arrive at the finish point ahead of schedule by an amount determined by the padding factor. Agents with access to uncertainty information estimated a distribution of arrival times. If this distribution was "too wide," they were able to perform "reconnaissance" to reduce the uncertainty, and then replan their routes. Their estimated travel time at the 90th percentile of the travel time distribution, and also applied a "padding factor." As shown in Figure 9, taken from Wright (2002), results of this experiment showed a

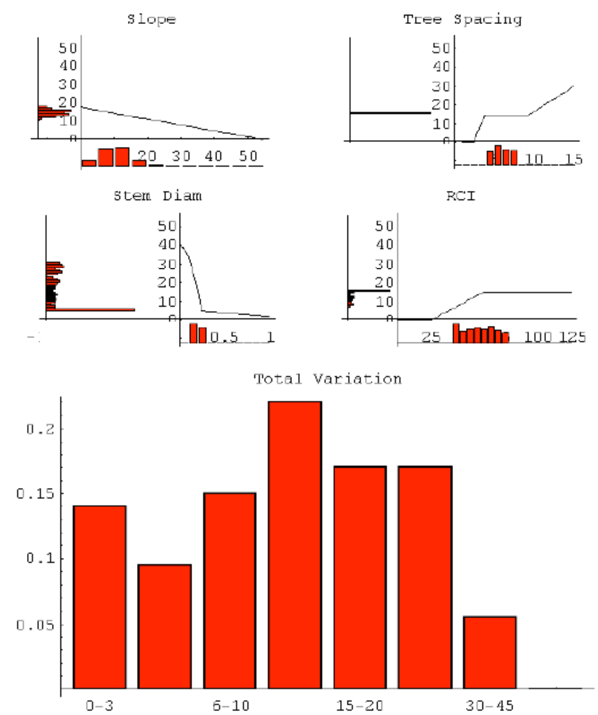

Figure 8: Visualizing The Influence of Terrain Factors on Total Uncertainty in CCM Predictions

dramatic improvement in the probability of arriving at the destination on time for agents that had access to the uncertainty information.

8 DISCUSSION

The examples demonstrate the importance of representing, properly managing, and communicating to decision makers information about uncertainty in the GIS products used for military planning. Several prerequisites are required. The quality of the geospatial data must be known, or techniques must be available to estimate data quality. If a "ground truth" data set exists, in which values are available for all random variables of the network, then straightforward parameter learning algorithms can be used to estimate the required parameters. Typically, though, some of the random variables will be unobserved hidden variables. In this case, more sophisticated algorithms are needed for learning in the presence of hidden variables

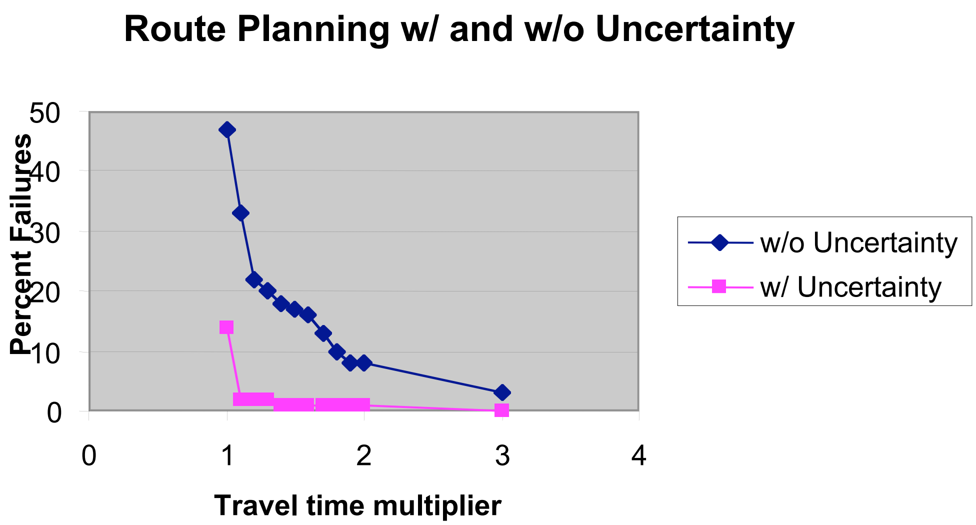

Figure 9. Results of the Value of Understanding Uncertainty experiment.

(e.g., Friedman, 1998; Laskey, et al, 2003). In addition to representing data quality, techniques must be available to propagate uncertainty of the data through GIS algorithms to estimate the uncertainty in the product. For example, the Bayesian network of Figure 7 was used to propagate uncertainty through the CCM model.

Different model views are appropriate for users playing different roles in the uncertainty management process shown in Figure 4. Model developers and implementers need access to the Bayesian network models of Figures 3 and 7, as well as to statistical models used to estimate the probability distributions that go into the models. End users need to see views of the model results that are tied to their familiar ways of interacting with the data. The displays of Figures 5 and 6 are constructed to be similar to traditional map displays, but to provide additional information about uncertainty as part of the display, and to allow users to drill down to a more detailed explanation of particular uncertainties. Figure 8 shows one kind of drill-down that decision makers might find useful.

The analyses and displays shown here were generated as stand-alone applications, and have not been incorporated into military geospatial analysis tools, into geospatial ontologies, or into decision support products. It is possible to carry out the kinds of analysis described in this paper with technology available today, however both impose costs on the production and use of geospatial data. So the final prerequisite is an organizational decision that providing information about the uncertainty of GIS products is important - that it provides benefits that exceed the costs.

References

Ames, D.P., *Bayesian Decision Networks for Watershed Management*. PhD Dissertation, Department of Civil and Environmental Engineering, Utah State University, 2002.

Costa, P.C.G., *Bayesian Semantics for the Semantic Web*. Doctoral Dissertation, School of Information Technology & Engineering, George Mason University: Fairfax, VA, USA, 2005.

Costa, P.C.G., Laskey, K.B., Laskey, K.J., and Wright, E.J. Probabilistic Ontologies: The Next Step for Net-Centric Operations, *Proceedings of the Twelfth International Command and Control Research and Technology Symposium*, 2007, to appear.

Cressie, N. A. C. *Statistics for Spatial Data*. New York, John Wiley & Sons, 1993.

DMS, "Procedural Guide for Preparation of DMA Cross-Country Movement (CCM) Overlays", Defense Mapping School Student Text DMS ST 330, Ft. Belvoir, VA, 1993.

Friedman, N. Learning Belief Networks in the Presence of Missing Values and Hidden Variables. Fourteenth International Conference on Machine Learning (ICML-97), San Mateo, CA: Morgan Kaufmann Publishers, 1998.

Goodchild, M.F., "Closing Report NCGIA Research Initiative 1: Accuracy Of Spatial Databases l", National Center for Geographic Information and Analysis, University of California, 1992.

Grêt-Regamey, A. and Straub, D. Spatially explicit avalanche risk assessment linking Bayesian networks to a GIS. *Natural Hazards and Earth System Science 6*(6), pp.911-926, 2006.

Jianwen, M., Qin, D. Migratory locust hazard monitoring and prediction using the Bayesian network inference. *Proceedings of the International Geoscience and Remote Sensing Symposium* (*IGARSS*) Page(s): 3623 – 3626, 2005

Kraak, M., and Ormeling, F. Cartography, Visualization of Spatial Data, Addison Wesley Longman Limited, Essex, England, 1996.

Laskey, K.B. and Myers, J. Population Markov Chain Monte Carlo, *Machine Learning 50*(1-2), 1993.

Lunetta, R.S., Congalton, E.S., Fenstermaker, Jensen, J.R., MCGwire, K.C., and Tinney, L.R. Remote Sensing and Geographic Information System Data Integration: Error Sources and Research Issues. PE&RS 57(Jun), 677-687, 1991.

Ryder, W.H. and Voyadgis, D.E., *Measuring the Performance of Algorithms for Generating Ground Slope*, USATEC Paper: Presented at DCAC TEM Feb 1996.

Walker, A.R., Pham, B., Moody, M. Spatial Bayesian Learning Algorithms for Geographic Information Retrieval, *Proceedings of the 13th annual ACM international workshop on Geographic information systems*. Bremen, Germany, 2005.

Wright, E. *Application of Bayesian Networks for Representing Uncertainty in Geospatial Data*. ASPRS Spring Convention, Tampa, FL, 1998

Wright, E.J., *Understanding and Managing Uncertainty in Geospatial Data for Tactical Decision Aids*. Doctoral Dissertation, School of Computational Sciences, George Mason University: Fairfax, VA, USA, 2002.