GIS 93

GIS'93
Symposium

Vancouver
British Columbia
February 1993

A decision support system for predicting and consolidating ecosystems from existing map data and classification systems

Jan A. Mulder, Ph.D.

Department of Advanced Computing and Engineering
Alberta Research Council
3rd Floor, 6815-8 Street, NE
Calgary AB T2E 7H7

Ian G. W. Corns

Northern Forestry Centre
Forestry Canada
Edmonton AB T6H 3S5

Abstract

The Naia program is concerned with the design and implementation of an ecologically-oriented spatial and knowledge-based framework to support forest and land resource management. As part of this program a decision support system was designed and implemented with the capability of representing the knowledge used by a forest ecologist to infer a forest ecosystem from a variety of data sources. The system has been designed as a knowledge and information framework (i.e. shell) which operates in conjunction with a GIS and a relational database. The shell consists of a knowledge base, a predictive mapping tool, and a consolidation tool. The predictive mapping tool implements a combination of deterministic and probabilistic inference mechanisms which predict ecosystems from topography, forest, and soil maps. The consolidation tool has the capability of consolidating ecosystem maps based on a variety of decision criteria. The tools have been tested in three different areas of west-central Alberta. Preliminary test results indicate that with the use of ecological knowledge of the area reasonably accurate ecosystem predictions can be made from regular map data. An ecologically-oriented predictive mapping technology provides forestry organizations with a set of tools and an effective means upon which to base management planning. Once in place, this technology will lead to improved resource decisions with significant opportunities to save costs.

Introduction

Due to environmental concerns and socio-political factors we are witnessing an increased pressure on forest companies to manage their forests in a way t[illegible] addresses the human econo[illegible] needs and appreciates non-timber [illegible] as wildlife, habitat, and rec[illegible] However, the information [illegible] management a[illegible] [illegible] of traditior[illegible] inventories [6]. Such inventories tend to be limited to information about physiographic conditions, soil types and commercial forest species. Information about non-commercial aspects of the forest [illegible] ity composition [illegible] [illegible], such infor[illegible] survival of rare [illegible] animals with narrow [illegible] ments. Within the fra[illegible] sustainable developm[illegible] harvest planner should be capable of determining how much of a particular tree species can be harvested annually in a certain area without depleting the inventory over a longer period. If there is vegetation in the area that is vital for sustaining the local wildlife, then a harvest plan must be generated that guarantees adequate quantities of the sustaining vegetation.

Increasingly, forest management is being replaced by ecosystem management in order to accommodate multiple and often conflicting demands. Ecosystem classifications tend to be hierarchical. At the uppermost level they discriminate *ecoregions* or biogeoclimatic zones. These units identify vegetation and soil development patterns resulting from a similar macroclimate [13]. Ecoregions are mapped at scales from 1 : 250,000 to one to several million.

At the bottom of the hierarchy are site specific land units belonging to the same soil series [3] with the potential of supporting vegetation belonging to the same plant association. Depending on the school of site classification, these units are referred to as ecosystem types of the biogeoclimatic school of site classification [8] as applied in British Columbia, and as eco-elements by the Canada Committee on Ecological Land Classification [5].

In the present study, the unit of greatest resolution identified is the *ecosystem association* which is an abstract taxonomic unit above the ecosystem type, within an ecoregion. Ecosystem associations are defined as land areas with the potential of supporting similar plant communities with similar successional development, belonging to the same plant association. More than one soil family may be represented. The ecosystem association corresponds to the *ecosite* of the Canadian Committee on Ecological Land Classification [5] and is the fundamental site unit described in [1]. The ecosystem associations provide information on tree dominants, understory covers, soil properties, physiographic features, productivity and a number of forest management interpretations. Knowledge acquired from sample plots in the Corns and Annas database [1] was encoded into the Naia system as described in this paper.

As mentioned before, most forest inventories do not include detailed forest ecosystem information. Forest companies will therefore soon face two important issues:

1. how to obtain such information in a cost efficient way
2. how in the short term to use such information in order to improve management decisions.

Hughes Aircraft of Canada Limited, Spatial Data Systems Division and the Alberta Research Council have undertaken a joint research venture with the purpose of designing a decision support system (DSS) for predicting and consolidating [illegible] from data available from existing map sources. [illegible] common sense would [illegible] believe that an accurate ecosystem identification can be made on location only. [illegible] the Naia project has taught us that this is not necessarily the case. A key result of the Naia study has been that in many cases ecosystems can indeed be predicted from regular map data with a reasonable degree of accuracy. In some cases it is even possible to exactly predict the ecosystem. Once such predictions have been made for a particular geographic area the information created can be spatially consolidated by means of a variety of decision criteria. For example, harvest planners may want to create an ecosystem map that aggregates the ecosystem types by their optimal season of harvest, or silviculturists may wish to map polygons with high vegetation competition hazard.

In order to design the Naia DSS new theoretical ground had to be broken. The forest ecologist's knowledge of ecosystems and their occurrence is in part deterministic (also referred to as symbolic) and in part probabilistic in nature. A great deal of work has been done in the field of Artificial Intelligence on symbolic and probabilistic constraint satisfaction [4,7,9,10,11]. However, the question of how to combine these two forms of constraint satisfaction has not been studied in depth. Probabilistic constraint satisfaction can be subdivided into attribute uncertainty and spatial uncertainty management. Spatial uncertainty addresses the question of the existence and location of spatial map features - point line polygons. Attribute uncertainty addresses the interpretation of such features / polygons. This paper deals with the deterministic and attribute side of uncertainty management only. Spatial uncertainty and its computation is addressed in the companion paper by Crain *et al.* [2].

The Naia DSS has been designed as a shell. A description of the ecoregions, ecosystem associations and ecosystems of a particular geographic area can be characterized within the shell together with a number of attributes and a range of possible values. The attribute values constrain the ecoregions and ecosystem associations possible in both a symbolic and probabilistic manner. In the next section we describe the shell's components. Section 3 discusses the implementation of the system, whereas section 4 illustrates the preliminary results of experimental tests. In section 5 we discuss some potential applications for the shell.

The Naia decision support shell

The shell consists of 3 components: a *knowledge base*, a *classification tool*, and a *consolidation tool.*

The knowledge base

The knowledge base contains object-centered descriptions of three types of objects: *polygons*, *attributes*, and *ecosystems*. The shell has been designed to operate in conjunction with a GIS and relational database. From the GIS the shell abstracts a set of polygons that constitute a map or the result of a map overlay. The attributes and values associated with each polygon are abstracted from the relational database. Ecosystems need to be defined inside the shell and are the result of a knowledge elicitation process with a domain specialist. As was already discussed in the introduction ecosystems have a hierarchical organization and are embedded in a *specialization* hierarchy. Such a hierarchy represents the object at different levels of refinement. Fig. 1 shows the appearance of such a hierarchy for ecosystems. Ecosystems are essentially described at three levels of refinement, an *ecoregion* level, an *ecosystem association* level, and an *ecosystem* level. The level of specialization used is determined by the amount of detail required for the planned management activity and by the availability of data. Such considerations determine the appropriate scale for mapping a site classification. The purpose of the DSS is to provide a meaningful representation format for ecosystem classifications, while taking into account data availability and reliability. The relationships (and constraints) between ecosystems and attributes are expressed in the form of *belief functions* which express the degree of belief of a (human) expert in the occurrence of an ecoregion, association, or ecosystem, given a particular value for an attribute. Belief functions represent the domain specialist's accumulated experience and enhanced confidence gained by making classifications in the field based on the (attribute) data provided.

The classification tool

Each polygon defined in the knowledge base is expected to have a series of attributes each with one or more values. The objective of the classification tool is to let these attributes act as constraints on the ecosystems possible and to compute a list of possible ecoregions and ecosystem associations ordered by a belief measure. The classification process forms a combination of symbolic and probabilistic reasoning and it is guided by two different principles: the principle of *least commitment* and the principle of *graceful degradation*. The first principle requires that at any time the attribute data are represented by an ecosystem at a specialization level that is appropriate for those data. The second principle implies that the classification process reflects the reliability and availability of attributes. More in particular, we do not want the classification process to break down completely, if one or more attributes are missing in a polygon, although we do allow the level of confidence in the results to decrease.

The principle of least commitment is enforced by specifically allowing certain attributes to constrain ecosystems at one particular level of specialization only (see Fig. 1). As an example, "elevation" as an attribute cannot by itself very accurately distinguish between ecosystem associations, but does have that power for separating ecoregions. Graceful degradation is reflected in two ways. First, the system always generates a classification, no matter how many or few attributes are available. Second, the system provides a confidence factor for the classification of each polygon which reflects both the quantity and reliability of the attributes available.

In analogy with many perceptual systems the classification process can be looked at conceptually as a cycle of processes. In the Computer Vision literature this cycle is often referred to as "the *cycle of perception*". Fig. 2 illustrates this cycle. It consists of four processes: *cue discovery, model invocation, model testing,* and *model elaboration*. In the Naia system cue discovery is essentially the process of gathering together all the necessary information such as the maps and attributes. It can therefore also be looked at

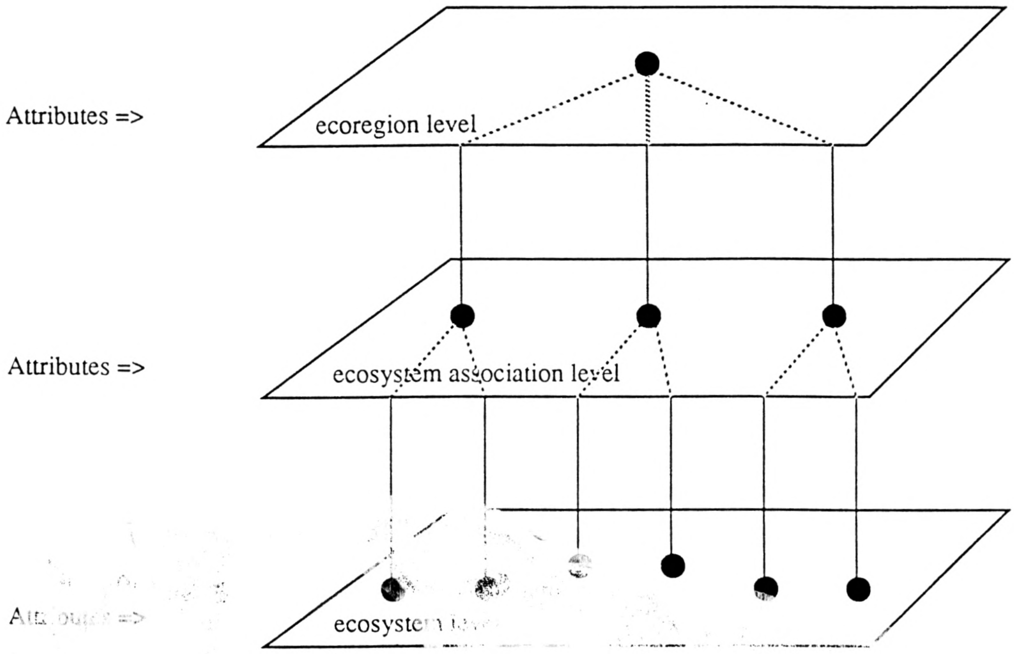

Figure 1: *The ecosystem specialization hierarchy*

as a *segmentation* process with the capability of creating map overlays. This task is performed by the GIS (See Skye's companion paper for a description of this process [12]).

The GIS polygon overlay process provides resultant polygons with attributes attached to each. These polygons constitute the finest resolution the system is capable of producing based on the spatial information available. Classification can also be looked at as a process of model refinement. Attributes serve as cues for the classification process. Model invocation is the start button that sets the classification process in motion. Based on the attributes found during cue discovery it creates a single (abstract) classification model for each polygon. Model testing is the process that computes the influence of each input attribute on this classification model. As a result the original model will be replaced by one or more specialized models. Model elaboration is the process that adapts the classification models of each polygon based on the classification models of their immediate neighbors. Model elaboration is the task performed by the *consolidation* tool. Consolidation often leads to a merging of neighboring polygons which effectively constitutes a *resegmentation* of the input maps, thus closing the cycle. The cycle is usually entered at cue discovery.

The consolidation tool

Consolidation represents the model elaboration and resegmentation stage of the cycle. The consolidation tool can compare interpretations of adjacent polygons and can line up adjacent polygons as merging candidates based on some merging criterion defined by the user. The system provides the capability of doing merging based on a variety of deterministic and probabilistic criteria. These include: merging based on the same classification models (or subset of models) occurring in two adjacent polygons, merging based on the same most likely classification models, and merging based on model management interpretation criteria. The consolidation tool passes on the appropriate information to the GIS which actually performs the resegmentation.

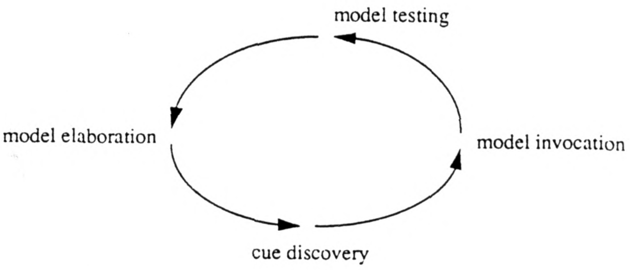

Figure 2

Implementation

The system has been implemented and tested in three different areas of west-central Alberta; one area near Grande Prairie managed by Canadian Forest Products, and two areas near Hinton managed by Weldwood, Hinton division. The models used formed a subset of ecosystems as defined in the "Field Guide to Forest Ecosystems of West-Central Alberta" [1]. This led to the creation of the hierarchical knowledge base illustrated in Fig. 3. Each ecosystem occurs within one of three ecoregions: Lower Boreal Cordilleran (LBC), Upper Boreal Cordilleran (UBC), and Subalpine (SA).

For each of the test areas topographic base, soil, and forest cover maps were obtained and overlaid by means of a GIS. All available attributes of the overlaid maps were

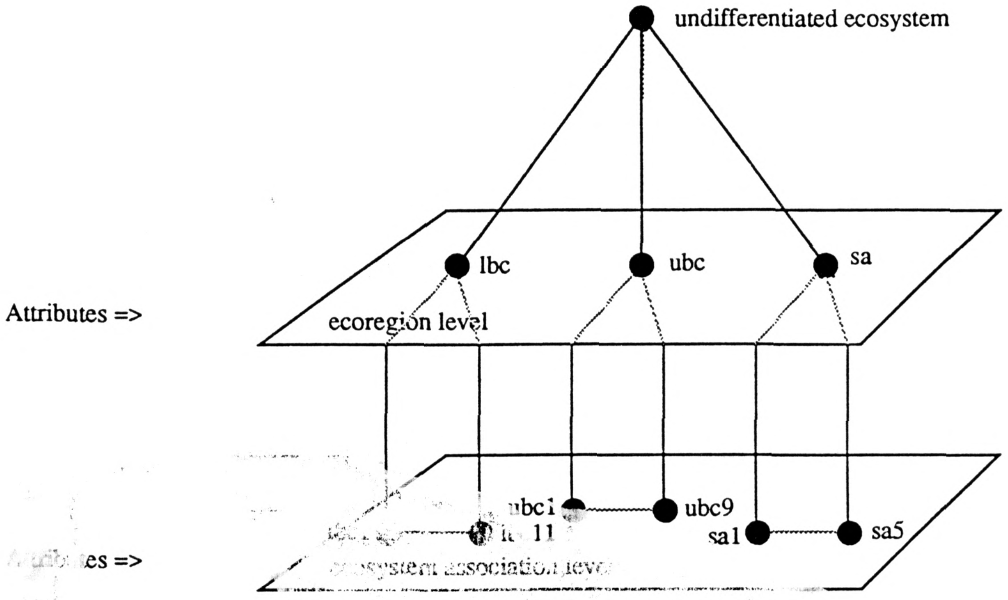

Figure 3 *The west-central Alberta knowledge base*

stored in a relational database. A characteristic subset of attributes was selected for use in the DSS. These include attributes such as elevation, aspect (from base map), tree species composition and abundance (from forest cover map), and texture, drainage (from soil map) (see also Skye's paper [12]). For the prototype we used ArcInfo™ as a GIS, and Oracle™ as a relational database. The DSS was implemented in Common Lisp Object System (CLOS).

The classification process is analogous to what forest ecologists do on the ground using conventional identification keys in field guides. Each polygon is at first interpreted as an undifferentiated ecosystem (which we know for certain). No attributes have been processed at this time. As a result our confidence in the current interpretation is zero. As we introduce the attributes one by one the ecosystem classification model is forced down the hierarchy (see Fig. 3). The elevation attribute, for example, may constrain the undifferentiated ecosystem to specialize in to *lbc* and *ubc*, each with an associated belief which expresses the system's relative confidence in each classification. Another attribute (e.g. aspect) may allow only some of *lbc*'s and *ubc*'s refinements and may force the classification down into the ecosystem association level (e.g *lbc11*and *ubc5*). The classification process thus takes the form of a constraint based process that propagates constraints in both a vertical direction (in between levels of specialization) and in a horizontal direction (at one specific level of specialization). The constraints are both symbolic (i.e each attribute value allows a specific subset of ecosystems only) and probabilistic (i.e each ecosystem comes with an associated belief that requires continuous updating as new attributes are brought in).

The results of the classification process are stored again in the relational database. A GIS with access to this database can then be used to display the results as a map of predicted ecosystem associations [12]. For example, the user can define a legend for each ecosystem association and display for each polygon the ecosystem with the strongest belief. For each polygon in the map the user can also bring up a window that displays all classification information associated with the polygon. An example of information contained in such a window is illustrated in Fig. 4. The window displays information about the polygons's attributes, classifications, and confidence. The classification results are given on a level by level basis (see Fig. 3). The confidence is expressed by a value between 0 and 1. It reflects the number of attributes that was processed relative to the total number of potentially available attributes.

```
Name: polygon 16
elevation: 950
aspect: Flat
texture: Organic
Drainage: Very poor

Classifications:

ecoregion level: [lbc 0.6][ubc 0.4]
ecosystem association level: [lbc11 0.6][ubc5 0.4]

confidence: 1.0
```

Figure 4: *Example output window*

The user is also given the possibility of looking at the classification "trace" for each polygon. That is, the system keeps a record of the complete classification process on an attribute by attribute basis. Thus, the user can see exactly what influence each attribute had on the classification. This facility provides an easy means of tracking down the [illegible] undesirable results. If, for example, the forest ecologist decides that given the current combination of [illegible] attribute values ecosystem association *ubc5* should have come out as the most desirable, then a close inspection of the trace will reveal immediately which attribute(s) caused *ubc5* to loose its strength. The knowledge base may need adjustment or it may raise questions about the quality of the source map information.

Validation

As the validation process was still in progress at the time of writing of this paper we can only report some preliminary (albeit encouraging) results. The knowledge base and classification tool have been validated in two different ways. In one of the test areas, for which an ecosystem classification was determined by a ground crew, a sample of 52 points was selected. The attribute data for each point were entered into the system after which the system generated a classification list. Several probabilistic update formulas were experimented with. The most powerful results were generated by means of a weighted averaging formula and the Dempster-Shafer formula for evidential reasoning [11]. Fig. 5 shows the results for both formulas. The results are reported in three different categories: the percentage of cases in which the most likely ecosystem was also the human expert's choice (1), the percentage of cases in which the expert's choice was among the first two most likely choices of the system (2+), and the percentage of cases in which the expert's choice was among the first three most likely choices of the system (3+).

A knowledge base which is partially based on fuzzy logic operations can be subject to what one may call "pathological suppression". An ecosystem association is pathologically suppressed if there is no combination of attribute values that will force that association to come out as the most likely classification. A second form of validation testing was therefore done to ensure that the knowledge base did not suffer from this phenomenon. The knowledge base passed the test.

Another encouraging observation has been that the predictions made by the system for the three test areas have been quite consistent over large areas. As a result of the overlaying process the number of polygons for each area went up as high as 10,000. However, consolidation

	Update formula	
category	weighted averaging	dempster/shafer
1	63%	58%
2+	79%	79%
3+	88%	94%

Figure 5: *Test point results.*

of polygons based on the most likely classification shows a very large reduction (by as much as 90% in one area) in the number of polygons.

Discussion and potential application

Many forest companies do not yet have a land inventory that includes the spatial distribution of forest ecosystems. Without further aids, the classification of local ecosystem associations will require a time consuming and costly ground truthing effort. It is here that the Naia DSS can become a worthwhile investment. The Naia DSS uses the map data already available in the current land inventories to predict the local ecosystem associations.

Depending on the attributes available and their values the Naia DSS will sometimes uniquely predict and map an ecosystem association whereas in other cases the result will be a series of classification models each with an associated belief. Forest ecologists can use the system to locate areas of homogeneity and disparity. In homogeneous areas a few ground sample points may suffice to verify the system's prediction, whereas in areas of disparity more intensive ground truthing may be necessary. The ecologist can also take a careful look at the belief values of each ecosystem. Less ground verification may be necessary in areas where one ecosystem stands distinctly above its competitors. Finally, the system may sometimes fail to come up with any interpretation at all. This by itself is useful information. It indicates that either there is a gap or incorrectness in the ecological models for the area, or it may indicate a problem with some of the attribute source data (i.e. the maps). Overall the Naia DSS can have the effect of reducing the cost of building a forest ecosystem inventory.

The Naia DSS also has applications for forest management. Because forest ecosystem classification systems typically include management interpretation tables. These interpretations can be used for the predictive mapping and consolidation process. Management interpretation tables include such things as considerations of season of harvest, logging method, site preparation intensity, soil compaction and puddling hazard, reforestation methods and many others. Any of these factors can potentially be used as a consolidation measure in an ecosystem map. The DSS thus has the capability of supporting more efficient planning and resource decisions.

As mentioned before, the Naia system is a DSS based on ecosystem classification knowledge. Two decision support tools have been built which assist in ecosystem prediction and consolidation. These tools can address some of the short term needs of forest companies intent on developing forest management strategies based on sustainable development and integrated resource management. In the longer term a variety of other tools will be needed. These include: predicting changes in ecosystems over time (e.g. ecological succession), predicting volumes based on different harvesting and reforestation procedures, suggesting potential harvest areas based on a variety of constraints such as minimally required quantities of certain forest vegetation, and many more. Future phases of Naia will be targeting such issues (see also Jones's companion paper [6]).

Acknowledgments

The research described in this paper has been supported in part by a grant from the Natural Sciences and Engineering Council (NSERC) of Canada (OGP0121247) to the first author. The Naia DSS was developed under a Joint Research Agreement between Hughes Aircraft of Canada Limited, Spatial Data Systems Division and the Alberta Research Council, with support from the University of Calgary, Department of Geomatics Engineering. The authors are grateful to Don Kuzenko for assisting in the implementation of the DSS and to Keith Jones for comments on a draft version of this paper. Doug Allan of Forestry Canada (Northern Forestry Centre), Canadian Forest Products, and Weldwood of Canada are gratefully acknowledged for providing the expertise and data necessary to validate the system. ARCINFO is a GIS marketed by ESRI. ORACLE is a relational database marketed [illegible] corporation.

References

[01] I.G.W. Corns, R.M. Annas, Field Guide to Forest Ecosystems of West-Central Alberta, Northern Forestry Centre, Canadian Forest Service, Edmonton, 1986.

[02] I.K. Crain, P. Gong, M.A. Chapman, S. Lam, J. Alai, M. Hoogstraat, Implementation Considerations for Uncertainty Management in an Ecologically-Oriented GIS, *Proceedings of GIS 93, Vancouver, Canada.*

[03] Expert Committee on Soil Survey, The Canadian System of Soil Classification, 2nd edition, Publication 1646, Agriculture Canada, Research Branch, Ottawa Ontario, 1987.

[04] J.Y. Halpern, R. Fagin, Two Views of Belief: Belief as Generalized Probability and Belief as Evidence, *Artificial Intelligence* **54**, (3), 275 - 317, 1992.

[05] G.R. Ironside, The Canada Committee on Ecological Land Classification (CCELC): History, Objectives and Activities in Land / Wildlife Integration, Ecological Land Classification Series no 11, Lands Directorate, Environment Canada, 1980.

[06] R.K. Jones, Next Generation Forest Site Classification: Ecologically - Oriented Predictive Mapping Technology, *Proceedings of GIS 93, Vancouver, Canada.*

[07] A.K. Mackworth, J.A. Mulder, W.S. Havens, Hierarchical Arc Consistency: Exploiting Structured Domains in Constraint Satisfaction Problems, *Computational Intelligence* **1**, no 3 - 4, 118 - 126, 1985.

[08] W.R. Mitchell, R.E. Green, Identification and Interpretation of Ecosystems of the Western Kamloops Forest Region, Land Management Handbook no 2, Province of British Columbia, Ministry of Forests, 1981.

[09] J.A. Mulder, Discrimination Vision, *Computer Vision, Graphics,* and *Image Processing* **43**, 313 - 336, 1988.

[10] J. Pearl, On Evidential Reasoning in a Hierarchy of Hypotheses, *Artificial Intelligence* **28**, 9 - 15, 1986.

[11] G. Shafer, R. Logan, Implementing Dempster's Rule for Hierarchical Evidence, *Artificial Intelligence* **33**, 271 - 298, 1987.

[12] D. Skye, Ecologically-Oriented Spatial and Knowledge-based Framework to support Forest and Land Resource Management, *Proceedings of GIS 93, Vancouver, Canada.*

[13] W.L. Strong, K.R. Leggat, Ecoregions of Alberta, Alberta Forestry, Lands, and Wildlife, Land Information Services Division, Edmonton AB, 1992.

The world's premier GIS Symposium on

forestry
natural resources
the environment

February 15 – 18, 1993
Vancouver, British Columbia

Symposium Proceedings • Volume 1

GIS
93

Volume 1

Volume 2

7th Annual Symposium on Geographic Information Systems in forestry, environment and natural resources management

February 15 - 18, 1993

Vancouver Trade and Convention Centre
Pan Pacific Hotel
Vancouver, British Columbia, Canada

This publication contains the texts of papers presented at GIS'93 that were received prior to February 1, 1993. These Proceedings are published in two volumes:

Volume 1
International • Forestry • Natural resource management • Environmental • Wildlife

Volume 2
Global monitoring • Corporate GIS • Developments and trends • Technologies, methodologies and techniques • LIS

GIS'93 is organized by Polaris Conferences and is sponsored in part by Forestry Canada and the British Columbia Ministry of Forests, with funding provided under the Canada-British Columbia Partnership Agreement on Forest Resource Development, FRDA II; and by the British Columbia Ministry of Environment, Lands and Parks.

Forestry Canada Forêt Canada

Province of British Columbia
Ministry of Forests
Ministry of Environment, Lands and Parks

Symposium man[illegible]

polaris [illegible]

305-1040 Ha[illegible] Canada, [illegible]

ISSN 0835-0752

ISBN 0-662-20392-5

Ministry of Supply and Services Canada, 1993
Cat. No. FO 18-16/1993E

Editing, copyfitting and desktop publishing by
Polaris Conferences, Vancouver, BC

Printing by Gastown Printers, Vancouver, BC