International Journal of Approximate Reasoning 51 (2010) 209–223

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Envisioning uncertainty in geospatial information

Kathryn Blackmond Laskey [a,*], Edward J. Wright [b], Paulo C.G. da Costa [a]

[a] SEOR Department and C4I Center, George Mason University, MS 4A6, 4400 University Drive, Fairfax, VA 22030, United States
[b] On Line Star, Incorporated, 2515 Red Cedar Drive, Bowie, MD 20721, United States

ARTICLE INFO

Article history:
Received 23 March 2008
Received in revised form 17 May 2009
Accepted 20 May 2009
Available online 21 June 2009

Keywords:
Geospatial reasoning
Geographic information systems
Geospatial metadata
Probabilistic ontologies
Multi-entity Bayesian networks

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 essential to using it intelligently. A systematic approach to data quality requires: estimating and describing the quality of data as they are collected; recording the data quality as metadata; 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. There are shortcomings in the state-of-the-practice in GIS applications in dealing with uncertainty. No single point solution can fully address the problem. Rather, a system-wide approach is necessary. Bayesian reasoning provides a principled and coherent framework for representing knowledge about data quality, drawing inferences from data of varying quality, and assessing the impact of data quality on modeled effects. Use of a Bayesian approach also drives a requirement for appropriate probabilistic information in geospatial data quality metadata. 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 been replaced by 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.

* Corresponding author. Tel.: +1 703 993 1644; fax: +1 703 993 1521.
E-mail addresses: klaskey@gmu.edu (K.B. Laskey), ewright@onlinestarinc.com (E.J. Wright), pcosta@gmu.edu (P.C.G. da Costa).
URLs: http://ite.gmu.edu/~klaskey (K.B. Laskey), http://c4i.gmu.edu/~pcosta (P.C.G. da Costa).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2009.05.011

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. AAGTs are changing the way military operations are conducted. The development of tools is shaped by military necessity. Furthermore, as tools become broadly available, the decision making process itself is 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 significant 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 attention in the academic GIS community (cf., [9]). All geospatial data contain errors. These include positional error, feature classification error, poor resolution, attribute error, data incompleteness, lack of currency, and logical inconsistency [13]). Studies have shown that 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 ([18]).

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., [6]), and these methods are widely applied to geospatial reasoning problems. There is a growing interest in the application of Bayesian networks (BNs) in the domain of geospatial reasoning. This is reflected by a number of specific applications of BNs to analyses of geospatial data, geographic information retrieval, and other geospatial reasoning problems (e.g., [12,21,22]). A Bayesian analysis plugin, based on the GeNIe/SMILE[1] Bayesian network system, has recently been released for the open-source MapWindow™ GIS system.[2] Applications of BNs to geospatial reasoning include avalanche risk assessment [10], locust hazard modeling [11], watershed management [2], and military decision support [15,23,24].

This paper steps back from specific applications to the broader problem of representing, reasoning about, and understanding uncertainty across the entire life cycle of geospatial data, and to the importance of model views in this broader problem. Graphical probability models provide a theoretically principled and computationally attractive methodology for representing uncertainty in geospatial problems. In his dissertation on the application of Bayesian networks to tactical military decision aids, Wright [24] 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. Within our specific application domain of geospatial analysis for military decision support, we present several examples illustrating the importance of representing and managing uncertainty. We also illustrate the use of model views to convey the uncertainty in geospatial information to decision makers. More generally, the lifecycle approach we advocate and the issues we discuss are relevant to a broad range of geospatial reasoning and decision support problems.

2. Example: cross country mobility

As an example illustrating 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. A CCM Tactical decision aid 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.

A wide range of military digital mapping products (digital terrain data) are available from the DoD National Geospatial Agency (NGA). Two common types of data used for military GIS analysis are elevation data (typically Digital Terrain Elevation Data (DTED)) and feature data (typically Interim Terrain Data (ITD) available in a variety of formats).

Elevation data is typically formatted as an array of elevation values that represent surface elevations of some portion of the world. Elevation values are provided on a grid with a defined spacing in the North South and East West directions.

Feature data, representing terrain features on the Earth's surface, is formatted as 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. Fig. 1 shows the kind of display typically used to depict the content of a military feature data set. 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 fea-

[1] http://genie.sis.pitt.edu/.
[2] http://www.mapwindow.org/.

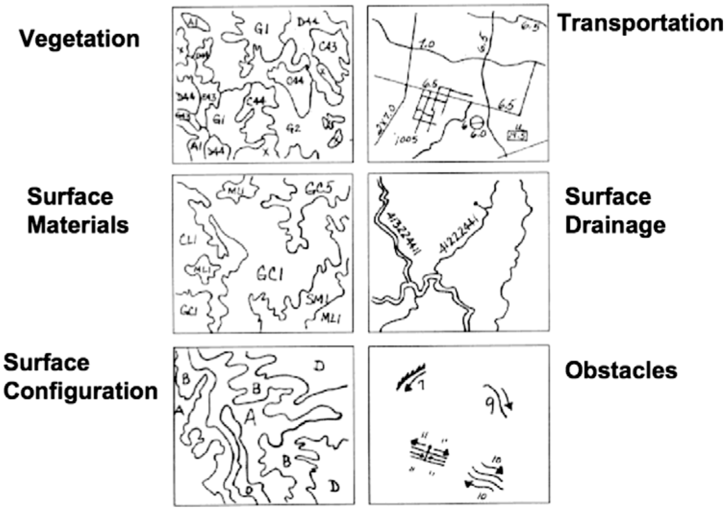

Fig. 1. Information content of Interim Terrain Data (ITD).

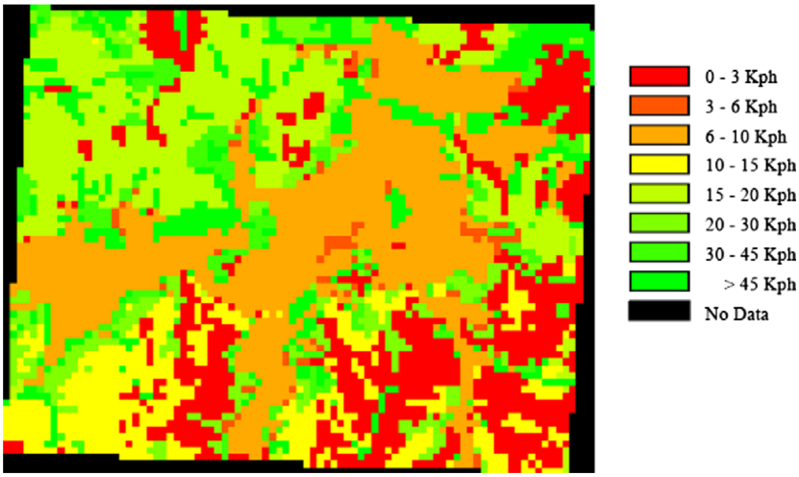

Fig. 2. Traditional CCM Product (M1 Tank, DMA Mobility Model, ITD Data, Korea).

tures 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. The surface configuration layer contains polygons for surface slope in defined categories. The obstacle layer contains information about other terrain features (such as ledges, fences, pipelines, cuts and fills) that may be obstacles to military mobility. The accuracy of the different thematic layers is in general unknown, although some studies have been done [20], and results from civilian studies may be used as a guide.

There are several CCM analysis models commonly in use by military organizations in the US and around the world. The CCM product of Fig. 2 was produced using the ETL CCM algorithm [19]. CCM products can be generated for specific vehicle types, for classes of vehicles, or for military unit types. The products may 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.

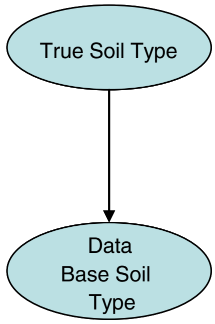

Fig. 3. A two-node Bayesian Network for a database entry as evidence for the true value of soil type.

	Reference Data				
Classified Data	A	B	C	D	row total
A	65	4	22	24	115
B	6	81	5	8	100
C	0	11	85	19	115
D	4	7	3	90	104
col total	75	103	115	141	434

Fig. 4. Notional error counts from estimating soil type from imagery.

3. Representing uncertainty

The data elements residing in a GIS are imperfect estimates of an uncertain reality. Knowledge about an uncertain data element can be represented as a probability distribution across a range of possible states.[3] Consider the example of soil type, one of the attributes of the polygons in the surface material layer of ITD. In ITD, and most other defense databases, soil types are defined using the USCS classification system (Unified Soil Classification System). This system defines 15 soil types with different engineering properties. For a particular surface material polygon in an ITD dataset, there is an attribute ‘Soil Type’ that contains one of the USCS symbols. Because there is uncertainty in every geospatial database, the reported values for this attribute are imperfect estimates of the true soil type. We can represent this situation with a two-node Bayesian network, as shown in Fig. 3.

In Fig. 3, the database value of the soil type attribute is evidence for the true soil type. To define this BN, we need a prior distribution on true soil type, which can be estimated from available geographic data, and the conditional probability distribution (CPD) for the database soil type given the true soil type. There are several approaches to specifying this CPD. Often, categorical geographic data are generated from imagery or other sources by applying a classification algorithm. In most cases, the classification algorithm has been evaluated against ground truth to generate a classification accuracy or error matrix. A simplified example of an error matrix is shown in Fig. 4. The counts in the error matrix can easily be converted to the probabilities needed to define the CPD required in the BN of Fig. 3. If an error matrix is not available, the CPD can be estimated by comparing the data generation process to similar processes, or by using judgment of human experts who are familiar with the process and its use in practice. For example, one could elicit the error matrix that the expert would expect to get if a rigorous comparison of a data layer to ground truth were available. Eliciting knowledge from human experts is itself an important research topic, a full treatment of which goes beyond the scope of the present paper.

To use this BN, the database is queried for the soil type attribute. The returned value is used to instantiate the ‘Data Base Soil Type’ node. The BN algorithm then propagates the evidence to the ‘True Soil Type’ node, applying Bayes rule to infer a probability distribution for the true soil type given the database soil type.

The simple BN of Fig. 3 can be extended to deal with some other common GIS operations. Fig. 5 shows a BN for combining two soil type data layers to generate an integrated estimate of the true soil type. Each database node has a CPD, derived from an error matrix or from expert judgment, to represent the conditional distribution of the database value given the true value of the corresponding feature. When the two database nodes of the BN are instantiated with the values from the two databases, the BN algorithm propagates the evidence to the true soil type node. Again, this results in a probability distribution across the range of possible values. This BN approach for data integration appropriately weights the results according to the accuracy of each of the input databases. The approach can easily be extended to include additional data sets.

This example can be extended to illustrate additional complications in combining geographic data. ITD uses the USCS soil classification system, but many other soil classification systems are in use around the world. Many of the other systems are designed to support agriculture applications, and so have different classification criteria from those used by the USCS. Sup-

[3] While this distribution can be discrete or numerical, this paper focuses primarily on categorical variables.

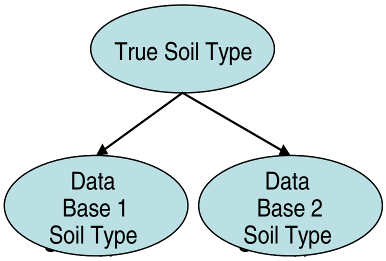

Fig. 5. A simplified BN for integrating two soil type data layers.

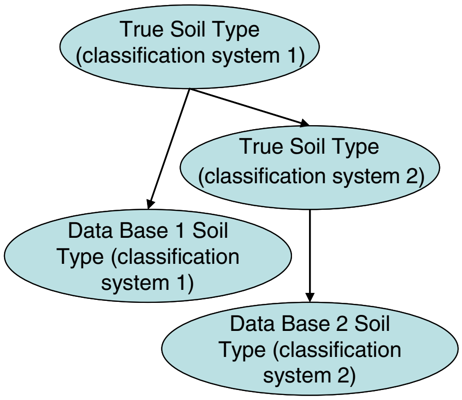

Fig. 6. A Bayesian network for combining multiple data layers across different classification systems.

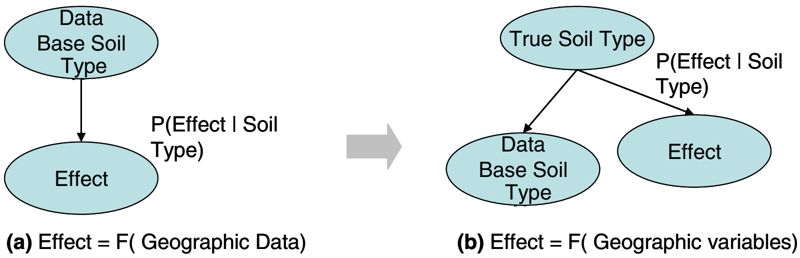

Fig. 7. Estimating effects based on geographic data (a) vs. appropriately propagating the data uncertainty into an estimate of the effects (b).

pose we wish to integrate two soil type data layers, but they are using two different classification systems. A common approach is to define a set of translation rules to convert one of the data layers into the desired classification system. Unfortunately, soil classification systems (and many other geographic themes) do not lend themselves to crisp conversions. There is uncertainty in the conversion process, and it should be represented. Fig. 6 is an example of how this can be done in a BN.

This BN is similar to the one of Fig. 5, with the addition of the extra node for the true soil type in the second classification system. To define the BN, this node needs a CPD that defines the conditional probability of the true soil type in the second classification system, given the true soil type in the first classification system. This kind of CPD is capable of expressing crisp conversions, where they exist, as well as probabilistic ones where there is uncertainty. The required CPD can be developed based on expert understanding of the geographic phenomena involved, the relevant specifications for the classification systems, and available empirical data.

Geographic data are used by the military to estimate the effects of the environment on military operations. Typically a geospatial model is used to estimate the effect of interest as a function of one or more geographic variables, *Effect* = *F* (*Geographic Variables*). This geospatial model may be a simple mathematical function. More generally, it can be an algorithm or process consisting of a complex sequence of geospatial operations. In practice, the true values of the geographic variables are often unknown. Therefore, effects are typically estimated from available data by inserting the data into the geospatial model in place of the true value of the geographic variable: *Effect* = *F*(*Geographic Data*). Our proposed Bayesian approach addresses uncertainty by explicitly inferring the unknown geographic variables from the data, and then using the inferred geographic variables to estimate the desired effect. Because we are using BNs, the uncertainties, expressed as local probability distributions at each node, are appropriately propagated to a probabilistic estimate of the effects. Fig. 7 provides an illustration.

Fig. 7a shows a simple two-node network for estimating an effect as a function of geographic data. Fig. 7b shows a three-node BN in which the technique from Fig. 3 is used to infer the true geographic variables, from which the effect is inferred. Because this BN explicitly accounts for the accuracy of the geographic database, the estimated effects will appropriately include the resulting uncertainty. This BN is very simple, involving only one geographic variable. A more complex example is presented in Section 6 of this paper.

Applying BNs to geospatial reasoning requires more than just integrating BN inference engines with GIS applications. Another essential requirement is representing the information required to define the conditional probability distributions. The above examples required information about how accurately database values reflected ground truth. They also required probabilistic information about the translation between different soil classification systems. We argue that the knowledge required to define CPDs should be represented as geospatial metadata, and that a natural way to encode this metadata is as geospatial probabilistic ontologies [16].

Ideally, metadata about data quality for a geographic dataset will contain information about fitness for use, and will enable users to appropriately weight the data, based on its quality, in any geospatial model or application. For use in a BN, the metadata should include the CPD which defines P(data value | true value), or an error matrix from which this CPD can be derived. Alternatively, the metadata should provide enough detail about the sources and processes used to generate the data, that a knowledge agent (human or automated) can reasonably construct a CPD. Current geospatial metadata standards allow, but do not require, this kind of explicit probabilistic representation of thematic (categorical) data. The current data quality standards are defined using free text fields, useful for human understanding, but difficult for algorithms to access. To implement the methodology of this paper, current standards should be extended to support probabilistic data quality assessments in a standard format that can be processed by a computer.

Not all the knowledge required to process uncertainty will be in the metadata of the individual data sets used to compute a geospatial decision support product. For example, knowledge of the probabilistic relationship between different soil classification systems would not be found in metadata for either of the two soils data layers. What is needed is a geographic ontology that contains information about soils and the ways they can be classified. This ontology should represent the relationships between the two classification systems, or should be readily extensible to add this new relationship. In addition, if a data set does not contain the explicit CPD needed, or an error matrix from which it can be derived, then the geographic ontology should contain sufficient information about geographic variables, sources, and geographic processes, that it can construct an appropriate CPD (perhaps with human assistance).

4. Propagating uncertainty

Fig. 8 shows an example, taken from Wright [24], of a Bayesian network (BN) for integrating data from different sources into a 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. The nodes in the Bayesian network are defined as follows:

- *Elevation* is a root node that influences Geology and Topography.
- *Slope* is also a root node that influences topography.

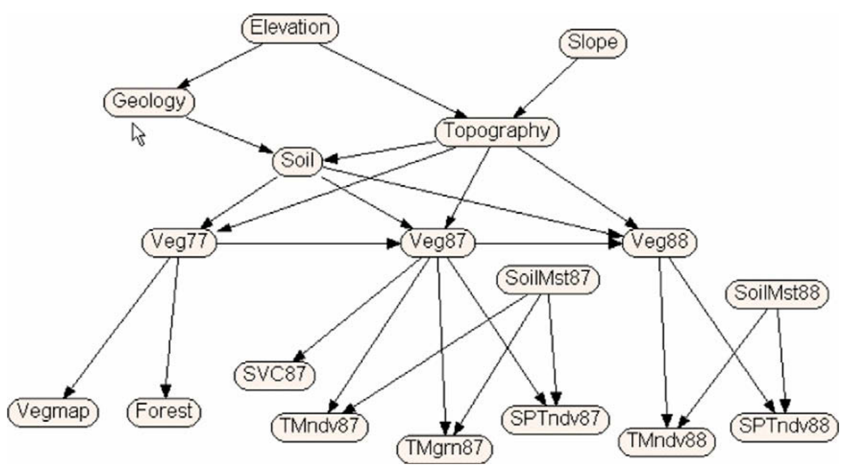

Fig. 8. Bayesian network for information integration.

- *Topography* depends on slope and elevations: possible states are bottom lands, slopes, and uplands.
- *Geology* is shown with an arc from elevation. This arc represents the ability to predict geology if elevation is known. In use, if geology is known from the geology map, elevation will have no effect on geology. If geology is not known (a void area) then elevation can be used to predict the geology.
- *Soil* represents the unknown soil type, which depends on geology and topography. Information on likely soil types based on the topography and underlying geology is provided in [7].
- The three vegetation nodes (*Veg77*, *Veg87*, *Veg88*) depend on the soil type and topography. That is, if soil type and topography are known, it is possible to predict likely vegetation cover. The links between the vegetation types (from *Veg77* to *Veg87*, and from *Veg87* to *Veg88*) represent the dependence of the vegetation type on the previous vegetation type. These arcs capture domain knowledge about potential transitions between vegetation classes. From *Veg87* to *Veg88* it is impossible for a grass area to become a forest area, but it may be possible for a grass area to become a crop area (this transition is more likely if the soil is appropriate, and the topography is bottom land).
- Information products previously derived from imagery, including a forest and vegetation map, as well as Normalized Difference Vegetation Index (NDVI) and Greenness (grn) products derived from the French Système Pour l'Observation de la Terre (SPOT) or Landsat Thematic Mapper (TM) imaging systems: *Forest*, *Vegmap*, *SVC87* (Spot Vegetation Classification), *TMndvi87*, *TMgrn87*, *SPTndvi87*, *TMndvi88*, and *SPTndvi88*. These are the information products that are being integrated. For each, the arcs from the true vegetation classes (*Veg77*, *Veg87*, *Veg88*) represent the accuracy of the product as could be expressed in an error matrix.
- The final two-nodes are soil moisture in 87 and 88 (*SoilMst87* and *SoilMst88*). These are necessary because different NDVI (and greenness) images may be different for the same vegetation class, if there is different soil moisture. For example if 1987 is a dry year and 1988 is a wet year, the NDVI may be very different even if the vegetation class is the same.

This Bayesian network applies to a single pixel, and is replicated for each pixel in the data set. The Bayesian network was used to infer soil type and vegetation for the three years represented in the figure (1977, 1987, and 1988) from the ten data sources. To compute these results, a custom application was written to apply the Bayesian network of Fig. 8 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™.

A more sophisticated model must be applied when errors at different pixels are not independent. There are many sources of spatial correlation in geospatial processes. Sources of correlation include blurring of an image, which introduces correlations between neighboring pixels, and registration errors, which introduce a bias affecting all pixels in a given region. It is possible to include these effects in the analysis, although the model and the computations required to compute with it will be substantially more complex. When there is spatial correlation, the distribution at each pixel depends on information from neighboring pixels. Therefore, applying the same Bayesian network independently at each pixel is inappropriate. Even if each pixel is correlated with only a few neighboring pixels, exact inference will generally be intractable. Wright [24] developed a graphical model for fusing elevation data that used undirected arcs to model spatial autocorrelation, and included random variables to represent vertical bias in elevation measurements. Because exact inference for this model was intractable, Gibbs sampling was used for inference. If spatial correlation and bias were considered serious sources of error, the model of Fig. 8 could be extended in a similar manner, but approximate inference methods would be required.

The BN of Fig. 8 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. Metadata enables producer and consumer to communicate information about data quality, including structural information relevant to constructing probabilistic models for combining data from different sources. We have argued elsewhere (e.g., [5]) that this information should be represented as a probabilistic ontology.

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 [3], traditional 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. Probabilistic ontologies, 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, probabilistic ontologies 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 probabilistic ontology 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 probabilistic ontology can represent probability distributions. This requires more than the simple ability to represent uncertainty about the attributes of entities of a given type. Probabilistic ontologies represent conditional dependencies on other attributes of the same or related entities, as well as uncertainty about the types of entities and the relationships in which they participate. PR-OWL [3] is an upper ontology, written in the Web Ontology Language (OWL), that enables an OWL ontology to represent relational uncertainty.

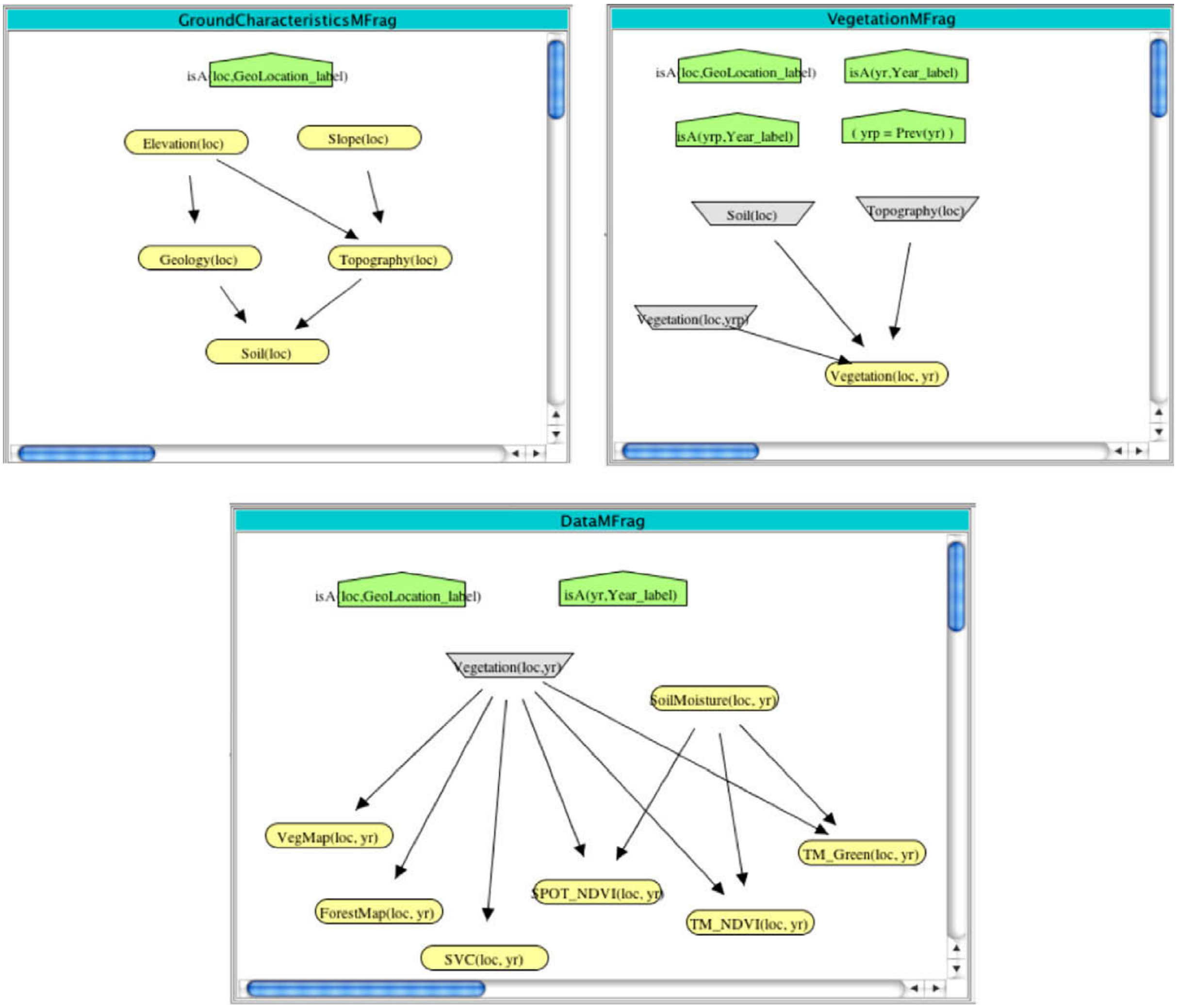

Fig. 9. Information integration model represented in MEBN and PR-OWL.

Reasoners capable of handling general-purpose relational probabilistic models are not yet generally available. However, an alpha version of an open-source graphical editor for PR-OWL ontologies has recently been released [4]. This system extends the UnBBayes[4] Bayesian network system to represent multi-entity Bayesian networks (MEBN), a first-order Bayesian logic that represents probabilistic knowledge as parameterized fragments of Bayesian networks [14]. UnBBayes includes a graphical editor for designing and editing probabilistic ontologies and saving them in PR-OWL[5] format. A PR-OWL ontology is specified as a collection of MFrags, or parameterized Bayesian network fragments that can be instantiated for multiple entities.

Fig. 9 shows screenshots from UnBBayes. The figure shows the MEBN/PR-OWL representation of a generic model from which many instances can be generated. In particular, this generic model can be instantiated repeatedly to generate the Bayesian network of Fig. 8 for each pixel in a database. The model of Fig. 9 shows three MFrags, each of which represents probabilistic knowledge about a specific aspect of the geospatial domain. The green pentagonal nodes are *context* random variables, which represent assumptions under which the distributions in the MFrag are valid. The gray trapezoids are *input* random variables, which point to random variables whose distributions are defined in other MFrags. The yellow ovals are *resident* random variables, whose distributions are defined in the MFrag.

The model represents two kinds of entities: *GeoLocations* (cells on a digital map) and *Years*. The upper left MFrag, *Ground-Characteristics*, represents attributes of a GeoLocation (which the MFrag's only context node links to the variable *loc*) as well as the interrelationships between those attributes, namely its elevation, slope, topography (bottom lands, slopes, and up-

4 http://sourceforge.net/projects/unbbayes.
5 http://www.pr-owl.org.

lands), the geology, and the soil type. These attributes are represented as resident nodes, and their conditional dependencies are represented through the arcs and Conditional Probability Tables (CPTs). The upper right MFrag, *Vegetation*, is a recursive MFrag that defines a distribution for vegetation for a specific GeoLocation and Year (respectively the variables *loc* and *yr* of the MFrag's only resident node) conditional on soil type, topography, and the previous year's vegetation. The conditioning variables are represented as input nodes. One of these, *Vegetation*(*loc*, *yr*), is recursively defined in this MFrag itself. CPTs for the other two are defined in the *GroundCharacteristics* MFrag. The third MFrag, *Data*, relates data sources on vegetation at a GeoLocation in a given Year to the ground truth value *Vegetation*(*loc*, *yr*). This latter random variable is resident in the Vegetation MFrag. In addition, a soil moisture node is included as a conditioning node, because the distribution for the satellite images may have different properties depending on the soil moisture [24].

MEBN/PR-OWL saves the generic model in a standardized PR-OWL syntax (an XML-compliant format, with RDF, OWL and PR-OWL restrictions) capable of capturing the semantics of the geospatial domain, its statistical regularities, and many other nuances a probabilistic ontology is capable of representing. The Bayesian network of Fig. 8 is constructed by instantiating Fig. 9's MFrags for one location at three different years. An automated system can store probabilistic knowledge as metadata in a PR-OWL probabilistic ontology, and use a reasoning tool such as UNBBayes-MEBN to construct a Bayesian network at each pixel to combine the data sources into a probability distribution for the soil type and vegetation at each pixel.

As another 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, this conclusion would be misleading if, for example, all three databases obtained their raw data 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. This example points to the need to represent not just a single credibility number, but information about how the credibility depends on the sensor and the data processing algorithm. If the source of the data in one of the databases is uncertain, 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 relationships. A probabilistic ontology could, provided that it is based on a sufficiently expressive probabilistic logic. Probabilistic ontologies provide a flexible means to express complex statistical relationships, a crucial requirement for dealing with uncertainty in geospatial systems.

5. Visualizing uncertainty

Visualization of uncertainty in GIS products is essential to communicating 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.

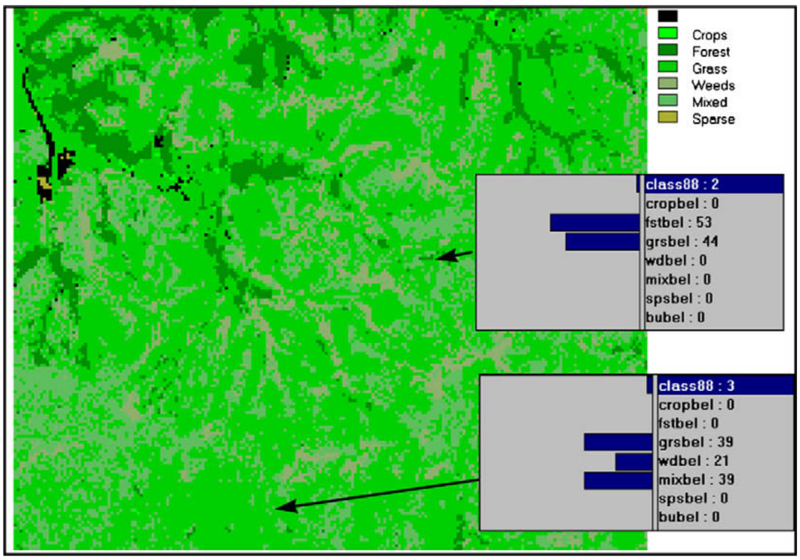

Fig. 10. Fused Vegetation Map for 1988.

Methods for visualizing uncertainty in geospatial data pose a difficult research challenge. GIS products are already complex displays, and they are just one of many inputs to a complex decision process. Current emphasis in military Command and Control is to reduce information overload. This generates understandable resistance to portraying even more information on a display already crowded with difficult-to-assimilate information. The ideal visualization may require a sophisticated set of alternative displays with which the user can interact to gain a thorough understanding of the impact of uncertainty.

A few examples of uncertainty visualization ideas, taken from [24], are presented here. Our purpose is to illustrate possibilities, not to present a definitive study on uncertainty visualization. Fig. 10 shows a fused vegetation map that displays the results of applying the Bayesian network model of Fig. 8 to each pixel on a map. The display shows color-coded highest probability classifications, and provides the ability to drill down to view the uncertainty associated with the fused estimate.

As another example, consider the cross country mobility example of Section 2. The CCM display of Fig. 2 was developed using a traditional CCM algorithm called the ETL algorithm [19]. This simple algorithm has well-known limitations. Although more complex and accurate CCM algorithms exist, they require detailed terrain data that frequently cannot be obtained. For this reason, simple algorithms like this one are often used for analysis of practical problems. We chose the ETL algorithm because it is used in practice, is complex enough to illustrate how uncertainty in terrain data can be propagated through computational GIS models, yet is simple enough that it can be presented in full.

The ETL algorithm is depicted in Fig. 11. To show how uncertainty in terrain data can be propagated through a computational GIS model, we implemented this algorithm as a Bayesian network, and then added additional nodes and arcs to represent the uncertain relationship between the true values of terrain variables and the database values. The resulting Bayesian network is shown as Fig. 12 [24].

The variables in this Bayesian network fall into four broad groups: true values of terrain variables, database values for terrain variables, vehicle parameters, and computational steps in the algorithm. The variables are described below:

- Nodes with labels that start with a 'T' (shown in light blue) represent the true values of terrain variables: Slope (TSlope), Vegetation Stem Spacing (TVegSS), Vegetation Stem Diameter (TVegSD), Ground roughness (TGR), and Soil Type (TSoil), Soil moisture (TSMoist), and Soil Strength (TSStrn). Each of these except soil strength has an associated database value. Because soil strength is not directly available in military terrain databases, it must be estimated from soil type and estimated soil moisture. The variable TSStrn is the estimate of true soil strength as measured by Rated Cone Index (RCI).
- Nodes with labels that start with a 'DB' (shown in orange) are values from the terrain database: Slope (DBSlope), Vegetation Stem Spacing (DBVegSS), Vegetation Stem Diameter (DBVegSD), Ground roughness (DBGR), Soil Type (DBSoil), and Soil moisture (DBSMoist). Soil moisture is not typically a data layer in military terrain databases, because it changes so quickly with variations in the weather. However, military terrain analysts have techniques for estimating the current soil moisture, which can be applied to populate a soil moisture layer. An additional variable in this category is a Boolean flag for the presence of vegetation (VegP).

Vehicle parameters
- VehSpd = vehicle max road speed
- VOffRdG = vehicle gradability
- Vwd = vehicle width
- VORD = Vehicle Over Ride Diameter
- VCI1 = Vehicle Cone Index 1 pass
- VCI50 = Vehicle Cone Index 50 passes

Terrain parameters
- TSlope = terrain slope
- TVegSS = Vegetation Stem Spacing
- TVegSD = Vegetation Stem Diameter
- TGR = Ground Roughness
- TSStrn = Soil Strength (Rated Cone Index)

Function CCM

s1 = VehSpd - (TSlope * VehSpd)/VOffRdG;	*Effect of slope*
f1 = (TVegSS - TVegSD - Vwd)/(2.0 * Vwd);	*Effect of maneuvering between trees*
f2 = 1.0 - ((TVegSD ^2)* Vwd)/((VORD ^2) * TVegSS);	*Effect of driving over trees*
s2 = s1 * max(f1, f2);	*Choose the faster speed*
s3 = s2 * TGR;	*Effect of ground / surface roughness*
f4 = (TSStrn - VCI1) / (VCI50 - VCI1);	*Effect of soil strength (note: strong dependence on soil moisture)*
CCM = s3 * f4;	*Max CCM speed based on combined effects of terrain features*

Fig. 11. ETL Cross-Country Mobility (CCM) model.

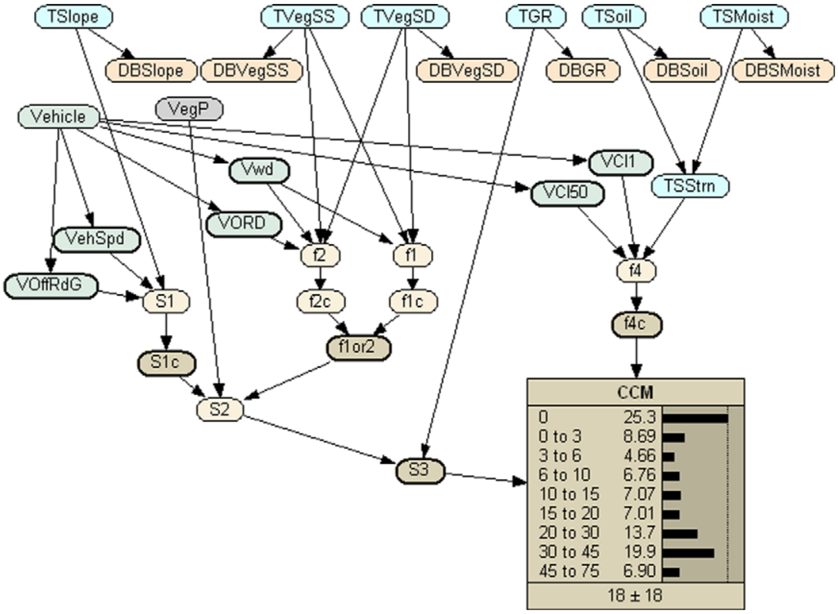

Fig. 12. Bayesian network implementation of CCM algorithm.

- Variables with labels that start with a 'V' (shown in light green) are attributes of the vehicle for which CCM is being estimated: top speed on level ground (VefSpd), Off road gradeability (VOffRdG), Override diameter – the diameter of the largest tree that the vehicle can over ride (knock down) (VORD), vehicle width (Vwd), and vehicle Cone Index for one pass (VCI1) and for 50 passes (VCI50). Cone index is a measure of soil strength. VCI1 is the required soil strength to support one pass of a vehicle; VCI50 is the required soil strength to support a column of 50 vehicles following in the same track. These vehicle parameters are available for all common military vehicles.
- The remaining variables are intermediate variables used to implement the ETL CCM algorithm:
  - The variable S1 is the vehicle speed as affected by ground slope. If the formula for calculating S1 results in a value less then zero, the variable S1c clips it to zero.
  - The variable f1 is an estimate of whether the vehicle can maneuver between tree trunks, and if so, the amount it will be slowed.
  - The variable f2 is an estimate of whether the vehicle can knock over trees, and if so, the amount it will be slowed.
  - The variables f1c, and f2c are intermediate variables that clip the values of f1 and f2 to be in the range [0, 1].
  - The variable f1or2 chooses the larger of f1c or f2c.
  - The variable S2 modifies the variable S1c by f1or2, but only if vegetation is present in the database.
  - The variable S3 modifies S2 based on the value of ground roughness.
  - The variable f4 represents the degree by which soil strength will slow vehicle movement. If the calculation for f4 results in a value less then zero (soil strength does not support the vehicle), then f4c clips the value to zero.
  - The variable CCMRng is the final result, the estimated CCM speed for the vehicle.

The BN of Fig. 12 uses deterministic CPTs to express the mathematical operations of the algorithm. Database terrain values are accepted as evidence, and uncertainty is propagated through the network to the CCM node. The result reflects the impact of the uncertainty in the terrain data on the estimated CCM results. Clearly, many steps in the above algorithm themselves are themselves subject to uncertainty. A more sophisticated treatment would also include uncertainties in these computations.

The relationships between database and true values of terrain variables were assessed judgmentally by military terrain analysts. The analysts are accustomed to thinking of database values as inputs transformed by the model into cross country speeds. They felt comfortable assessing uncertainty by providing judgmental assessments on the uncertainty in this transformation, but did not feel comfortable providing assessments in the causal direction from true value to database value. The Bayesian network of Fig. 12 was created by reversing the arcs on the judgmental assessments provided by the analysts. This transformation was straightforward in this case because each database value of Fig. 12 has only one parent in the Bayesian network.

This example demonstrates that transforming a deterministic geospatial algorithm into a Bayesian network is straightforward, provided that the information needed to construct the CPDs is available and is captured as part of the metadata.

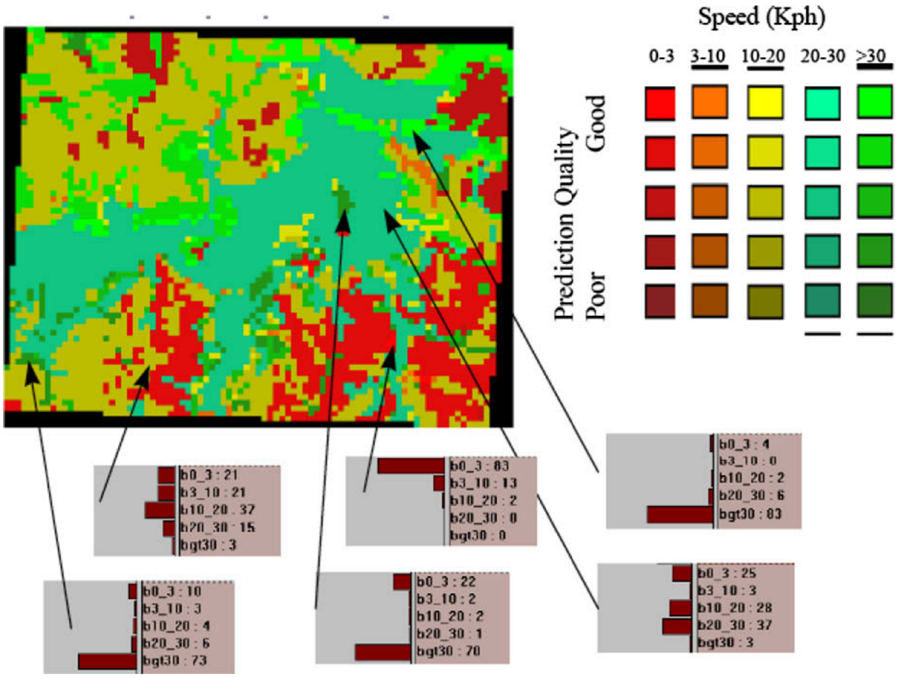

Fig. 13. CCM product with visualization of uncertainty in the CCM prediction.

Additional modeling is required when required inputs are not available. As discussed above, the presence of strong spatial correlation would require more sophisticated modeling and inference, together with additional data to estimate the correlations.

Fig. 13 shows a visual display of a CCM product with associated uncertainty. This display was created by applying the BN of Fig. 12 to each pixel. In the figure, CCM uncertainty is shown in two ways, through the display coloring and via interactive histograms that the user can control. The predicted CCM speed range is coded by color. The quality of the color represents the quality of the prediction: bright colors represent low uncertainty, and muddy colors represent high uncertainty. There is enough information in the legend that it is difficult to interpret the product colors. This intricacy is exacerbated by the difficulty of matching colors from computer monitor to printed hardcopy. To offset the difficulty in interpretation, user controlled popup histograms are provided on the digital display. Several examples are shown in Fig. 13. The popup histograms are useful to illustrate how the legend works:

- For each pixel in the display, the Bayesian network was applied to produce a probability distribution for predicted CCM speed.
- The pixel color (legend column) was selected that corresponds to 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.
- Note that the probability distribution of predicted speeds may be bimodal. This illustrates an important aspect of the visualization challenge. A simple display (e.g., best case, worst case) could be easily misinterpreted as representing the tails of a single mode (or even Normal) distribution.

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 Fig. 13. 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?" Fig. 14 shows an additional visualization that makes it possible to answer this query.

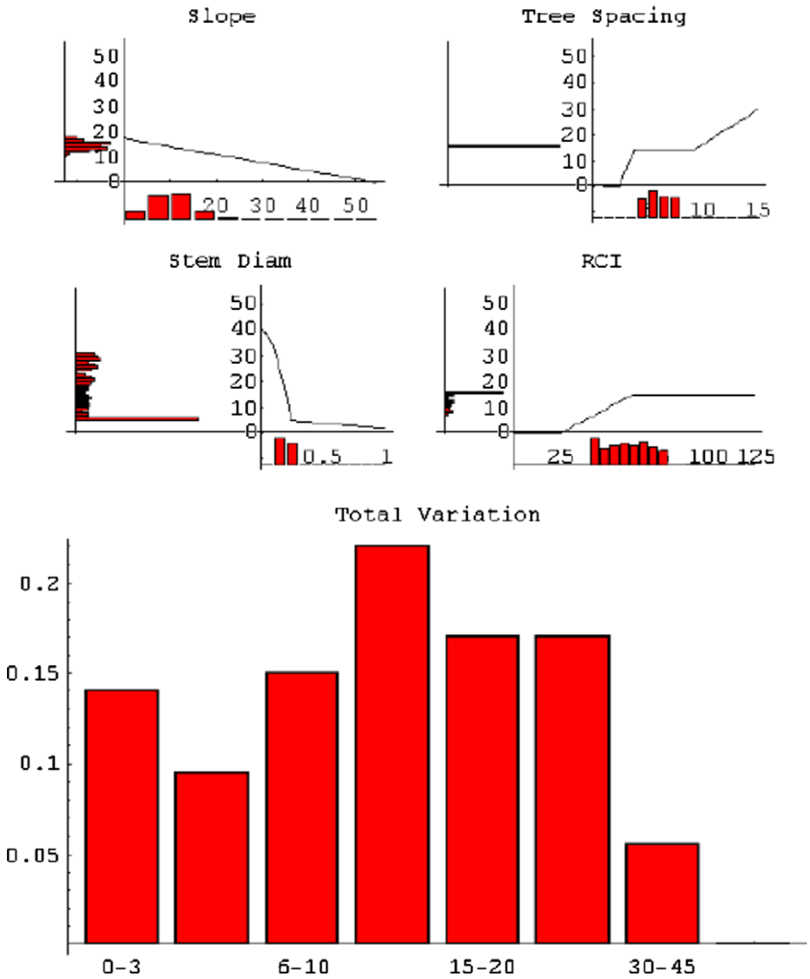

Fig. 14. Visualizing the influence of terrain factors on total uncertainity in CCM predictions.

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 associate variation in terrain inputs with 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. Each graphic also contains 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.

The above ideas regarding possible visualizations of uncertain, incomplete data uncover another vital issue for a successful geospatial system – the ability to meet the specific knowledge requirements of different types of user. The issue of concern is more complex than simple GUI customizations. The multitude and diversity of users relying upon a wide spectrum of possible features of a geospatial system suggests the need to customize features such as display formats for query responses, information granularity, and varieties of drill-down capability. 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.

An approach to addressing this 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, timeliness, expected granularity of information, most important factors for CCM predictions, etc. It could 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).

Geospatial reasoning systems are increasingly employed as web-based applications. In the military domain, the Department of Defense has mandated a new doctrine of network-centric operations. The objective of network-centric operations is to translate information superiority into a competitive military advantage through the use of well-informed, networked, geographically dispersed forces [1]. To achieve network-centric operations, information processing functions are packaged as services, which can be distributed over a network, and can be combined and reused to create more complex applications. For example, the function of creating a cross country mobility product might be packaged as a service and invoked across a network by a unit planning its next mission. Network-centric operations require interoperability among a diverse collection of services. Conformance to standard data exchange formats is insufficient – it is necessary for consumer and producer of a service to agree on the semantics of the information being exchanged. Ontologies have been promoted as a means to semantic interoperability. Costa et al. [5] advocate probabilistic ontologies as a means to semantic interoperability among multiple, distributed information sources, repositories, and users of a geospatial system.

Communicating potential uncertainties to decision makers enables them to make better decisions. A simulation experiment reported by Wright [23] demonstrates the importance of properly accounting for uncertainty in CCM calculations. Results of this experiment showed a dramatic improvement in a simple military planning scenario for agents that had access to the uncertainty information. For example, a traditional CCM product (without an uncertainty estimate) may show a potential maneuver corridor as "Go" terrain, supporting high speed cross country movement. Based on the prediction, the commander would feel justified in selecting the corridor as a critical avenue of approach for his plan. However, it is possible that the CCM product gives a false sense of certainty, and the corridor actually is not trafficable. As a result, the commander's plan may fail. If the original CCM product displays the uncertainty, the commander is able to make a better decision. For example, he might select an alternate avenue with little uncertainty, or he might allocate reconnaissance assets to collect additional terrain information and reduce the uncertainty.

6. Discussion and future work

It is important to represent, manage, and communicate to decision makers information about uncertainty in the GIS products used for military planning. Several prerequisites are required to achieve this goal. Methods must be available to estimate the quality of available geospatial data. 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 (e.g., [8,17]). In some cases, expert judgment is necessary to estimate data quality. 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 Fig. 12 was used to propagate uncertainty through the CCM model.

Different model views are appropriate for users playing different roles in the uncertainty management process. Model developers and implementers need access to the Bayesian network models of Figs. 8 and 12, 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 Figs. 10 and 12 are constructed to be similar to traditional map displays, while providing additional information about uncertainty. In addition to depicting uncertainty as part of the display, users can drill down to a more detailed explanation of particular uncertainties. Fig. 13 shows one kind of drill-down that decision makers might find useful. These model views must be managed by the GIS and the decision support tools built on the GIS.

Applying BN technology to geospatial reasoning in a principled way so to achieve the benefits we present here requires the ability to convey a complex information set that includes:

- Data for defining the local conditional distributions, which can be as simple as the example of Fig. 4 or as complex as first-order expressions.
- Structural information on the relationships among relevant parameters that can be as simply depicted as the BNs of Figs. 3, 5, 6, and 7, or in a much more elaborate scheme such as the first-order MEBN representation of Fig. 9, which can be used with a probabilistic reasoner to instantiate as many versions of the former as needed.

As we explained in Section 4 above, the appropriate representation for this information set is a geospatial probabilistic ontology, which provides a standardized, XML-based scheme for expressing geospatial metadata that can be directly processed by computers. In other words, probabilistic ontologies are sufficiently expressive to allow models to capture the semantics of the geospatial domain, error representations for each data source, and other probabilistic information. In short, BN technology in conjunction with probabilistic ontologies promises not only to improve the usefulness and reliability of geospatial models (i.e., by considering their implicit uncertainty), but also to enable a probabilistic reasoner to automate many complex inferential tasks that would otherwise require tedious, error-prone, resource-expensive processes.

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, this imposes costs on both production and use of geospatial data. In addition, implementation is hampered by the lack of appropriate industry standards for geospatial data quality, especially thematic data quality. Thus, an additional prerequisite is an organizational decision that the benefits of providing information about the uncertainty of GIS products exceed the costs.

A number of issues need to be addressed to address limitations in the methods described here. First, additional research is needed on usability of displays that incorporate uncertainty. Empirical research is needed on methods to communicate uncertainty in geospatial information to humans in a manner that supports decision-making. Second, additional research is needed to assess the true costs of ignoring uncertainty in typical kinds of problems encountered in applications. Several examples were given in this paper of how ignoring uncertainty can lead to suboptimal decisions. Extending these examples to a systematic study of the costs of ignoring uncertainty would provide empirical support for the need to treat uncertainty in geospatial decision support products. Third, additional research is needed on a number of modeling and computational issues. These include research on the impact of simplifying assumptions (e.g., ignoring spatial correlation), models and algorithms for relaxing simplifying assumptions made here (e.g., methods for addressing spatial correlation), research on the scalability of the algorithms, and research on architectures for distributed computation of geospatial products with associated uncertainty.

References

[1] D.S. Alberts, J.J. Garstka, F.P. Stein, Network Centric Warfare: Developing and Leveraging Information Superiority, CCRP Publications, second ed., 1999.
[2] D.P. Ames, Bayesian Decision Networks for Watershed Management. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Utah State University, 2002.
[3] P.C.G. Costa, Bayesian Semantics for the Semantic Web. Doctoral Dissertation, School of Information Technology and Engineering, George Mason University, Fairfax, VA, USA, 2005.
[4] P.C.G. Costa, M. Ladeira, R. Carvalho, K.B. Laskey, L. Santos, S. Matsumoto, A first-order Bayesian tool for probabilistic ontologies, in: Proceedings of the Florida AI Research Symposium (FLAIRS), 2008.
[5] P.C.G. Costa, K.B. Laskey, K.J. Laskey, E.J. Wright, Probabilistic Ontologies: the next step for net-centric operations, in: Proceedings of the 12th International Command and Control Research and Technology Symposium, 2007.
[6] N.A.C. Cressie, Statistics for Spatial Data, John Wiley and Sons, New York, 1993.
[7] DMS, Analysis of Natural Features, Defense Mapping School (DMS) Student Text, Ft. Belvoir, VA, Defense Mapping School, 1982.
[8] N. Friedman, Learning belief networks in the presence of missing values and hidden variables, in: 14th International Conference on Machine Learning (ICML-97), Morgan Kaufmann Publishers, San Mateo, CA, 1998.
[9] M.F. Goodchild, Closing Report NCGIA Research Initiative 1: Accuracy Of Spatial Databases l, National Center for Geographic Information and Analysis, University of California, 1992.
[10] A. Grêt-Regamey, D. Straub, Spatially explicit avalanche risk assessment linking Bayesian networks to a GIS, Natural Hazards and Earth System Science 6 (6) (2006) 911–926.
[11] M. Jianwen, D. Qin, Migratory locust hazard monitoring and prediction using the Bayesian network inference, in: Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), 2005, pp. 3623–3626.
[12] V. Kocabas, S. Dragicevic, Coupling Bayesian networks with GIS-based cellular automata for modeling land use change, in: Proceedings of the Fourth International Conference Geographic, Information Science (GIScience 2006), Münster, Germany, September 20–23, 2006.
[13] M. Kraak, F. Ormeling, Cartography Visualization of Spatial Data, Addison Wesley Longman Limited, Essex, England, 1996.
[14] K.B. Laskey, MEBN: A language for first-order bayesian knowledge bases, Artificial Intelligence 172 (2–3) (2008).
[15] K.B. Laskey, P. Costa, E.J. Wright, K.J. Laskey, Probabilistic ontology for net-centric fusion, in: Proceedings of the 10th International Conference on Information Fusion, July 2007.
[16] K.B. Laskey, E.J. Wright, P. Costa, Envisioning uncertainty in geospatial data, in: Proceedings of the Fifth UAI Bayesian Modeling Applications Workshop, 2007. Available from: <http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-268/>.
[17] K.B. Laskey, J. Myers, Population Markov Chain Monte Carlo, Machine Learning 50 (1–2) (2003).
[18] R.S. Lunetta, E.S. Congalton, Fenstermaker, J.R. Jensen, K.C. MCGwire, L.R. Tinney, Remote sensing and geographic information system data integration: error sources and research issues, PE&RS 57 (1991) 677–687.
[19] A.R. Pearson, J.S. Wright, Synthesis Guide for Cross-Country Movement, Engineer Topographic Laboratories Report No. ETL-0220, Ft. Belvoir VA, 1980.
[20] W.H. Ryder, D.E.Voyadgis, Measuring the Performance of Algorithms for Generating Ground Slope, USATEC Paper: Presented at DCAC TEM February 1996.
[21] A. Stassopoulou, M. Petrou, J. Kittler, Application of a Bayesian network in a GIS based decision making system, International Journal of Geographical Information Science 12 (1) (1998) 23–45.
[22] A.R. Walker, B. Pham, M. Moody, Spatial Bayesian learning algorithms for geographic information retrieval, in: Proceedings of the 13th Annual ACM International Workshop on Geographic Information Systems, Bremen, Germany, 2005.
[23] E. Wright, Application of Bayesian Networks for Representing Uncertainty in Geospatial Data, ASPS Spring Convention, Tampa, FL, 1998.
[24] E.J. Wright, Understanding and Managing Uncertainty in Geospatial Data for Tactical Decision Aids Doctoral Dissertation School of Computational Sciences, George Mason University, Fairfax, VA, USA, 2002.