

APPLICATION OF BAYESIAN NETWORKS TO  
LARGE-SCALE PREDICTIVE ECOSYSTEM MAPPING

by

Adrian Walton

BES, University of Waterloo, 1995

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF NATURAL RESOURCES AND ENVIRONMENTAL STUDIES

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

March 2005

© Adrian Walton, 2005

**UNIVERSITY of NORTHERN  
BRITISH COLUMBIA  
LIBRARY  
Prince George, B.C.**

## ABSTRACT

Large-scale ecosystem maps are essential tools for managers of forest-related activities. In British Columbia, the prevailing approach for ecosystem mapping has been to use an expert system that captures expert knowledge in the form of a belief matrix. In this project, I replaced the belief matrix with a Bayesian network in an attempt to overcome some of the drawbacks of the belief-matrix approach. I created a Bayesian-network knowledge base applied to an area encompassing Prince Rupert and part of the following three biogeoclimatic units: Coastal Western Hemlock Very Wet Maritime Montane, Coastal Western Hemlock Very Wet Maritime Submontane and, Coastal Western Hemlock Very Wet Hypermaritime Central. Using each knowledge base, I produced a map of grouped site series. Accuracy assessments performed on each of the maps of grouped site series revealed that the maps poorly predicted the spatial distribution of rare and very wet site-series groups. The results of the map-accuracy assessment, however, were consistent with those resulting from a belief-matrix approach. I considered the Bayesian-network knowledge bases easier to develop, interpret and update than belief matrices.

## TABLE OF CONTENTS

ABSTRACT.....	ii
LIST OF FIGURES .....	v
LIST OF TABLES.....	vii
ACKNOWLEDGEMENTS.....	viii
1 INTRODUCTION .....	1
1.1 Statistical Methods.....	4
1.2 Geostatistical Methods.....	5
1.3 Artificial Neural Network Methods.....	6
1.4 Decision Tree Analysis .....	8
1.5 Expert Systems.....	9
1.6 Hard Versus Soft Classification.....	15
1.7 Accuracy Assessment .....	18
1.8 Statement of the Problem and Objectives.....	21
2 METHODS .....	24
2.1 Study Area .....	24
2.2 Methods Overview.....	27
2.3 GIS Database Development.....	29
2.4 Knowledge Base Construction.....	43
2.5 Final Map Generation .....	64
2.6 Map Accuracy Assessment.....	64
3 RESULTS .....	66
3.1 Knowledge Base Structure.....	66
3.2 Predictive Ecosystem Maps .....	70

3.3	Map Accuracy Assessment .....	74
4	DISCUSSION .....	80
5	CONCLUSION AND RECOMMENDATIONS.....	96
6	REFERENCES .....	99

## LIST OF FIGURES

Figure 1 – A graphical representation of the semantic-import model used to define the relationship between slope gradient and the semantic descriptions of flat and sloping.	17
Figure 2 – The study area which was used to conduct large-scale predictive ecosystem mapping.....	25
Figure 3 – This diagram illustrates the methodology used to meet the objectives of this study. There were three general sections in the methodology that also comprises most expert systems: create the database, develop the knowledge base and perform inference.....	28
Figure 4 – This sample map (a) demonstrates the difference between the upper-landform elements as defined by Jones (2003) (b) and the upper-landform elements as defined by the continuous-landform-elements layer (c). .....	34
Figure 5 – The Bayesian network of continuous landform elements, which was developed in Netica version 2.17, was developed to predict the spatial arrangement of landform elements. ....	37
Figure 6 – Four examples of the GIS data layers used to create the database used in this study. This diagram highlights the different spatial scales that occurred within the GIS database.....	42
Figure 7 – The influence diagram created to explain the relationship between the GIS database used in this study and soil-moisture regime. ....	48
Figure 8 – The influence diagram created to explain the relationship between the GIS database used in this study and soil-nutrient regime.....	49
Figure 9 – The Bayesian network used to define the relationships between the GIS data layers used in this study, and soil-moisture and soil-nutrient regimes within the CWHvm2 BEC unit. This network was created by combining the soil-moisture regime and soil-nutrient regime influence diagrams in Figure 7 and Figure 8.....	50
Figure 10 – Edatopic grids and corresponding site-series groups used for the CWHvm1, CWhvm2 and CWHvh2 Bayesian networks.....	55
Figure 11 – The portion of the CWHvm2 Bayesian-network, which illustrates the relationship between site-series group ( <b>Site Series Edatopic</b> ), soil-moisture regime ( <b>Soil Moisture</b> ) and soil-nutrient regime ( <b>Soil Nutrient</b> ).....	56
Figure 12 – The portion of the Bayesian network for the CWHvm2, which illustrates the influence of slope ( <b>Slope Group2</b> ) on the previous estimation of site-series-group ( <b>Site Series Edatopic</b> ) to produce a new prediction of site-series group ( <b>Site Series Slope Modifier</b> ).....	57

Figure 13 – The portion of the Bayesian network for the CWHvm2, which illustrates the influence of forest-stand productivity ( <b>Site Productivity</b> ) and forest-stand species ( <b>Species Influence</b> ) on the previous estimation of site-series group ( <b>Site Series Slope Modifier</b> ) to produce a final estimation of site-series group ( <b>Site Series Veg Modifier</b> ) for the CWHvm2.....	57
Figure 14 – The portion of the Bayesian network for the CWHvm1, which illustrates the influence of floodplains ( <b>Floodplain</b> ) on the previous estimate of site-series group ( <b>Site Series Veg Modifier</b> ) to produce a final estimate of site-series group ( <b>Site Series Fluvial Modifier</b> ) for the CWHvm1. ....	59
Figure 15 – The Bayesian network for the CWHvh2, which illustrates the influence of ocean wind ( <b>Salt Spray Shoreline</b> ) and ocean estuaries ( <b>Gully Adjacent to Shore</b> ) on the previous estimation of site-series group ( <b>Site Series Fluvial Modifier</b> ) to produce a final estimation of site-series group ( <b>Site Series Shore Modifier</b> ) for the CWHvh2....	60
Figure 16 – The complete Bayesian network developed to predict site-series group in the CWHvm2 BEC unit of this study. ....	67
Figure 17 – The complete Bayesian network developed to predict site-series group in the CWHvm1 BEC unit of this study. ....	68
Figure 18 – The complete Bayesian network developed to predict site-series group in the CWHvh2 BEC unit of this study.....	69
Figure 19 – An overview map of predicted site-series groups in all three BEC subzone/variants in the study area. This map shows the relative position and abundance of each of the predicted site-series groups within each of the three BEC subzone/variants.....	71
Figure 20 – A perspective view illustrating the relative terrain position of the final site-series-group predictions for the CWHvm2 and CWHvm1 BEC units. The wet site-series groups, such as 05, appear to be correctly placed within the gullies and toe slopes, and the dry site-series groups, such as 03, appear to be correctly placed on ridges and shoulders. ....	72
Figure 21 – A perspective view illustrating the relative terrain position of the final site-series-group predictions for the CWHvh2 BEC unit. The wet site-series groups, such as 07 and 14, appear to be correctly placed in the gullies and toe slopes, and the dry site-series groups, such as 03, appear to be correctly placed on ridges and shoulders. ....	73

## LIST OF TABLES

Table 1 – A summary of the Jones (2003) rule-based logic used to create the TIN/DEM dependant GIS layers used in this project. ....	31
Table 2 – The GIS layers of bedrock, stream density, and ocean wind, which were derived using the procedures of Jones (2003), were not derived solely from the TIN or raster DEM, but rather depended on other data sources. ....	32
Table 3 – The conditional-probability table used to define the relationship between landform elements, terrain context and form elements. ....	39
Table 4 – The heuristic rule-base used to convert input values into fuzzy landform attributes. ....	40
Table 5 – A list of the data layers contained within the GIS database, along with comments regarding each layer’s ability to assist site series prediction. ....	41
Table 6 – Site-series groups used for the CWHvm2 BEC unit. CWHvm2 site-series units of Meidinger and Pojar (1991) were grouped into simpler units due to the low thematic and spatial accuracy of some GIS layers. ....	44
Table 7 – Site-series groups used for the CWHvm1 BEC unit. CWHvm1 site-series units of Meidinger and Pojar (1991) were grouped into simpler units due to the low thematic and spatial accuracy of some GIS layers. ....	45
Table 8 – Site-series groups used for the CWHvh2 BEC unit. CWHvh2 site-series units of Meidinger and Pojar (1991) were grouped into simpler units due to the low thematic and spatial accuracy of some GIS layers. ....	46
Table 9 – The accuracy-assessment contingency-matrix for the CWHvm2, which includes percentage correct, user’s and producer’s accuracy.....	75
Table 10 – The matrix of average posterior probabilities for the CWHvm2, which includes the average posterior probabilities for each mapped site-series group. ....	75
Table 11 – The accuracy-assessment contingency-matrix for the CWHvm1, which includes percentage correct, user’s and producer’s accuracy.....	76
Table 12 – The matrix of average posterior probabilities for the CWHvm1, which includes the average posterior probabilities for each mapped site-series group. ....	76
Table 13 – The accuracy-assessment contingency-matrix for the CWHvh2, which includes percentage correct, user’s and producer’s accuracy.....	78
Table 14 – The matrix of average posterior probabilities for the CWHvh2, which includes the average posterior probabilities for each mapped site-series group. ....	79

## ACKNOWLEDGEMENTS

It is a pleasure to thank the many people that made this degree possible over the past six years. Foremost, I wish to state my gratitude to my MNRES supervisor, Roger Wheate. His patience allowed me to explore many possible research topics and degrees within the College of Science and Management. I would also like to express my gratitude to Michael Gillingham for being a part of my supervising committee and providing thorough reviews.

I thank my co-workers for providing me with useful advice during the completion of my final project. The greatest contribution came from Del Meidinger. He not only agreed to work closely with me to create the ecological models but also agreed to be part of my supervising committee. This project would not have been possible without his advice and support. Marvin Eng, my work supervisor, very graciously allowed me to focus my past year on the completion of my final project. I would not have completed this project without this opportunity. I also thank Peter Ott for providing me with statistical advice and Shirley Mah for helping me with my writing —both valuable contributions.

I wish to thank my siblings and extended family. My sisters, Heidi Robinson and Liesl Walton, and my brother-in-law, Michael Robinson, were particularly supportive and patient during the past six years.

Also very important to me was the direct support I received at home from my partner, Jodi Mucha. She graciously accepted the impact that combined academic and work life had on our life at home and provided support during times of frustration.

Lastly, and most importantly, I thank my parents, Christine Walton and Colin Walton. They raised me and supported me, which allowed me to be who I am today. To them I dedicate this degree.

# 1 INTRODUCTION

It is essential for land managers to have information regarding the nature and distribution of ecosystems to manage forest-dependant activities. In order to facilitate this need, a working definition of ecosystems is required, including inventories of the components that comprise ecosystems, and an understanding of the spatial pattern of ecosystems, their formation and interrelationship.

Hierarchical systems that classify land based on ecological principles have been developed for scales ranging from global to local. At the global scale, also referred to as small-scale, several researchers have developed ecological land classifications using a bioclimatic approach: e.g., Udvardy (1975), Walter and Box (1976), and Bailey (1989a; 1989b). In the United States, Driscoll (1984), Omernik (1987), and Cleland *et al.* (1997) developed ecologically-based classifications at the regional scale. In New Zealand, Huser (1999) proposed a hierarchical ecosystem-classification scheme ranging from the regional scale to the local scale, also referred to as large scale. In Canada, nation-wide ecosystem classifications have been created by Wiken (1986), the Ecological Stratification Working Group (1996), and Marshall and Schuts (1999). Large-scale ecological-classification schemes have been presented in the United States (Barnes *et al.* 1982), New Brunswick (Zelazny *et al.* 1989), Ontario (Jones 1983; Sims *et al.* 1989; Chambers *et al.* 1997), Manitoba (Zolaneski *et al.* 1995) Saskatchewan (Beckingham *et al.* 1996b) and Alberta (Beckingham & Archibald 1996; Klappstein & Corns 1996; Beckingham *et al.* 1996a). In the province of British Columbia, Pojar *et al.* (1987) formalized the Biogeoclimatic Ecosystem Classification (BEC) approach to ecosystem classification.

The BEC system is an integrated hierarchical classification scheme that combines climate, vegetation and site classifications. Ecosystems ranging in scale from regional to local are defined using zonal (climatic) concepts and the vegetation of zonal climax or near-climax ecosystems. In the BEC scheme, the basic unit of zonal or climatic classification is the subzone. Grouping subzones forms zones, while subzone divisions define variants (Pojar *et al.* 1987). Site series, which are the largest-scale ecosystems unit of the BEC system, are defined within variants, or subzones if variants are not described for a subzone. Site-series units describe all land areas that are capable of supporting a specific climax vegetation within a climatic area (subzone or variant) (Pojar *et al.* 1987).

The ecosystem-forming model of Pojar *et al.* (1987) is based on the vegetation and soil-formation theory of Major (1951): vegetation and soils are products of climate, organisms, topography, parent material, and time. Plants and soil, considered simultaneously, integrate all the components of an ecosystem and reflect ecosystem formation. As a result, an ecological unit can be described by the plant community (a volume of relatively uniform vegetation) and the soil polypedon (a volume, to the depth of the solum, of relatively uniform soil) upon which the plant community occurs (Pojar *et al.* 1991).

Delineating the geographical distribution of the ecological units that belong to an ecological classification scheme is broadly referred to as ecosystem, or ecological, mapping. The approaches used to delineate ecological units on the landscape are very similar to those employed in vegetation (Skidmore 1989; Kimes *et al.* 1996; Carpenter *et al.* 1997; Carpenter *et al.* 1999; Miller & Franklin 2002; Linderman *et al.* 2004) and soils mapping (McCracken & Cate 1986; Burrough 1989; Moore *et al.* 1993; Skidmore *et al.* 1996; MacMillan *et al.*

2000; Zhu *et al.* 2001; Qi & Zhu 2003). This is because all three domains rely on similar relationships between topography, climate, parent material, organisms and time (Jenny 1941; Major 1951; Pojar *et al.* 1987). In addition, large-scale mapping techniques employed in each domain have evolved similarly over time (Scull *et al.* 2003).

Traditionally, large-scale ecosystem-, soil- and vegetation-mapping were derived through a site-survey approach (Scull *et al.* 2003). This mapping approach involved conducting intensive field surveys over the entire area to be mapped. Although the survey approach is useful for small spatial extents, high costs and time are required to conduct intensive field surveys over large areas. Instead, most current mapping projects of large-scale units over large spatial extents must rely on information gathered remotely.

For at least two decades large-scale ecosystem, soil and vegetation mapping projects covering large extents were based on manual procedures for delineating ecological units from remotely gathered aerial photos (Klinka & Skoda 1977; Mitchell & Eremko 1987; RIC 1998; Scull *et al.* 2003). This required a human interpreter to view and analyze aerial photographs and infer the classification units using visible landscape attributes such as slope gradient, slope position, surface curvature, drainage and vegetation (RIC 1998). Recognizing the high cost (\$3.00 - \$8.00/ha) and time required to manually produce large-scale ecosystem maps from aerial photos, however, research began to focus on producing maps using computer-based predictive approaches (Mulder & Corns 1994; Biggs *et al.* 1997; Jones *et al.* 1999). Computer-based approaches were seen as an attractive alternative to ecosystem maps derived from aerial photo due to their low reliance on field surveys and low costs, which ranged from \$0.20 to \$2.00/ha (Biggs *et al.* 1997).

In BC, procedures that use a computer to stratify the landscape into large-scale ecological units are referred to as predictive ecosystem mapping (PEM) (Jones *et al.* 1999). Only BC (RIC 1999) and Alberta (Mulder & Corns 1994; Beckingham *et al.* 1999) generate large-scale, ecosystem-unit maps using computer-based predictive approaches, although ecosystem-classification systems exist that define large-scale ecological units in the United States (Cleland *et al.* 1997) and New Zealand (Huser 1999), BC (Pojar *et al.* 1987), Alberta (Beckingham & Archibald 1996; Klappstein & Corns 1996; Beckingham *et al.* 1996a), Saskatchewan (Beckingham *et al.* 1996b), Manitoba (Zolaneski *et al.* 1995), Ontario (Chambers *et al.* 1997) and New Brunswick (Zelazny *et al.* 1989).

Although the application of computer-based predictive approaches to large-scale, ecological-unit mapping is limited, it is closely related to the field of predictive vegetation mapping, and predictive soil mapping. All these terrain-related predictive-mapping domains usually start with the development of a numerical or statistical model of the relationship between environmental variables and ecosystem properties, which is then applied to a geographic database to create a predictive map (Jones *et al.* 1999). The methods used to predict ecosystem, soil and vegetation distribution could generally be divided into five separate types: statistical, geostatistical, artificial neural networks, decision trees and expert systems (Scull *et al.* 2003).

## 1.1 STATISTICAL METHODS

Statistical methods are used to explore and model the relationship between quantifiable environmental variables and ecological properties to create predictive maps (Scull *et al.* 2003). Statistical models that have successfully been applied to predictive soils

mapping include linear regression (Moore *et al.* 1993), logistic regression (King *et al.* 1999), and exponential regression (Bell *et al.* 2000). Miller and Franklin (2002) successfully applied generalized linear models to predictive vegetation mapping, and Beckingham *et al.* (1999) used a combination of linear regression and correspondence analysis to successfully predict ecological unit distribution.

Most statistical methods, excluding generalized linear models, however, are limited by their assumption that the data are normally distributed (Scull *et al.* 2003). It is more likely that the mapped units do not fit a normal distribution, but rather fit a multinomial distribution (Foody 2002a). Statistical methods are also limited by their high data requirements. Also, the inflexibility of these statistical methods prevent integration with a variety of potential data sources, such as expert knowledge (Scull *et al.* 2003).

## 1.2 GEOSTATISTICAL METHODS

Similar to statistical methods, geostatistical methods are another approach employed in creating predictive maps. Geostatistical methods are a form of statistics that deal with spatial data and interpolation (Scull *et al.* 2003). Their primary purpose is to spatially interpolate values for locations where values are not known using locations where values are known. Kriging, which was developed by Krige (1963), is the main approach used for interpolation of unknown values. Using this approach, geostatistics have been applied to predictive soil mapping through the work of such authors as: Burgess and Webster (1980), Burrough (1989), McBratney *et al.* (1991), Webster (1994), and McBratney *et al.* (2000).

Criticisms of geostatistic methods, however, include their heavy reliance on data, requiring mapping projects to have a large number of closely spaced data points.

Geostatistical methods work best in simple terrain because they rely on a high amount of spatial autocorrelation, but perform poorly in complex terrain where there are abrupt changes in environmental variables (Zhu 1997). Furthermore, like traditional statistics, geostatistics do not incorporate expert knowledge (Scull *et al.* 2003).

### 1.3 ARTIFICIAL NEURAL NETWORK METHODS

Artificial neural networks (ANNs) are another method employed in the creation of ecosystem, soil and vegetation maps through predictive methods. ANNs are a type of classification model that attempt to emulate the learning process of the human neural network (Atkinson & Tatnall 1997). The term artificial neural network is really a broad term that encompasses a number of approaches that attempt to emulate the neural network process. ANN models first start by creating relationships between a set of known attributes mapped and a set of known outcomes. This data set, which is commonly referred to as a training data set, trains the neural network to establish relationships between the known attributes and the known outcomes. The artificial neural network can then be used to predict outcomes from known attributes once it has established the relationships using the training data (Atkinson & Tatnall 1997). The most commonly used ANN technique for vegetation mapping is the multilayer perception neural network, but other successfully employed techniques include the self-organizing feature-map neural network and the adaptive-resonance-theory network (Atkinson & Tatnall 1997; Carpenter *et al.* 1999; Foody 1999).

The popularity of ANNs has stemmed from their ability to integrate data acquired at a low level of precision, to handle categorical and continuous data sources, and their freedom from linear or simple non-linear distribution assumptions (Atkinson & Tatnall 1997; Zhang

& Foody 2001). The process of training an ANN is computationally intensive and can take considerable time, but once trained an ANN can quickly process new input data (Kimes *et al.* 1996). Criticism of ANNs include their potential to terminate model development at a “local minimum” and produce a sub-optimal model (Kanellopoulos & Wilkinson 1997). On the other hand, ANNs can also produce a model that “over-fits” the training data and produces poor predictions on other independent data sets (Park & Vlek 2002; Qi & Zhu 2003). The potential of over-fitting to the training data increases as the training data increasingly misrepresents the conditions in the data set to be classified (Qi & Zhu 2003).

ANNs have been applied extensively to land-cover classification and predictive-vegetation mapping, primarily where satellite imagery was the main component of the geographic database. Kimes *et al.* (1996) used a multilayer perception neural-network to create a model of forest age using Landsat satellite imagery and topographic data. Linderman *et al.* (2004) successfully used Landsat Thematic Mapper imagery to predict the presence of understory bamboo. Carpenter *et al.* (1997; 1999) successfully applied adaptive-resonance-theory networks to vegetation mapping using Landsat remote sensing imagery and digital terrain data.

The application of ANNs to predictive soil mapping has been less extensive. Qi and Zhu (2003) used neural networks to model the relationships between existing soil maps and terrain variables. Park and Vlek (2002) compared the performance of neural networks, regression trees and general linear models at predicting soil variability using soil type, vegetation attributes and terrain attributes.

## 1.4 DECISION TREE ANALYSIS

Decision-tree (DT) analysis is similar to the ANN approach to predictive mapping. However, DT analysis uses a multi-stage approach when making a prediction, unlike ANNs, which use all the attributes simultaneously and make a single prediction. The prediction process is considered a chain of simple decisions, where tests are applied at each node and each branch represents the resulting decision path. The “leaves” (or branch termini) of the tree represent the prediction labels (Pal & Mather 2003).

Like artificial neural networks, decision trees are developed using a training data set and then applied to a data set of attributes for which predictions are to be generated (Qi & Zhu 2003). Manually “pruning” the tree is often required to prevent the DT from overfitting the training data and to reduce tree complexity (Qi & Zhu 2003; Pal & Mather 2003). Tree pruning is accomplished by merging two leaf nodes into a single node, usually under the direction of expert knowledge (Qi & Zhu 2003).

Advantages of DT analysis over traditional statistical methods include: its ability to handle missing data values and outliers, capture nonadditive and nonlinear behaviour, the lack of assumptions about the data distribution, and ease of updating when new data become available (Moore *et al.* 1991). Also, expert knowledge can be indirectly applied to DTs through tree pruning. This highlights a significant advantage DTs have over neural networks. With DT methods, the relationships established between the input variables and the predicted outcomes are explicit and interpretable by the analyst, unlike neural networks where the relationships are hidden. This explicit presentation of the relationships can help the expert understand which environmental variables influence the relationship between input

and output variables and make adjustments to the tree structure if necessary (Pal & Mather 2003).

DT models, however, are negatively affected by outliers and can produce very different results if outliers are included. Maps produced using DTs are more likely to retain the spatial pattern of the input data because DT models partition the data using one input variable at a time (Miller & Franklin 2002). Pal and Mather (2003) also found that the classification accuracy of decision-tree methods decreased relative to neural networks when the number of input variables used during the model training step exceeded 25. They still achieved high classification accuracy (86%) when classifying land-cover type from satellite imagery. Cialella *et al.* (1997) applied decision-tree analysis to predictive soil mapping to model soil drainage from satellite imagery and digital elevation models.

## 1.5 EXPERT SYSTEMS

Expert systems differ substantially from the statistical, geostatistical, ANN, and decision-tree approaches to predictive mapping. They are computer programs that are built using a human expert's knowledge to simulate the expert's reasoning (Stock 1987). In predictive mapping the purpose of an expert system is to capture the information a surveyor accumulates while working in the field and integrate that knowledge into a predictive model (McCracken & Cate 1986, as cited in Scull *et al.* 2003). Expert systems are generally comprised of data, a knowledge base, and an inference engine (Skidmore *et al.* 1996). The knowledge base contains a set of rules or decisions that define the relationships between the input variables and the output prediction (Stock 1987). The inference engine applies the

input data, which would comprise the GIS database in the case of PEM, to the knowledge base to infer a solution, which in the case of PEM would be the ecological unit.

What makes expert systems different from conventional statistical methods is their ability to store qualitative information and that, essentially, they are meta-models where the knowledge is separate from the model (Davis 1993). This allows expert models to apply the relevant input at different stages of the modelling process, and allows for the input to be updated easily (Scull *et al.* 2003).

Expert systems offer many benefits over traditional statistical approaches. Some of these advantages include the ability to integrate qualitative information (Scull *et al.* 2003), the ability to integrate expertise from multiple sources (Renooij 2001), the ability to capture knowledge before it becomes unavailable, and improved knowledge transfer to novice domain specialists (Zhu *et al.* 2001). Despite these advantages, the expert system does have some limitations. These limitations include being unsuitable in areas where there is little expert knowledge regarding the relationships between input variable and output conditions (Scull *et al.* 2003), and that experts may have difficulty expressing their own expertise or eliciting a set of rules without bias (Renooij 2001).

A criticism of decisions derived from expert opinion is that the result is unfalsifiable because it contains unknown assumptions, limitations and accuracy (Scull *et al.* 2003). For the results from expert systems to be falsifiable the elicitation of expert knowledge must be presented transparently and explicitly (Varis & Kuikka 1997).

Expert systems have been applied to a wide range of predictive-mapping projects. Examples include forest-fire modelling (Davis *et al.* 1986), forest-vegetation mapping

(Skidmore 1989), forest-soil mapping (Skidmore *et al.* 1991), terrain mapping (MacMillan *et al.* 2000; Schmidt & Hewitt 2004), mineral exploration (D'Ercole *et al.* 2000), and land-use planning (Zhu *et al.* 1998). These may employ one of four types of knowledge bases, as described below.

### *Rule-Based Knowledge Base*

Most expert systems use a rule-based knowledge method. In a rule-based system, the relationships between the input variables and output predictions are derived through a series of if-then-else statements. Simple rule-based knowledge bases are binary, where a condition is either true or false; partially-true and partially-false conditions are not allowed. In a rule-based expert system there are two general methods in which the inference engine can make a decision: forward chaining or backward chaining (Moon 1999). In forward chaining, the inference engine compares the attributes of the unknown object to the rule base until arriving at a prediction. This approach is sometimes referred to as data-driven reasoning. In backward chaining, a prediction is hypothesised for an unknown object and the inference compares the attributes of the unknown object to the attributes of the prediction. This approach is sometimes referred to as goal-driven reasoning (Teng & Fairbairn 2000).

### *Belief Matrix Knowledge Base*

A less popular knowledge-base method used in expert-based systems is that of the belief or decision matrix. The belief matrix records the belief that an event will occur given a set of conditions (Moon 1999; Hailu & Sommer 1999). For example, in PEM, a belief matrix would record the belief that an ecological unit occurs on slopes over 50%. The values in a belief matrix can be any positive or negative value, where a highly positive value

indicates a high belief that an event could occur given a condition, and a highly negative value indicates a low belief or impossibility that an event could occur. The inference engine produces a prediction by summing, for each possible output class, the belief values associated with the input conditions. High sums of belief values indicate high belief that the output event occurs given the input conditions, while low belief-value sums indicate low belief.

### *Fuzzy Logic Network Knowledge Base*

A third type of knowledge-base method used in expert based systems is that of the fuzzy-logic network. The fuzzy-logic network uses the theory of fuzzy logic, which attempts to recognize the concept of partial truth developed by Zadeh (1965). The theory of fuzzy logic permits an object to belong to more than one class, recognizing that objects in nature rarely fit exactly the classification types to which they are assigned (Scull *et al.* 2003). The fuzzy-logic network extends the rule-based knowledge base to include partial membership to decision paths. When the inference engine processes the input variables through a fuzzy-logic network, the output is a membership value, usually ranging from zero to one, for each possible output class. Membership values of zero indicate no membership while values of one indicate full membership.

### *Bayesian Network Knowledge Base*

Lastly, Bayesian networks (BN) are another common type of knowledge-base method used in expert-based systems. Bayesian networks, also known as belief networks, Bayesian belief networks, Bayes networks and causal probabilistic networks, were first introduced to probability-based decision-systems through the work of Pearl (1988). BNs are so called

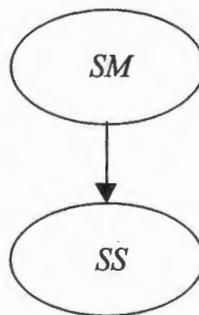
because they commit to Bayes' rule for probabilistic inference. This rule, developed by mathematician and theologian Reverend Thomas Bayes, generally states that our belief that an event will occur is based on the number of times the event has or hasn't occurred in the past given similar conditions. Bayesian methods have been gaining use in applied science, including ecology. Fisheries biology and management has, by far, seen the greatest use of Bayesian methods (Marcot *et al.* 2001). There has been to date, however, no application of Bayesian networks to predictive ecosystem mapping.

Bayesian networks, like fuzzy-logic networks, are numerical models that are primarily built as graphical networks, sometimes referred to as directed-acyclic graphs (Jensen 1996). Bayesian networks consist of the following: a set of variables, also referred to as nodes, which have a finite set of mutually exclusive states, and a set of directed edges, or links, among the variables. When focusing on a single node, parent nodes are those nodes that are at the beginning of a link, which is pointing at the node of interest. Child nodes are those nodes that are at the end of a link, which is exiting the node of interest. BNs cannot contain feedback loops, where the child of a node is also the node's parent or grandparent in some way (Jensen 1996). Those nodes that have no parent are commonly referred to as root or input nodes, while those nodes with no children are referred to as leaf or output nodes (Freidman *et al.* 1997).

Bayesian networks are similar to belief matrices, however, instead of the knowledge base containing a single large table, the BN simplifies the problem by breaking the single large table into several smaller tables. A conditional-probability table is associated with each variable with parents, while an unconditional-probability table is associated with each variable with no parent (Jensen 1996). A complete Bayesian network (i.e., nodes, links and

conditional probabilities) can wholly be defined using expert knowledge. When the inference engine processes input variables through a Bayesian network the output is a posterior-probability distribution for the output variable, which contains the occurrence likelihood of each output class.

A simple Bayesian network could be envisioned as one parent node, which could represent some environmental variable such as soil moisture (SM), and one child node, such as site series (SS).



According to Bayes theorem:

$$P(SS | SM) = \frac{P(SM | SS)P(SS)}{P(SM)}$$

Using this theorem, the probability of a given site-series (SS) outcome given soil moisture (SM) is dependent on the likelihood of SM given a specific value of SS ( $P(SM|SS)$ ) and on the prior probabilities of both SS ( $P(SS)$ ) and SM ( $P(SM)$ ). The result is the “posterior” probability for site series (SS).

An important aspect of Bayesian networks is the ability to update the conditional probabilities using a “learning” algorithm such as “expectation maximization” (Dempster *et al.* 1977). Learning generally involves using observations to update the conditional probabilities of some or all of non-parent nodes to reflect the conditions encountered in the

observations (Plach 1999). The values for variables that have no findings are determined through causal inference (Norsys 1997).

## 1.6 HARD VERSUS SOFT CLASSIFICATION

Statistical, geostatistical, ANNs, decision-trees, and expert-systems approaches to predictive mapping can all produce hard classifications (sometimes referred to as crisp classifications). Some of these approaches can also produce soft classifications (sometimes referred to as fuzzy classifications). In a hard classification, only one output class is permitted per observation or location. Soft classification, however, allows an observation or location to belong to several or all possible output classes. The soft-classification result records how strongly an observation or location represents each class (MacMillan *et al.* 2000). Bayesian networks and fuzzy-logic models are both examples of approaches that can produce both hard and soft classifications.

Predictive-mapping projects are increasingly producing soft-classification results (MacMillan *et al.* 2000; Zhang & Foody 2001; Foody 2002a; Foody 2002b). Soft classification can capture prediction uncertainty whether it is a result of class-definition (thematic) uncertainty or spatial uncertainty (Zhang & Foody 2001). In predictive mapping, there are mainly two advantages that soft classification has over hard classification. Firstly, many predictive-mapping projects attempt to map concepts that are not discretely defined within the landscape. Instead, many class concepts occur as a continuum across the landscape, which overlap with other class concepts. For example, when mapping landscape elements such as ridge, hill, valley, etc., rarely do these semantically defined features occur discretely or mutually exclusively across the landscape. Some locations will strongly match

the semantic definition of a ridge, some locations will weakly match the concept of a ridge, although some locations will not match the concept of a ridge at all. Also, a location may weakly represent the concept of a ridge but strongly represent the semantic definition of a hill. Soft classification can capture this continuum of landscape features that are non-mutually exclusive, but hard classification can not. The second advantage of soft classification involves the scale of the observation or input data. If the scale or resolution of the input data is such that a boundary separating two different feature classes occurs within the minimum-mappable area, then soft classification can capture the proportional occurrence of each class within that area. The occurrence of multiple features within the minimum-mappable area is affected by the latter's size. The larger the minimum-mappable area the greater the chance that more than one class exists within its area boundary. This is sometimes referred to as a "mixed pixel" in satellite-imagery classification, because more than one feature can exist within an image pixel (Foody 1996).

Similar to soft-classification output, which allows the *outputs* to contain fuzzy membership, is the semantic-import model of Burrough (1989), which allows the *inputs* to contain fuzzy membership. Using the semantic-import model, the hard thresholds between classes can be converted to fuzzy boundaries. For example, in a hard classification the classification of slope gradient into flat and sloped would occur at a single slope-gradient value. Using the semantic-import model, the belief that a location belongs to either the flat or sloped class would vary according to a range of slope-gradient values. With the semantic-import model, instead of the transition between sloped and flat being determined by a single value, the transition from one class to another occurs over a zone defined by two values ( $ST_1$  and  $ST_2$  in Figure 1). The advantage of the semantic-import model over the hard-class

approach is that the semantic-import model captures the uncertainty inherent in the semantic definition of landscape features (Burrough 1989). This uncertainty in class membership may result from either the continuous nature of the input or the varying semantic definitions of the input classes.

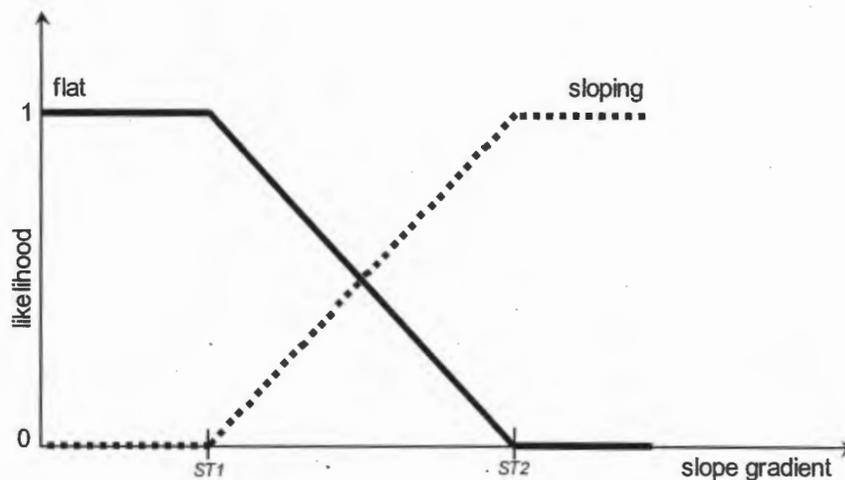


Figure 1 – A graphical representation of the semantic-import model used to define the relationship between slope gradient and the semantic descriptions of flat and sloping. Slope gradients between zero and *ST1* are considered flat, and slope gradients greater than *ST2* are considered sloping. As slope gradients increase from *ST1* to *ST2*, the likelihood that the slope is considered flat decreases linearly to zero while the likelihood that the slope is considered sloping increases linearly to one. Flat and sloping semantic-descriptions are considered equally likely at the mid-point between *ST1* and *ST2*, although both likelihoods are 0.5.

Although maps resulting from soft classification are gaining popularity, in some circumstances it may be most appropriate to produce a hard classification. For example, decision makers often find probability distributions unclear and prefer a single answer upon which to make decisions. In addition, accuracy-assessment procedures are often based on hard-classification results. To this end, it is not necessary to solely produce hard-classification results, but rather the results from a soft-classification model can be “hardened”. Classification “hardening” often involves taking the dominant class from each soft-classification result (Foody 1996). This usually means choosing the class with the

highest likelihood (Skidmore *et al.* 1996), but in the case of fuzzy logic, the dominant class is the class with the highest membership value (Zhang & Foody 2001).

## 1.7 ACCURACY ASSESSMENT

Most assessments of map accuracy are built around hard-classification results. The contingency table, which is also known as the contingency matrix, error matrix and confusion matrix, is currently at the core of accuracy-assessment procedures for land-cover classification and is used extensively for accuracy assessment in the field of remote-sensing classification (Congalton & Story 1986; Foody 2002b). The contingency table is simply a square array that relates the number of sample units assigned to a particular class in one classification relative to the number of sample units assigned to a particular class in another classification (Congalton & Green 1999). Contingency tables are almost exclusively applied to hard-classification results, because they are unable to handle soft-classification results. Commonly, one of the classifications is considered correct and is assigned to the matrix columns while the predicted classification is assigned to the matrix rows. The major diagonal then indicates the agreement between the two data sets.

The contingency table is an effective way to portray individual accuracies of each class along with both the errors of inclusion (commission errors) and errors of exclusion (omission errors). Commission errors result from putting an area into a class to which it does not belong: committing the act of getting it wrong. Omission errors result from excluding an area from a class to which it does belong: omitting the act of getting it right. Omission and commission errors are related, because every error is both an omission from the correct class and a commission to the incorrect class. Other accuracy-assessment measures commonly

derived the contingency matrix are: percentage correct, producer's accuracy, user's accuracy and kappa analysis (Congalton & Story 1986).

*Percentage correct* is commonly used to represent overall map accuracy (Congalton & Story 1986). To derive this measurement the diagonal elements of the contingency table are summed and divided by the total number of samples.

*Producer's accuracy* is a class-specific accuracy-assessment measure that reports how well a specific landscape feature can be mapped (Congalton & Story 1986). Its calculation involves taking the total number of correctly classified samples in a class and dividing by the total number of reference samples in the same class. The result indicates the probability that a sample is placed in the correct class, expressed as a percentage. The producer's accuracy for each class essentially measures the error of omission; the proportion of samples that are omitted from the correct class, where the error of omission is equal to one minus the producer's accuracy.

*User's accuracy* is a class-specific accuracy-assessment measure that essentially measures the error of commission. It is closely related to the producer's accuracy because every error of omission in one class is an error of commission into another class. The user's accuracy is so named because it measures the reliability that the map actually represents what is really on the ground. It is calculated by taking the total number of correctly classified samples in a class and dividing by the total number of samples that are placed in that class. The percentage generated by this calculation indicates to the map user the probability that a location on the classified map actually represents that category on the ground (Congalton & Story 1986).

*Kappa analysis* is a discrete multivariate technique that expresses the degree of agreement between two classifications while compensating for chance agreement (Congalton & Green 1999). The result of performing a Kappa analysis is a KHAT statistic, also denoted  $\hat{K}$ , which is an estimate of Kappa.

The KHAT statistic was introduced into the field of land-cover classification through the work of Congalton and Mead (1983), and has been used extensively to assess the accuracy of land-cover classifications derived from remotely sensed imagery (Foody 2002b). KHAT values range from  $-1$  to  $1$ . Positive values KHAT values, however, should be expected, because there should be a greater than random, or chance, agreement between the classification map and observation data. Congalton and Green (1999) suggest using three KHAT value classes to express agreement. KHAT values greater than  $0.8$  represent strong agreement, values between  $0.4$  and  $0.8$  represent moderate agreement and values less than  $0.4$  represent poor agreement. In addition to the KHAT statistic, the KHAT variance (Congalton & Green 1999) and Z statistic (Devore 2000) can be calculated to determine if the generated classification map is significantly greater than a random, or chance, classification.

For map accuracy assessment, Congalton and Green (1999) recommend 50 reference samples per class, however, they also recognize that it may not be possible to obtain 50 samples for rare classes. In the case of rare classes it is acceptable to collect less than 50 samples as long as a minimum number of samples collected per class (Congalton & Green 1999). If obtaining 50 samples per class is not possible then Congalton & Green (1999) recommend a balance between the project's time and cost limitations, and the need to adequately populate the contingency matrix.

## 1.8 STATEMENT OF THE PROBLEM AND OBJECTIVES

In BC, ecosystem maps have been applied most widely to forest-related management activities. Managers of forest-related activities routinely rely on detailed, large-scale ecosystem maps where the base ecological unit typically occupies no more than 10 ha. Forest-development planning, wildlife-habitat suitability/capability mapping, and ecologically-based yield analysis with site-based site-index curves are all example uses of large-scale ecosystem maps for forest-related management (Biggs *et al.* 1997).

During the past six years, predictive mapping of large-scale ecological units has been developing in BC to meet the management needs of forest-related activities. All of the completed PEM projects have attempted to predict the spatial distribution of site-series ecological units from the biogeoclimatic ecosystem-classification scheme of BC. The first pilot projects covered small areas and employed expert-based systems, utilising either a rule-based or a belief-matrix knowledge-base approach (Jones 1999). Since that beginning, over 40 PEM projects have been completed. Although plans exist to complete large-scale ecosystem mapping of the entire province, the completed projects now cover approximately 15% of the province's land base – primarily in non-coastal regions. Of the 32 completed project reports available for review, 28 reported using an expert-based system with a belief-matrix knowledge base. Each of the completed projects have first stratified the landscape using the mapping method of Eng and Meidinger (1999), which delineates biogeoclimatic subzone/variants, and then developed a separate expert-based system for each subzone/variant stratum.

All of the completed large-scale predictive ecosystem maps have been evaluated using the 65% correct *de facto* standard established by Meidinger (2003). Based on the review of completed projects, results of map-accuracy assessment have been mixed, with approximately half of the mapped BEC-subzone/variants only just meeting the 65% correct target. Approximately 25% of the mapped subzone/variants achieved percent-correct scores well above the 65% target and approximately 25% did not meet this target. Achieving 65% correct targets, however, does not guarantee the applicability of the map product to all forest-related management activities. Meidinger's (2003) 65% correct criterion was established to evaluate whether a large-scale predictive ecosystem map could be included in a timber supply analysis. A predictive site-series map may not be suitable for such purposes as establishing silviculture prescriptions or identifying rare ecosystems, especially if the map only just achieves the 65% correct target.

Efforts to create large-scale predictive-ecosystem maps in BC (e.g., Jones 1999; Timberline Forest Inventory Consultants Ltd. 2000; Rosen *et al.* 2001; Atticus Resource Consulting Ltd. 2001; Ketcheson *et al.* 2001a; Ketcheson *et al.* 2001b; Ketcheson *et al.* 2001c; Jones & McGregor 2002; Sulyma & Alward 2004; Jones 2004), however, have encountered several key problems, specifically: 1) the procedure used to model the relationships between environmental variables and ecological units has been difficult to understand for anyone other than those individuals who created the model; 2) the model of the relationships between environmental variables and ecological units could not be automatically updated with field observations; and 3) the resulting spatial database did not report prediction uncertainty.

To address the problems encountered by earlier PEM projects, and to meet the demand for large-scale ecosystem maps by managers of forest-related activities, this project aimed to develop and evaluate a modelling procedure that:

- 1) predicted the spatial distribution of large-scale ecological units, with a minimum 65% correct, through the incorporation of expert knowledge, but still allowing automated model adjustment based on field gathered observations;
- 2) improved the interpretability of the ecological relationships for users other than the ecologist used to develop the model; and
- 3) retained prediction uncertainty.

## 2 METHODS

### 2.1 STUDY AREA

The study area was located along the west coast of mainland BC, centred around Work Channel, and falling roughly between 54°4'N 130°31'W and 54°49'N 129°51'W (Figure 2). The study area was approximately 150,000 ha and was situated between Portland Inlet and the mouth of the Skeena River. It was defined by the biogeoclimatic subzone/variant units: Coastal Western Hemlock Very Wet Maritime Montane (CWHvm2), Coastal Western Hemlock Very Wet Maritime Submontane (CWHvm1), and Coastal Western Hemlock Very Wet Hypermaritime Central (CWHvh2), that fell within six BC Ministry of Sustainable Resource Management landscape units: Tuck, Kaien, Quottoon, Khutzeymateen, Somerville and Union.

I chose this study area for two reasons. Firstly, the area appeared to contain an adequate amount of georeferenced field data to aid model development and perform accuracy assessment. Secondly, the study area overlapped with the North Coast PEM Project study area (Jones 2003), which was a previously completed PEM project that employed a belief matrix expert system. I felt that overlapping with a previously completed PEM project would increase the availability of the GIS data required to build the GIS database.

The study area lies within the Coast and Mountains Ecoprovince of the Humid Maritime and Highlands Ecodivision of the B.C. Ministry of Sustainable Resource Management's (MSRM) ecoregional classification (Demarchi 1995). The climate is generally temperate and rainy with typically warm summers. The climate is dominated by

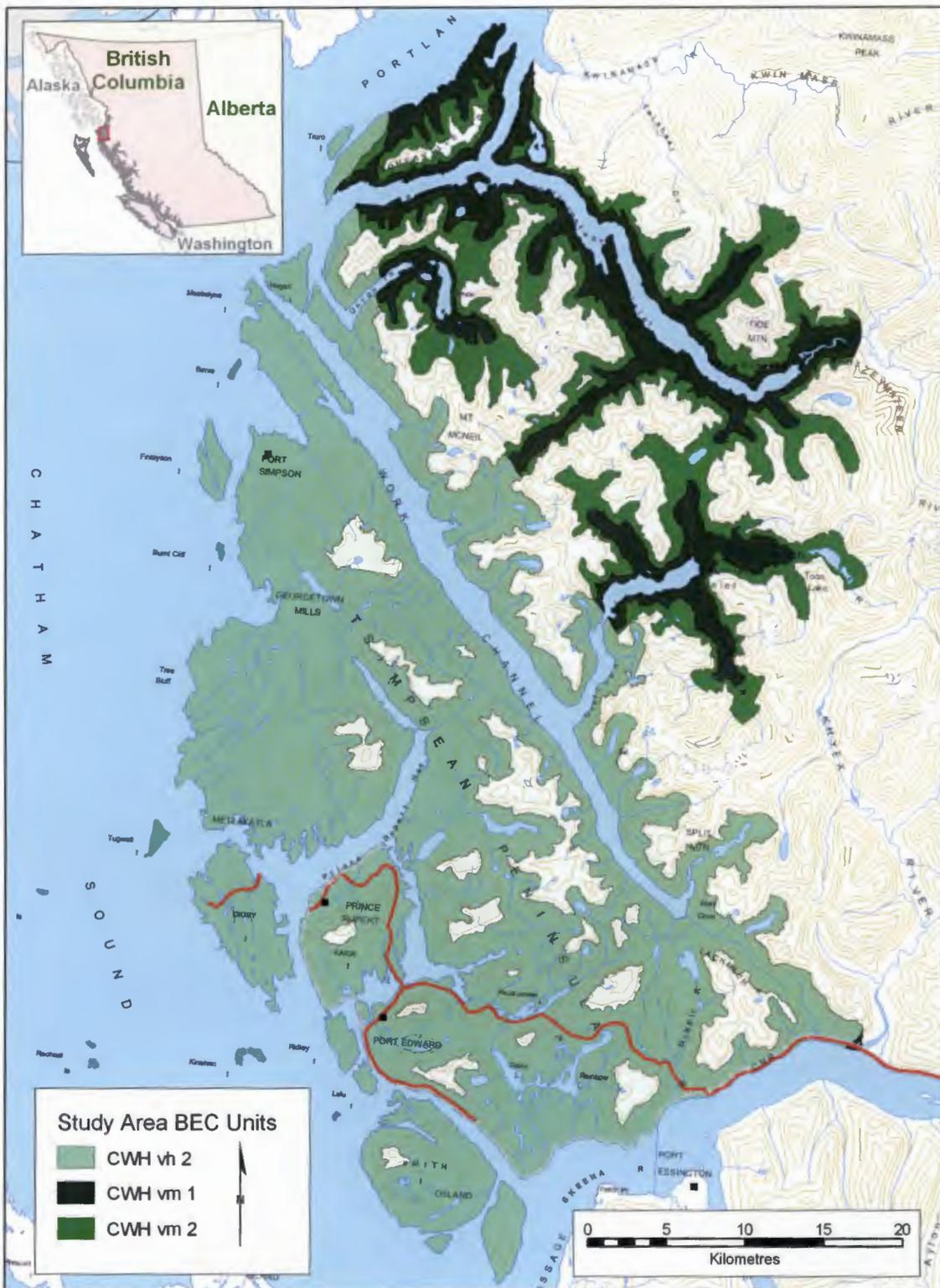


Figure 2 – The study area which was used to conduct large-scale predictive ecosystem mapping. The study area was comprised of the CWHvm2, CWHvm1 and CWHvh2 BEC units within the BC Ministry of Sustainable Resource Management Tuck, Kaien, Quottoon, Khtuzeymateen, Somerville and Union landscape units. It was approximately 150,000 ha and centred on Work Channel in North-Coastal BC.

the arrival of frontal systems from the Pacific Ocean and their subsequent lifting over the Coast Mountains. As a result, rain is abundant throughout the year, but markedly lower in the summer months. Ecoprovinces are divided into ecoregions, and the study area overlapped with both the Coastal-Gap Ecoregion and the Northern Coastal Mountains Ecoregion. The Northern Coastal Mountains Ecoregion, which comprises the northern third of the study area, contains typically large, rugged, ice-capped mountains that rise sharply from the ocean. The Coastal Gap Ecoregion portion of the project area contains somewhat lower and more rounded mountains, although valley sides are still rugged and steep (Campbell *et al.* 1990).

Starting at sea level, the Coastal Western Hemlock (CWH) zone occupies elevations less than 600 m near the ocean and less than 800 m inland. Western hemlock (*Tsuga heterophylla*) is usually the most common tree species although western redcedar (*Thuja plicata*) is abundant throughout the Coastal Gap Ecoregion. Amabilis fir (*Abies amabilis*) and yellow-cedar (*Chamaecyparis nootkatensis*) are both found in wetter sites, but with amabilis fir dominating the zone's high elevations and northern latitudes. Douglas-fir (*Pseudotsuga menziesii*) is abundant in drier parts of the zone, which lie south of the study area (Meidinger & Pojar 1991).

The hypermaritime portion of the study area contains two distinct areas: the lowland areas and the mountainous areas. The lowland area is dominated by subdued terrain occurring adjacent to the ocean, while the mountainous area is dominated by rounded mountains occurring adjacent to Work Channel (Banner *et al.* 1993). Bog woodlands, bog forests and swamp forests occur extensively in the low-lying areas of the CWHvh2. These

sites are characterized by extremely wet soils with poor nutrient availability and poor soil-moisture transmissivity, especially at low slope gradient. The upland sites of the CWHvh2 are dominated by wet and mesic soils with good soil transmissivity (Banner *et al.* 1993).

The maritime portion of the study area occurs inland from the CWHvh2 and is dominated by large rugged mountains. In the CWHvm2 and CWHvm1 BEC units, which occur in the maritime climate, soil transmissivity is generally constant with slope gradient. This is in contrast to the CWHvh2, which has extremely low soil-moisture transmissivity on low slope gradients (Banner *et al.* 1993).

The subalpine elevations that lie above the CWH zone are occupied by the Mountain Hemlock (MH) and Alpine Tundra (AT) biogeoclimatic zones (Meidinger & Pojar 1991). The MH zone was excluded from the study area due to a lack of field data. The AT zone was excluded because it typically contains very little forest and did not fit with the project's objective of producing ecosystem maps that aid the management of forest related activities.

## 2.2 METHODS OVERVIEW

I used a multistage methodology (Figure 3) to create the expert system and meet the objectives in this study. Identifying the required GIS database layers was my primary step in the methodology. To verify the spatial and thematic accuracy of the GIS layers I used expert opinion and comparisons to other spatial-data sets.

To build the Bayesian-network knowledge base I used Norsys Software Corporation's Netica (version 2.17) software application (Norsys 1997). Building a Bayesian network in Netica generally involved creating: 1) input, output and intermediate nodes, 2) links between

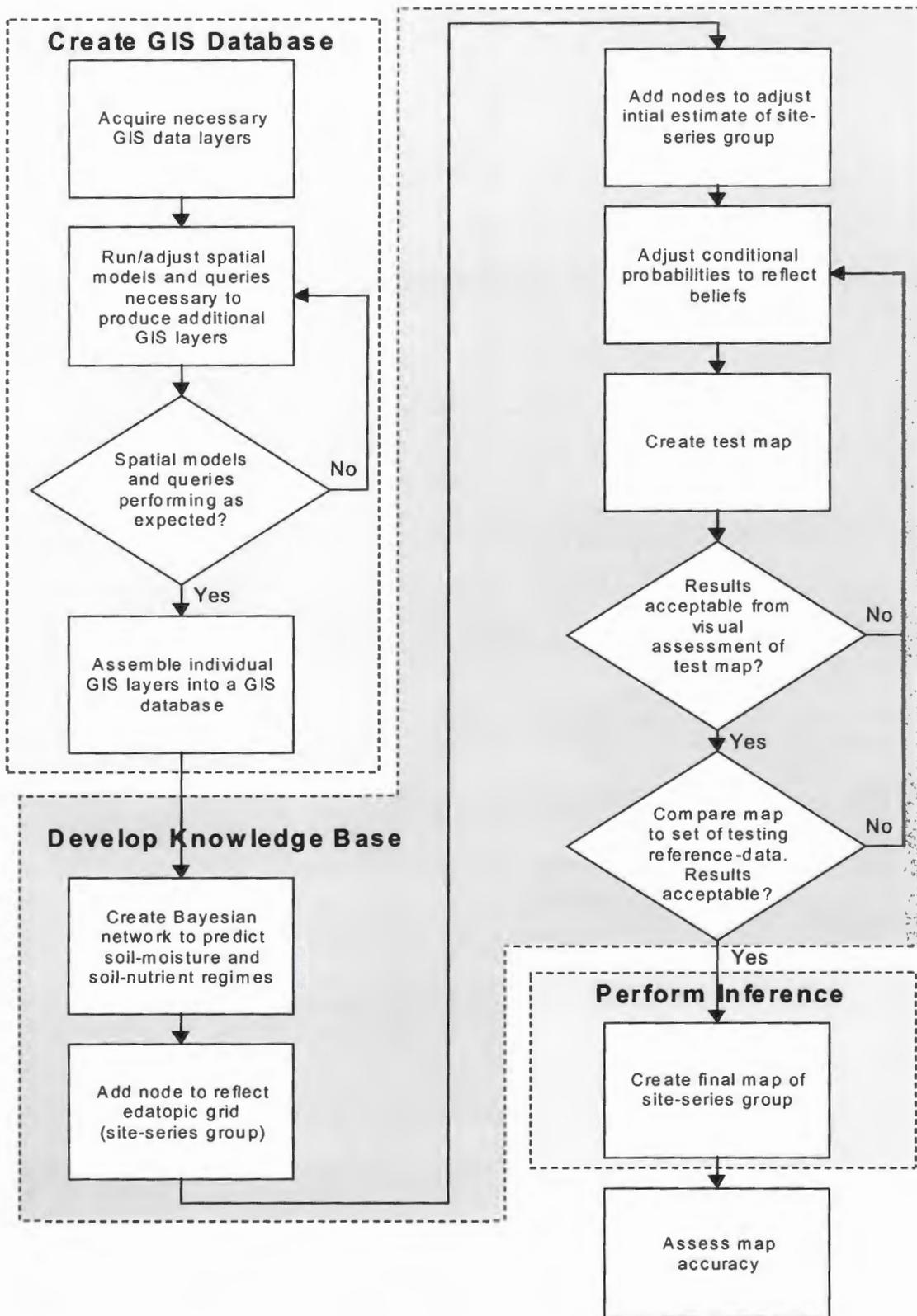


Figure 3 – This diagram illustrates the methodology used to meet the objectives of this study. There were three general sections in the methodology that also comprises most expert systems: create the database, develop the knowledge base and perform inference.

the nodes 3) classes for each node, and 4) conditional-probability tables.

I created an initial Bayesian network to predict soil-moisture and soil-nutrient regimes. I made a preliminary site-series-group prediction by adding a node to reflect the edatopic-grid relationships between soil-moisture and soil-nutrient regimes in each BEC subzone/variant. By adding new nodes and node relationship, I adjusted the initial site-series-group prediction, which was derived from the edatopic grid. Using the advice of the ecological-domain expert, I adjusted the conditional probabilities by assessing those posterior probabilities that reflect the entire study-area input conditions.

I used two procedures to test the knowledge base: 1) visual assessment of a map of predicted site-series group that covered a portion of the study area, and 2) comparison of predicted site-series group with a reference-data set. If the results of either of these two testing procedures were not satisfactory then the knowledge base received further conditional-probability value adjustments.

Using Netica as the inference engine, I processed the entire GIS database through the knowledge base to generate a final predictive ecosystem map. To assess the accuracy of the final ecosystem maps, I compared the map results to an independent reference-data set using several statistical procedures in Microsoft Excel 97.

### 2.3 GIS DATABASE DEVELOPMENT

Prior to Bayesian-network knowledge base construction, the ecological-domain expert, Del Meidinger of the BC Ministry of Forests, and I reviewed a belief matrix developed for the North Coast PEM Project by Jones (2003). Although encompassing a

larger geographic area and a greater number of BEC subzone/variants, the North Coast PEM Project contained predictive site-series mapping of the CWHvm2, CWHvm1 and CWHvh2 BEC units. This review of the North Coast PEM Project allowed us to understand how existing and modelled GIS layers influenced site-series prediction, and which GIS layers could be available for the Bayesian-network knowledge base.

Based on that review, we identified 14 GIS layers that we considered ecologically significant for site-series prediction. Four of these layers were directly available from an existing GIS database (described below). The remaining 10 were a product of applying spatial models to existing GIS databases (described below).

The four GIS layers directly available from an existing GIS database were forest-stand age, forest-stand height, forest-stand species, and forest-stand crown closure. I acquired all four forest-stand attributes from the B.C. Ministry of Sustainable Resource Management's (MSRM) vegetation resource inventory (VRI) GIS database.

The 10 GIS layers that were a product of applying spatial models to existing GIS databases were: ridge, hill, river bench, lake/wetland bench, gully, slope gradient, bedrock, stream density, ocean wind, and toe slope. I derived these GIS layers using the same logic as Jones (2003). Using Arc/Info version 8.3 GIS software from Environmental Systems Research Institute (ESRI), I created a triangulated irregular network (TIN) from MSRM's terrain resource information management (TRIM) digital elevation points and air-photo-interpreted ridges and gullies. I modelled the GIS layers of slope gradient, ridge, river bench, lake/wetland bench, and gully using the TIN digital terrain model (DTM) (Table 1).

Table 1 – A summary of the Jones (2003) rule-based logic used to create the TIN/DEM dependant GIS layers used in this project.

<b>GIS Layer</b>	<b>Rule-Based Logic</b>
Slope Gradient	A natural component of a TIN. Slope gradient measures the maximum rate of change in elevation across the triangle.
Ridge	Ridge - TIN polygons with a slope gradient greater than 40% and that were adjacent and within 20m of a TRIM ridge
	Ridge Buffers - TIN polygons with a slope gradient greater than 40% and that were between 20m to 40m of a TRIM ridge
River Bench	Low River Bench - TIN polygons with a slope gradient less than 5% and that were adjacent and within 100m of a fourth order TRIM definite stream
	High River Bench - TIN polygons with a slope gradient less than 5% and between 100m and 200m from a fourth order TRIM definite stream
	River Terrace - TIN polygons with a slope gradient between 5% and 10% and that were between 200m and 500m from a fourth order TRIM definite stream. Or, polygons with a slope gradient less than 10% and that were between 500m and 1000m from a fourth order TRIM definite stream.
Gully	Gully - TIN polygons with a slope gradient greater than 30% and that were adjacent and within 20m of a TRIM gully
	Gully Buffer - TIN polygons with a slope gradient greater than 30% and that were between 20m and 40m of a TRIM gully
Lake/Wetland Bench	TIN polygons with a slope gradient less than 5% and that were adjacent and within 200m of a TRIM lake or wetland
Hill	Hill - Hills were derived from TRIM contour lines. A hill was represented by the largest outside contour line, that would form a closed area, and was less than or equal to 1200m in length, but which was not a depression.
	Hill Buffer - TIN polygons with a slope greater than 30% slope gradient and were adjacent and within 40m of a hill polygon
Toe Slope	Using a digital elevation model, a location was considered to be a toe slope if slope gradient was less than 25% and was down-slope and within 100 metres of an area with a slope gradient greater than 40%

Toe slope was generated from a raster digital elevation model (DEM), which was produced from the TIN (Table 1). During the creation of a DEM from a TIN, I set the raster resolution at 625 m<sup>2</sup>, or 1/16 ha, which was the product of a 25 x 25 m pixel. I chose this size because it approximated the minimum geographic spacing between TRIM elevation points along ridge and gully terrain features. This resolution also matched the MSRM provincial raster DEM product. The remaining GIS layers that I created using Jones (2003) procedures were not derived directly from a TIN or DEM but were generated from other data sources (Table 2).

Table 2 – The GIS layers of bedrock, stream density, and ocean wind, which were derived using the procedures of Jones (2003), were not derived solely from the TIN or raster DEM, but rather depended on other data sources.

GIS Layer	Procedure
Bedrock	The B.C. Ministry of Energy and Mines' 1:250000 bedrock geology GIS database was reclassified into 5 classes: limestone, volcanic, rich, intermediate, and poor.
Stream Density	The length of MSRM TRIM stream that intersects a stream density "base-unit" polygon. The "base-unit" polygon was defined by the intersection of VRI polygons with TIN slope gradient polygons that has been simplified into five classes, and with TIN slope-aspect polygons that has been simplified into two classes.
Ocean Wind	Those areas of land that were eastwardly visible from a sea-level line running down the middle of Chatham Sound, to a maximum elevation of 50 m.

In addition to those GIS layers generated using the procedures of Jones (2003), the ecological-domain expert and I identified five other GIS layers useful in predicting site series. These GIS layers were floodplains, fans, soil moisture, shoreline and a continuous landform-elements layer.

The floodplains layer and the continuous-landform-elements layer, which delineates ridges, shoulders, spurs, hollows, back-slopes, terraces, foot-slopes and valley-bottoms, were created to supplement the individual GIS landform layers (ridge, gully, hill-top, etc.) and the river-benches layer that were derived using the Jones (2003) procedures. This supplementation was conducted due to the ecological-domain expert's concern with some of the Jones (2003) procedures used to generate landform elements and river floodplains (see Figure 4 for an example). The ecological-domain expert and I felt that by incorporating two methods of modelling landform elements and river floodplains within the knowledge base we could increase our chances of successfully predicting the location of landform elements and river floodplains.

I modelled floodplains, fans and soil moisture in the GIS software application SELES (Spatially Explicit Landscape Event Simulator), which was available from Gowlland Technologies Inc., using routines developed by Fall (2002). To generate these knowledge base inputs, I used the same DEM that was used to generate the toe slopes.

I used the SELES' floodplain model (Fall 2002) to generate floodplains. Moving outward from the location of approximate fourth-order streams, the floodplain model accumulated cost at a rate of  $(1 + \textit{percent slope gradient})^{1.5}$  until the cost exceeded 100. I divided the resulting floodplain-cost layer into low-river bench, high-river bench and river terrace based on a visual comparison with air photos, satellite imagery and DEM terrain. I considered floodplain-cost values less than 10 to represent low-river benches; cost values between 10 and 35 to represent high-river benches; and cost values greater than 35 to represent river terraces.

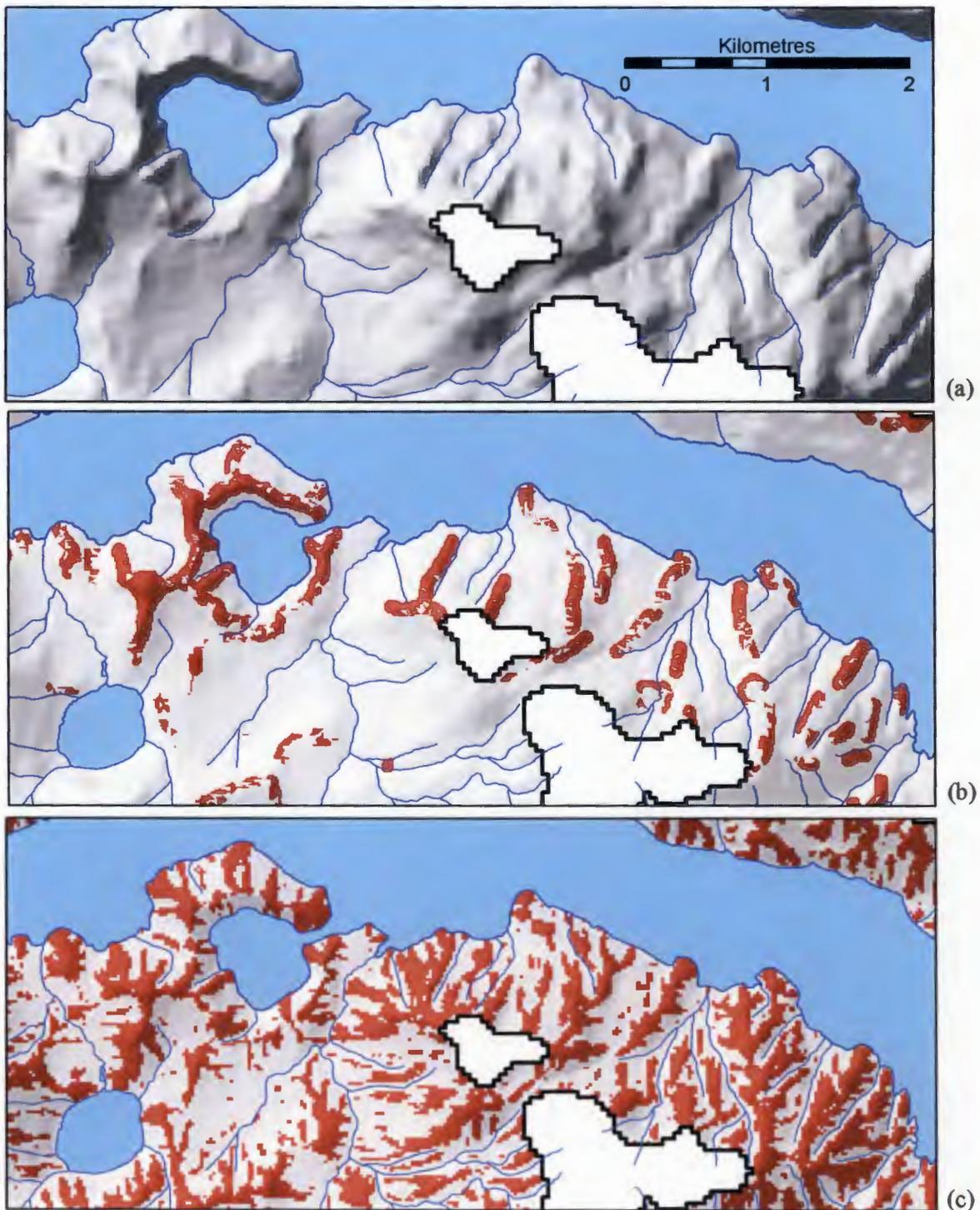


Figure 4 – This sample map (a) demonstrates the difference between the upper-landform elements as defined by Jones (2003) (b) and the upper-landform elements as defined by the continuous-landform-elements layer (c). After reviewing the results of the Jones (2003) ridges and hill layers (b), the ecological-domain expert and I felt that these layers were underestimating upper-landform elements (convex areas that shed water). Additional upper-landform elements (c) were identified using the continuous-landform-elements layer, which was based on the Schmidt and Hewitt (2004) procedure.

I delineated alluvial fans using the SELES alluvial-fan model (Fall 2002). The model first identifies confluence points between medium-sized streams and floodplains or larger bodies of water. Starting from these confluence points, the fan model spreads across slope gradients between 10% and 20% without accumulating cost. Cost will accumulate if slope gradients exceed 20% or drop below 10%. To control fan size, I set the maximum allowable accumulated-cost to be 500 or 1000 m from the confluence point.

I estimated soil moisture using the SELES soil-moisture model (Fall 2002), which implements the SINMAP algorithm developed by Pack *et al.* (1998). In the SINMAP soil moisture algorithm, a raster cell's soil moisture is related to the amount of upstream area that is feeding into the cell and the cell's slope (Pack *et al.* 1998). Although this algorithm can incorporate local rates of soil transmissivity and recharge, these rates were unknown within the study area. Consequently, the default values were used for transmissivity and recharge.

I created the shoreline GIS layer using raster analysis functions in Arc/info. This GIS layer represented all areas less than 5 m above-sea-level that were adjacent to the ocean. Based on the advice of the ecological-domain expert, I chose 5 m because it approximated the influence of tidewater and large ocean swells on the shoreline.

I developed the layer of continuous landform elements using an approach similar to Schmidt and Hewitt (2004), which uses the semantic-import model of Burrough (1989). Using the raster DEM, each cell was assigned a membership value for the landform elements: ridges, shoulders, spurs, hollows, back-slopes, terraces, foot-slopes and valley-bottoms. To assign membership values, however, I used a Bayesian network, instead of the fuzzy-logic approach used by Schmidt and Hewitt (2004). The follow section describes the

Bayesian network of continuous landform elements, which was created for this study.

During this description, and during subsequent descriptions of Bayesian networks, node names will be displayed in **bold** and node classes will be displayed with an underline.

Using the Schmidt and Hewitt (2004) approach, six GIS input layers were required to produce the continuous-landform-elements layer: tangential curvature (Dikau 1989), profile curvature (Dikau 1989), maximum and minimum curvature (Wood 1996), slope gradient and terrain context. I generated each of the four types of curvature, plus slope gradient, using Wood's (1996) multi-scale geomorphic characterization of DEMs. I created terrain context, which describes whether a raster cell was a valley or hill, using the TOPHAT function of Rodriguez *et al.* (2002).

Using the logic of Schmidt and Hewitt (see Figure 3 in Schmidt & Hewitt 2004), I used combinations of tangential curvature (**TangentialCurve**) and profile curvature (**ProfileCurve**) to assign sloping areas into one of nine local form elements: shoulder (Shoulder), shoulder slope (ShoulderSlope), hollow shoulder (HollowShoulder), spur (Spur), planar spur (PlanarSpur), hollow (Hollow), spur foot (SpurFoot), foot slope (FootSlope) and hollow foot (HollowFoot). This combination of tangential curvature and profile curvature was represented in the Landform BN by the **SlopingShape** node, which was conditional on the **TangentialCurve** node and the **ProfileCurve** node (Figure 5). Again using the logic of Schmidt and Hewitt (see Figure 3 in Schmidt & Hewitt 2004), I used combinations of maximum curvature (**MaxCurve**) and minimum curvature (**MinCurve**) to assign flat areas into one of six local form elements: peak (Peak), ridge (Ridge), plain (Plain), saddle (Saddle), channel (Channel) and pit (Pit). Figure 5 depicts this relationship, with the **FlatShape** node being conditional on the **MaxCurve** and the **MinCurve** nodes.

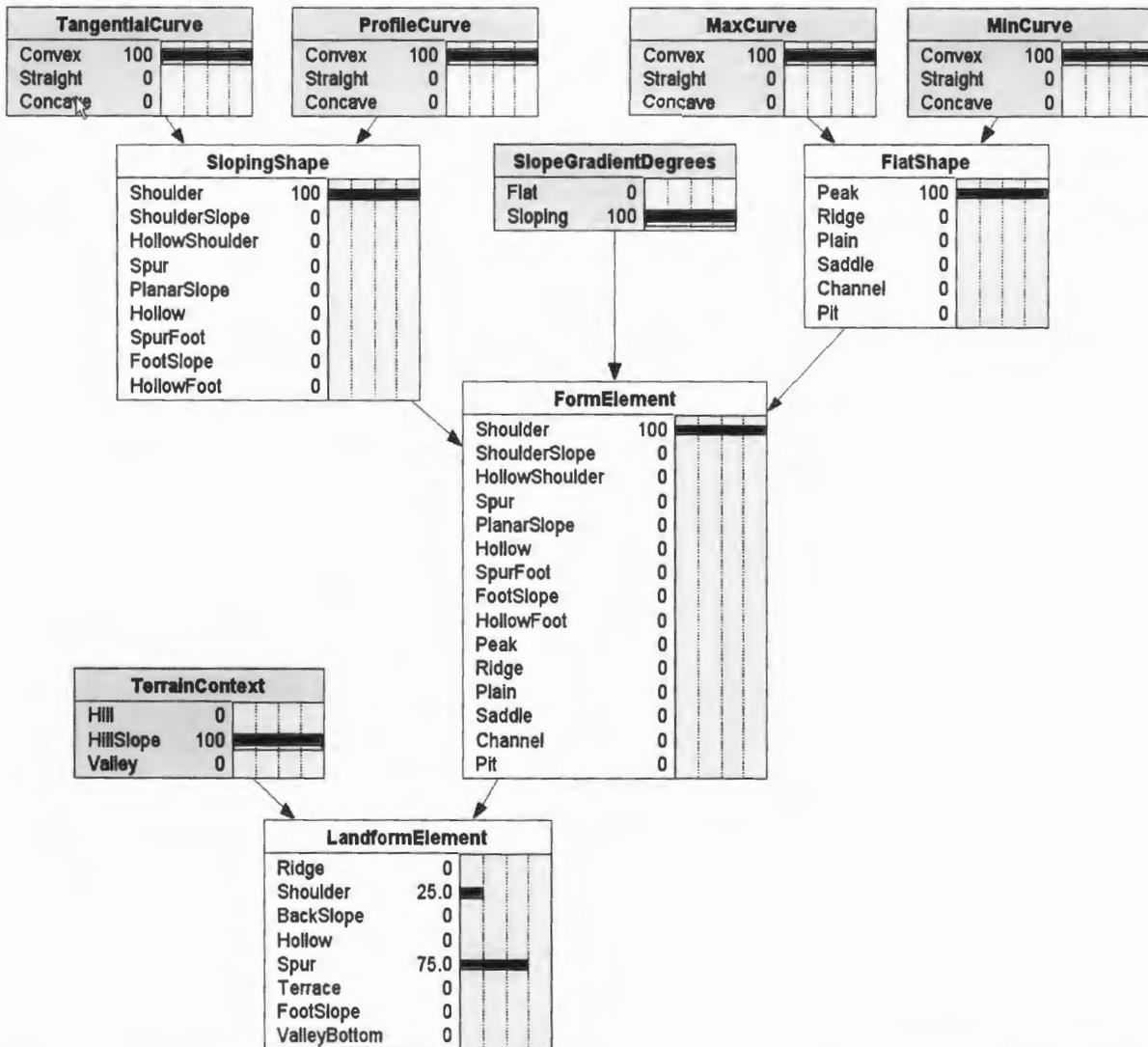


Figure 5 – The Bayesian network of continuous landform elements, which was developed in Netica version 2.17, was developed to predict the spatial arrangement of landform elements. In this diagram, the boxes represent nodes and the arrows represent the relationship between nodes. The input nodes, which represent GIS database layers and are not related to a parent node, were **TangentialCurve**, **ProfileCurve**, **MaxCurve**, **MinCurve**, **SlopeGradientDegrees** and **TerrainContext**. The output node, which is not related to a child node, was **LandformElement**. The remaining nodes, which were **SlopingShape**, **FlatShape**, and **FormElement**, were intermediate nodes. The conditional probability tables for the intermediate nodes and output nodes were defined using expert opinion and the logic of Schmidt and Hewitt (2004). For each node, the node name is provided at the top of the box with the node classes listed on the left. The thick horizontal black bars and preceding number indicate the probability, or belief, that a class occurs given the parent-node conditions. For example, given that the form-element node (**FormElement**) was 100% Shoulder and the terrain-context node (**TerrainContext**) was Hillslope, there was 25% probability that the landform element (**LandformElement**) was Shoulder and 75% probability that it was Spur.

Once each raster cell was assigned to either a sloping-form element or a flat-form element through the **FormElement** node, I assigned each cell a landform-element probability (**LandformElement**) based on terrain context (**TerrainContext**) and **FormElement**. The result was that each cell had a probability of belonging to each of the eight landform elements: Ridge, Shoulder, Backslope, Hollow, Spur, Terrace, Footslope and Valley Bottom. To produce these fuzzy landform-element probabilities I incorporated some uncertainty into the **LandformElement** node's conditional probabilities (Table 3). In Table 3 and Figure 5, for example, if a cell had a Shoulder **FormElement** and fell on a Hillslope in **TerrainContext**, then the cell was considered to have a 75% probability of being called a Spur **LandformElement** and a 25% probability of being called a Shoulder.

To capture the uncertainty in defining flat, slope or curved, I used the semantic-import model of Burrough (1989) for each of the input nodes. The semantic-import models allowed the BN inputs to reflect the class uncertainty, or fuzzy membership, resulting from either the continuous nature of the input or the varying semantic definitions of the input classes. Table 4 lists the thresholds I used to convert the continuous input values into soft classes. I heuristically determined these thresholds by visually assessing the spatial arrangement of each class relative to terrain. I ran the continuous-landform-element BN three times to capture the multiple spatial scales at which a semantically-defined landform element can exist. Each time I used a different search-window radius for curvature and slope gradient: 49 m, 74 m and 99 m. During all three runs, the search-window radius for the

Table 3 – The conditional-probability table used to define the relationship between landform elements, terrain context and form elements.

This table determines the probability of occurrence of each landform-element class given the input conditions of terrain context and form element. For example, if the **TerrainContext** was Hill and **FormElement** was Shoulder then there was a 60% probability that **LandformElement** node was Ridge and a 40% that it was Shoulder.

Terrain-Context	FormElement	LandformElements							
		Ridge	Shoulder	BackSlope	Hollow	Spur	Terrace	FootSlope	ValleyBottom
<u>Hill</u>	<u>Shoulder</u>	0.6	0.4	0	0	0	0	0	0
<u>Hill</u>	<u>ShoulderSlope</u>	0	1	0	0	0	0	0	0
<u>Hill</u>	<u>HollowShoulder</u>	0	0.6	0	0.4	0	0	0	0
<u>Hill</u>	<u>Spur</u>	0.6	0.2	0	0	0.2	0	0	0
<u>Hill</u>	<u>PlanarSlope</u>	0	0	1	0	0	0	0	0
<u>Hill</u>	<u>Hollow</u>	0	0	0	1	0	0	0	0
<u>Hill</u>	<u>SpurFoot</u>	0.6	0	0	0	0.4	0	0	0
<u>Hill</u>	<u>FootSlope</u>	0	0	1	0	0	0	0	0
<u>Hill</u>	<u>HollowFoot</u>	0	0	0.4	0.6	0	0	0	0
<u>Hill</u>	<u>Peak</u>	1	0	0	0	0	0	0	0
<u>Hill</u>	<u>Ridge</u>	1	0	0	0	0	0	0	0
<u>Hill</u>	<u>Plain</u>	1	0	0	0	0	0	0	0
<u>Hill</u>	<u>Saddle</u>	0.5	0	0	0.5	0	0	0	0
<u>Hill</u>	<u>Channel</u>	0	0	0	1	0	0	0	0
<u>Hill</u>	<u>Pit</u>	0	0	0	1	0	0	0	0
<u>HillSlope</u>	<u>Shoulder</u>	0	0.25	0	0	0.75	0	0	0
<u>HillSlope</u>	<u>ShoulderSlope</u>	0	0.25	0.75	0	0	0	0	0
<u>HillSlope</u>	<u>HollowShoulder</u>	0	0.25	0	0.75	0	0	0	0
<u>HillSlope</u>	<u>Spur</u>	0	0	0	0	1	0	0	0
<u>HillSlope</u>	<u>PlanarSlope</u>	0	0	1	0	0	0	0	0
<u>HillSlope</u>	<u>Hollow</u>	0	0	0	1	0	0	0	0
<u>HillSlope</u>	<u>SpurFoot</u>	0	0	0	0	0.75	0	0.25	0
<u>HillSlope</u>	<u>FootSlope</u>	0	0	0.75	0	0	0	0.25	0
<u>HillSlope</u>	<u>HollowFoot</u>	0	0	0	0.75	0	0	0.25	0
<u>HillSlope</u>	<u>Peak</u>	0.25	0	0	0	0.75	0	0	0
<u>HillSlope</u>	<u>Ridge</u>	0.25	0	0	0	0.75	0	0	0
<u>HillSlope</u>	<u>Plain</u>	0	0	0	0	0	1	0	0
<u>HillSlope</u>	<u>Saddle</u>	0	0	0	0.5	0	0.5	0	0
<u>HillSlope</u>	<u>Channel</u>	0	0	0	1	0	0	0	0
<u>HillSlope</u>	<u>Pit</u>	0	0	0	0.5	0	0.5	0	0
<u>Valley</u>	<u>Shoulder</u>	0	0	0	0	0.75	0	0.25	0
<u>Valley</u>	<u>ShoulderSlope</u>	0	0	0	0	0.5	0	0.5	0
<u>Valley</u>	<u>HollowShoulder</u>	0	0	0	0.35	0.35	0	0.3	0
<u>Valley</u>	<u>Spur</u>	0	0	0	0	0.5	0	0.5	0
<u>Valley</u>	<u>PlanarSlope</u>	0	0	0.5	0	0	0	0.5	0
<u>Valley</u>	<u>Hollow</u>	0	0	0	0.35	0	0	0.65	0
<u>Valley</u>	<u>SpurFoot</u>	0	0	0	0	0.25	0	0.75	0
<u>Valley</u>	<u>FootSlope</u>	0	0	0	0	0	0	1	0
<u>Valley</u>	<u>HollowFoot</u>	0	0	0	0.25	0	0	0.75	0
<u>Valley</u>	<u>Peak</u>	0	0	0	0	0.1	0	0	0.9
<u>Valley</u>	<u>Ridge</u>	0	0	0	0	0.1	0	0	0.9
<u>Valley</u>	<u>Plain</u>	0	0	0	0	0	0	0	1
<u>Valley</u>	<u>Saddle</u>	0	0	0	0.1	0.1	0	0	0.8
<u>Valley</u>	<u>Channel</u>	0	0	0	0	0	0	0	1
<u>Valley</u>	<u>Pit</u>	0	0	0	0	0	0	0	1

terrain context node remained fixed at 1875 m. My choice of search-window radii, which attempted to capture the scale that site series were influenced by local terrain geometry, was based on the advice and opinion of the ecological-domain expert. I created the final layer of continuous landform elements by taking the arithmetic mean from the output of the three continuous-landform-element runs.

Table 4 – The heuristic rule-base used to convert input values into fuzzy landform attributes. For example, slope gradients less than 4° had full membership to the Flat class. Membership to the Flat class decreased linearly to zero between 4° and 6° slope-gradient. Above 6° slope-gradient there was no membership to the Flat class.

Node	Class	Threshold	
		Full Membership	No Membership
Curvature (Tangential, Profile, Minimum and Maximum)	Concave	< -0.002	> -0.0004
	Straight	> -0.0004 and < 0.0004	< -0.002 and > 0.002
	Convex	> 0.002	< 0.0004
SlopeGradientDegree	Flat	< 4	> 6
	Sloping	> 6	< 4
TerrainContext	Valley	< -0.6	> -0.3
	Hillslope or Plain	> -0.3 and < 0	< -0.6 and > 0.4
	Hill	> 0.4	< 0

I combined all of the separate input GIS layers into a single GIS database using a raster data model. The raster-cell dimensions were matched to the SELES raster dimensions, which were 25 x 25 m. I converted the vector data sets of bedrock, stream density, forest-stand height, forest-stand age, forest-stand crown-closure and forest-stand species to raster using a majority rule. With a majority rule, the value of a raster cell reflects the vector polygon that occupies the largest portion of the cell. Table 5 summarizes the GIS database's contents along with comments regarding each layer's accuracy as it may pertain towards predicting site series. Figure 6 illustrates the varying spatial scales of the GIS input layers.

Table 5 – A list of the data layers contained within the GIS database, along with comments regarding each layer’s ability to assist site series prediction.

<b>GIS Layer</b>	<b>Confidence in the GIS layer’s spatial and thematic content with respect to the needs of site series prediction</b>
Ridge	High for sharp ridges
Hill Top	Medium – identifies some but not all
Gully	High
River Floodplains	Low
Lake/Wetland Benches	Low
Slope Gradient	High
Bedrock	Low
Stream Density	Medium
Ocean Wind	Medium
Shoreline	High
Toe Slope	Low
SELES Floodplains	Medium
SELES Alluvial Fans	Low
SELES Soil Moisture	Medium in high relief areas, low in low relief areas
Continuous Landform Elements	Medium
Forest Tree Age	High (but low for old age trees)
Forest Tree Height	High
Forest Tree Crown Closure	High
Forest Tree Species	High

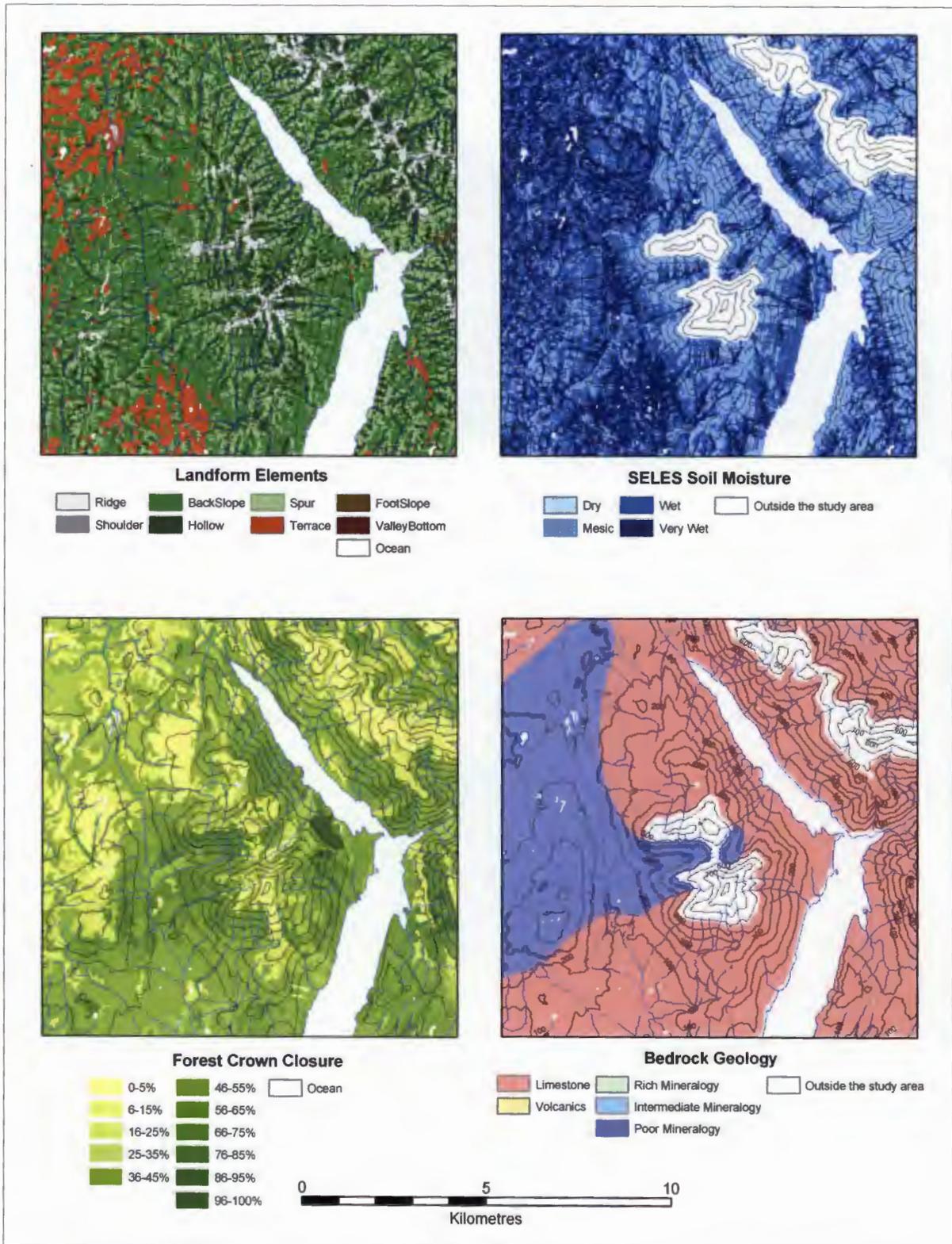


Figure 6 – Four examples of the GIS data layers used to create the database used in this study. This diagram highlights the different spatial scales that occurred within the GIS database. For example, the landform-elements layer has a large-spatial scale and the bedrock-geology layer has a small-spatial scale.

contents along with comments regarding each layer's accuracy as it may pertain towards predicting site series. Figure 6 illustrates the varying spatial scales of the GIS input layers.

Slope-aspect is considered an important ecological factor in relation to site series in some area of the province, however, the ecological domain expert did not consider slope-aspect ecologically important within the study area. The high latitude and the presence of heavy cloud cover for most of the year effectively removes the influence that sun-incidence angle can have on soil moisture within the study area.

## 2.4 KNOWLEDGE BASE CONSTRUCTION

Similar to the traditional approach of predictive site-series mapping (RIC 1999), I first stratified the study area using the BC Ministry of Forest's medium-scale subzone/variant mapping of the biogeoclimatic ecosystem classification. I constructed a separate site-series knowledge base for each of the three resulting strata (CWHvm2, CWHvm1 and CWHvh2) using the advice and knowledge of local ecological-domain expert, Del Meidinger.

A review of the input GIS database contents revealed that, based on the ecological-domain expert's opinion, the spatial and thematic accuracy of the GIS database was not sufficient to accurately map all of the site-series units. The solution was to group similar site series into site-series groups. Although not desirable, site-series grouping is not uncommon in predictive site-series mapping (Timberline Forest Inventory Consultants Ltd. 2000; Jones 2003; Jones 2004). Grouping site series reduced the total number of mapped units from 42 down to 23, and resulted in six site-series groups for CWHvm2 (Table 6), seven site-series groups for CWHvm1 (Table 7), and 10 site-series groups for CWHvh2 (Table 8).

Table 6 – Site-series groups used for the CWHvm2 BEC unit. CWHvm2 site-series units of Meidinger and Pojar (1991) were grouped into simpler units due to the low thematic and spatial accuracy of some GIS layers.

Original Site Series	Final Site-Series Group
01 – HwBa – Blueberry 06 – HwBa – Deer fern	01
02 – HwPl – Cladina 03 – HwCw – Salal	03
04 – CwHw – Sword fern 05 – BaCw – Foamflower 08 – BaSa – Devil’s club	05
09 – CwYc – Goldthread (Bog forest)	09
10 – Pl – Sphagnum (Bog woodland)	10
11 – CwSs – Skunk cabbage (Swamp forest)	11

Table 7 – Site-series groups used for the CWHvm1 BEC unit. CWHvm1 site-series units of Meidinger and Pojar (1991) were grouped into simpler units due to the low thematic and spatial accuracy of some GIS layers.

Original Site Series	Final Site-Series Group
01 – HwBa – Blueberry 06 – HwBa – Deer fern	01
02 – HwPl – Cladina 03 – HwCw – Salal	03
04 – CwHw – Sword fern 05 – BaCw – Foamflower 08 – BaSa – Devil’s club	05
09 – Ss – Salmonberry (high fluvial bench) 10 – Act – Red-osier dogwood (middle fluvial bench) 11 – Act – Willow (low fluvial benches)	09
12 -- CwYc – Goldthread (bog forest)	12
13 – Pl – Sphagnum (bog woodland)	13
14 – CwSs – Skunk cabbage (swamp forest)	14

Table 8 – Site-series groups used for the CWHvh2 BEC unit. CWHvh2 site-series units of Meidinger and Pojar (1991) were grouped into simpler units due to the low thematic and spatial accuracy of some GIS layers.

Original Site Series	Final Site Series-Group
01 – CwHw – Salal	01
02 – PIYc – Rhacomitrium	03
03 – CwYc – Salal	
04 – HwSs – Lanky moss	04
05 – CwSs – Sword fern	07
06 – CwSs – Foamflower	
07 – CwSs – Devil’s club	
08 – Ss – Lily-of-the-valley (high fluvial benches)	08
09 – Ss – Trisetum (middle fluvial benches)	
10 – Dr – Lily-of-the-valley (low fluvial benches)	
11 – CwYc – Goldthread (bog forest)	11
12 – PIYc – Sphagnum (bog woodland)	12
13 – CwSs – Skunk cabbage (swamp forest)	13
14 – Ss – Salal (Rock headlands and beach plains)	14
15 – Ss – Kindbergia (old beach plains)	
16 – Ss – Reedgrass (rocky headlands, colluvium, and old dunes)	
17 – Ss – Swordfern (marine terraces and scarps)	
18 – Ss – Slough sedge (strongly fluctuating brackish water table)	18
19 – Ss – Pacific crab apple (sites affected by brackish water)	

### *Predicting soil moisture and soil nutrient regimes*

My first step in creating the BN knowledge bases was to predict soil-moisture and soil-nutrient regimes using the variables in the GIS database. I developed a preliminary influence diagram using the ecological-domain expert's opinion on the relationships between soil-moisture regime, soil-nutrient regime and the layers in the GIS database (Figure 7 and Figure 8). To turn the influence diagram into a BN, I added classes to each of the variables using the ecological-domain expert's advice. I also merged the two influence diagrams into one BN and allowed the soil-moisture and soil-nutrient sub-models to share common variables (Figure 9). I populated each of the conditional probability tables using the ecological-domain expert's advice and supplemented with information published in the site identification field guide (Banner *et al.* 1993).

In the soil-moisture-regime sub-model (Figure 7) I used two environmental factors that, based on the advice of the ecological-domain expert, were believed to be related to soil-moisture regime: terrain shape, and surface-visible waters. I estimated the component of soil moisture that was related to terrain shape through the **Ground Wetness** intermediate node, while the component of soil moisture that was related to surface-visible water was estimated through the **Surface Wetness** intermediate node.

To predict soil wetness related to terrain shape (**Ground Wetness**), I used a combination of variables that described terrain shape or position (**Landform Elements, Hill, Ridge, SELES Floodplain, Lake/Wetland Bench, River Bench, Toe Slope and Gully**) and SELES soil moisture (**Flow Accumulation Wetness**). I divided the landscape into areas that

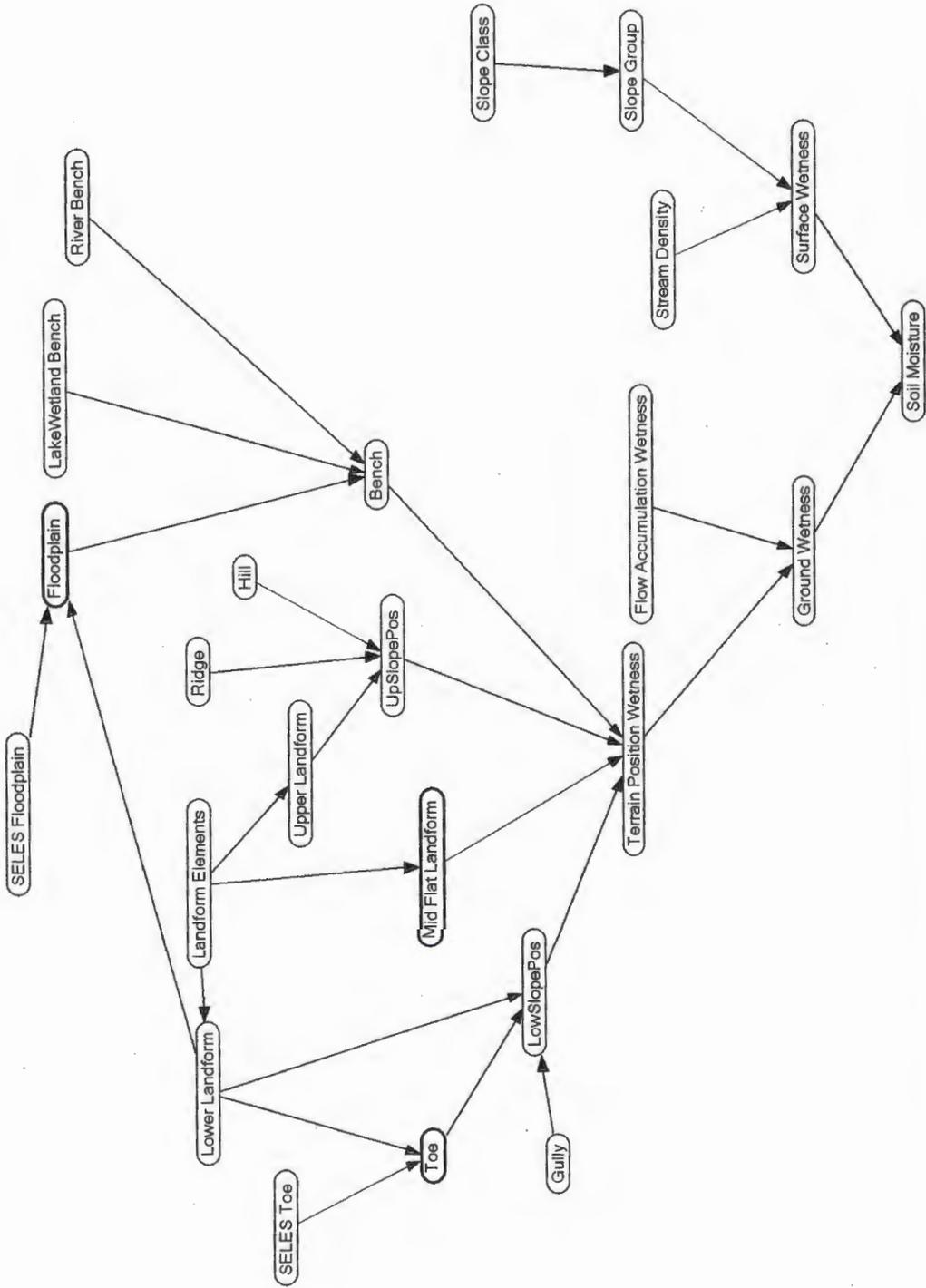


Figure 7 – The influence diagram created to explain the relationship between the GIS database used in this study and soil-moisture regime. Like the description of the landform-elements Bayesian-network in Figure 5, the input nodes represent the GIS data layers and the intermediate nodes help by simplifying the relationship between input nodes and output nodes by grouping related nodes. Arrows represent the relationships between nodes.

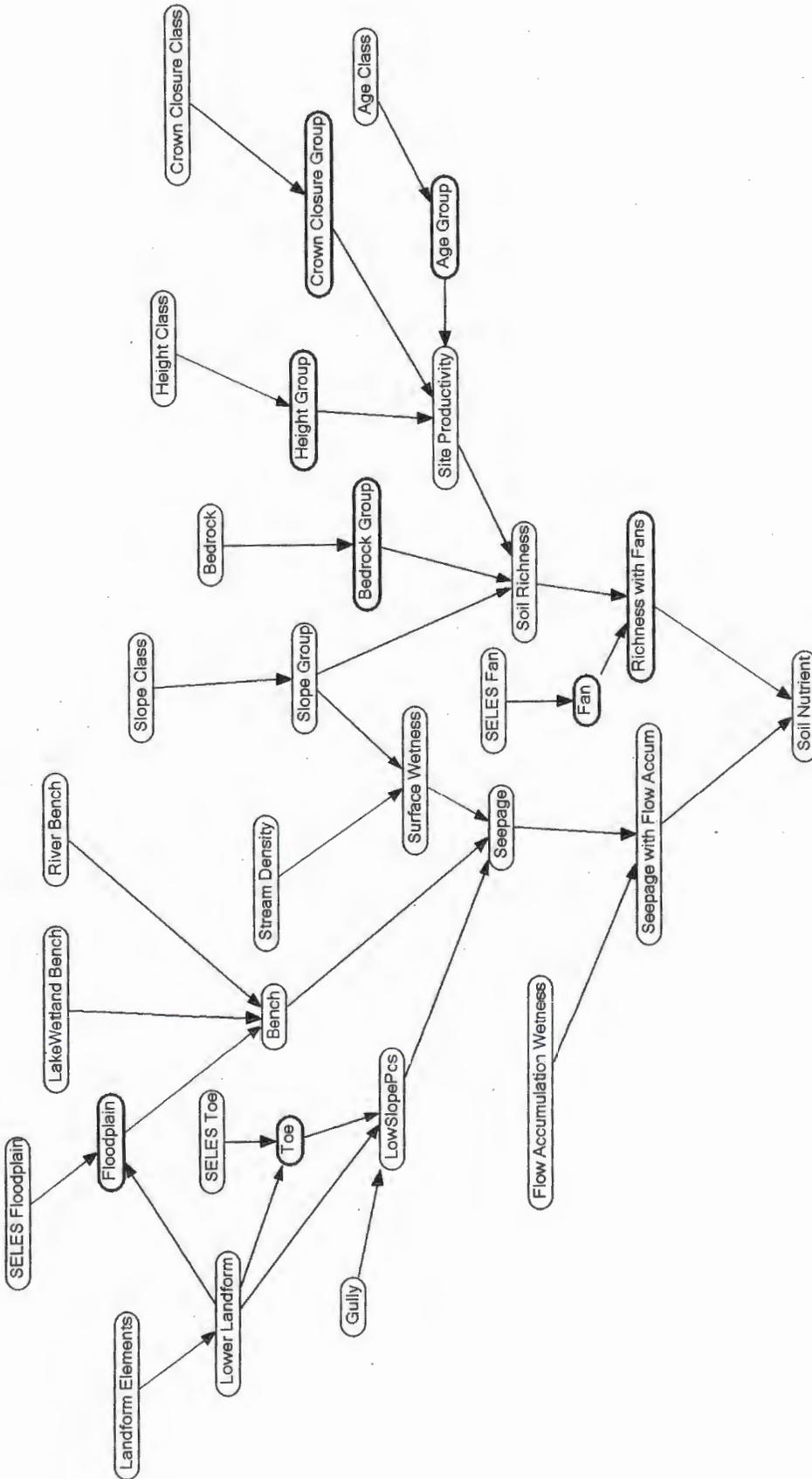


Figure 8 – The influence diagram created to explain the relationship between the GIS database used in this study and soil-nutrient regime. Like the description of the landform-elements Bayesian-network in Figure 5, the input nodes represent the GIS data layers and the intermediate nodes help by simplifying the relationship between input nodes and output nodes by grouping related nodes.



I considered water shedding (**UpSlopePos** - upland areas where water is moving out quicker than being replenished), water accumulating (**LowSlopePos** - lowland areas where water is moving out slower than being replenished), benches adjacent to water bodies (**Bench**), and areas between upland and lowland (**Mid Flat Landform**). For each combination of upper-slope position, mid-slope position, lower-slope position and bench, I assigned a conditional probability of belonging to each class in the **Terrain Position Wetness** intermediate node. To produce **Ground Wetness**, I created a conditional-probability table, which took into account the **Terrain Position Wetness** intermediate node and the SELES' soil moisture input node (**Flow Accumulation Wetness**).

Water visible on the earth's surface, such as streams and rivers, was the other environmental factor that the ecological-domain expert and I believed reflected soil moisture. I estimated surface-visible soil moisture (**Surface Wetness**) using local stream density combined with slope gradient, where higher stream density was assumed to signify greater soil moisture. When comparing areas of equal stream density, however, areas of high slope would be drier than areas of low slope.

To estimate soil-nutrient regime (Figure 8) I used two environmental factors that, based on the ecological-domain expert's opinion, were believed to influence soil-nutrient regime. These environmental factors were: soil seepage (**Seepage with Flow Accum**) and soil richness (**Soil Richness**). Seepage was used as an influence on soil richness because it was believed that increased seepage contributes to increased soil richness.

I derived the seepage intermediate node (**Seepage with Flow Accum**) from lowland, bench, high stream-density and high flow-accumulation areas because, based on the advice of

the ecological-domain expert, high soil-moisture levels affect seepage rates. I used two stages to estimate seepage rates: first seepage (**Seepage**) was estimated from the intermediate nodes representing low-terrain position (**LowSlopePos**), fluvial benches (**Bench**) and surface-visible soil-moisture (**Surface Wetness**). This estimate of seepage was subsequently modified by the SELES estimate of soil moisture (**Flow Accumulation Wetness**) to produce a final seepage estimate (**Seepage with Flow Accum**).

The other influence on soil-nutrient regime was soil richness. Based on the opinion of the ecological-domain expert, soil richness was related to four environmental variables: alluvial fans, slope gradient, underlying bedrock and forest. I created an initial soil-richness intermediate node (**Soil Richness**) using slope (**Slope Group**), bedrock geology (**Bedrock Group**) and forest productivity (**Site Productivity**). The ecological-domain expert defined the conditional probabilities for this initial soil-richness node. Based on the ecological-domain expert's opinion, forest productivity was given far greater influence on the conditional probabilities of soil richness than bedrock geology or slope. Slope was considered to have only a slight influence on soil richness, and bedrock geology was mapped at such a coarse scale that confidence in its predictions was low. Generally, areas of high forest productivity were considered by the ecological-domain expert to have high soil richness, and areas of low forest productivity were considered to have low soil richness.

Forest productivity was seen by the ecological-domain expert as an important effect of soil richness. Unlike any of the previous intermediate nodes, however, which relied on terrain variables, the estimate of forest productivity (**Site Productivity**) relied solely on the MSRM VRI forest vegetation inventory. Using classes of forest-stand age (**Age Group**), height (**Height Group**) and crown closure (**Crown Closure Group**), the ecological-domain

expert assigned each class combination a probability of being high (SP\_High), medium (SP\_Medium) or low productivity (SP\_Low). The ecological-domain expert's belief was that forested areas with short, old-aged trees with open crown-closure indicated areas of low productivity, while areas with medium or old-aged tall trees with closed crowns indicated areas of high productivity.

Once the initial soil-richness node (**Soil Richness**) was defined, I added the influence of alluvial fans (**Fans**) on soil richness to produce a final soil-richness node (**Richness with Fans**). Although occurring infrequently across the landscape, alluvial fans were seen by the ecological-domain expert as an important indicator of rich soils. To reflect this belief, I created the conditional-probability table for **Richness with Fans** with increased probability of rich soils occurring on sites with alluvial fans.

Once I built the BN knowledge base to predict soil-moisture and soil-nutrient regimes, I adjusted the conditional-probability tables to reflect the unique attributes of each BEC subzone/variant. In the CWHvm1 and CWHvm2 BEC units, I considered the soil transmissivity to be uniform across the landscape. In the CWHvh2, however, many of the areas of low slope-gradient hold moisture due to their low soil-transmissivity. To reflect this difference in soil transmissivity between the hypermaritime BEC subzone/variants and the maritime BEC subzone/variants, I adjusted the implementation of the SINMAP algorithm in SELES to increase the influence of low slope gradients on soil moisture. This adjustment allowed the SELES flow-accumulation GIS layers to reflect the ecological-domain expert's belief that low slope-gradients within the CWHvh2 BEC unit were much wetter than low slopes in the either the CWHvm1 or CWHvm2 BEC units.

In addition to the changes to the soil-moisture algorithm in the CWHvh2, I also adjusted two of the nodes required to predict soil nutrient in the CWHvh2. I gave the **Crown Closure Group** and the **Site Productivity** intermediate nodes an additional class to distinguish sites with very sparse crown closure.

*Adding a node to reflect the edatopic grid*

Once I developed a BN knowledge base for each BEC subzone/variant to reflect the ecological-domain expert's initial beliefs around soil moisture and soil nutrient, I used the domain expert's advice and the edatopic grids (Figure 10) to combine soil-moisture and soil-nutrient regimes into an initial site-series-group prediction (**Site Series Edatopic**) (Figure 11). Although I used soil-moisture and soil-nutrient regimes to place each location in the appropriate general area on the edatopic grid, the ecological-domain expert recognized that other variables were required to adjust site-series prediction. Adjustments were necessary either because of inaccuracies in the soil-moisture and soil-nutrient sections of the model, or because more than one site-series group could occur at a given edatopic-grid position. Based on advice from the ecological-domain expert I used such variables as vegetation, slope, floodplain and shoreline to refine the site-series prediction. For each BEC subzone/variant, the BN knowledge base required different adjustments to the post-edatopic site-series prediction depending on the unique environmental conditions of each BEC subzone/variant.

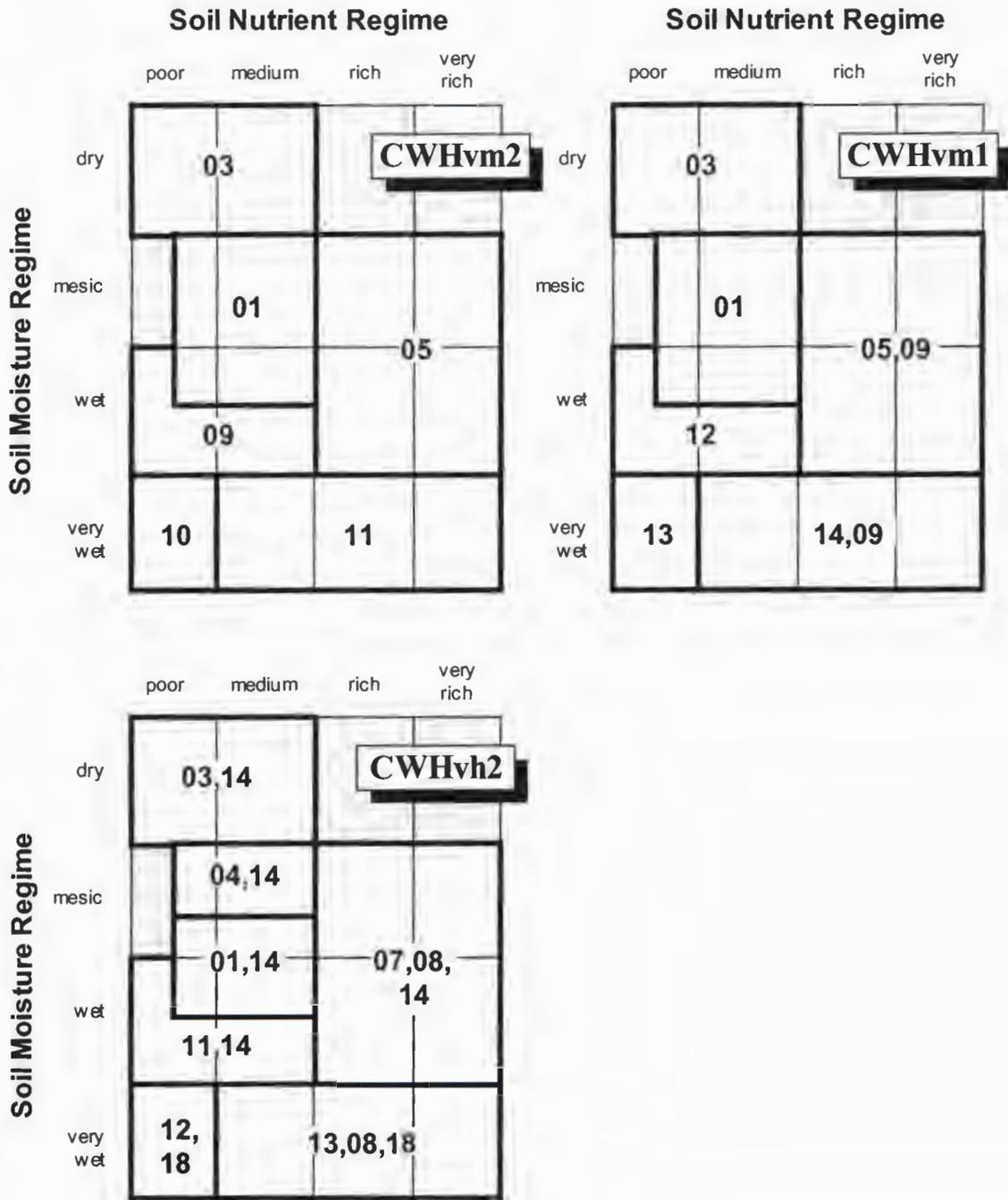


Figure 10 – Edatopic grids and corresponding site-series groups used for the CWHvm1, CWHvm2 and CWHvh2 Bayesian networks. Some areas of the edatopic grids were comprised of only one site-series group. For example, only site-series-group 10 in the CWHvm2 was comprised of poor soil-nutrient and very wet soil-moisture. Some areas of the edatopic grids, however, were comprised of more than one site-series group. For example, site-series groups 12 and 18 in the CWHvh2 were both comprised of poor soil nutrient and very wet soil moisture.

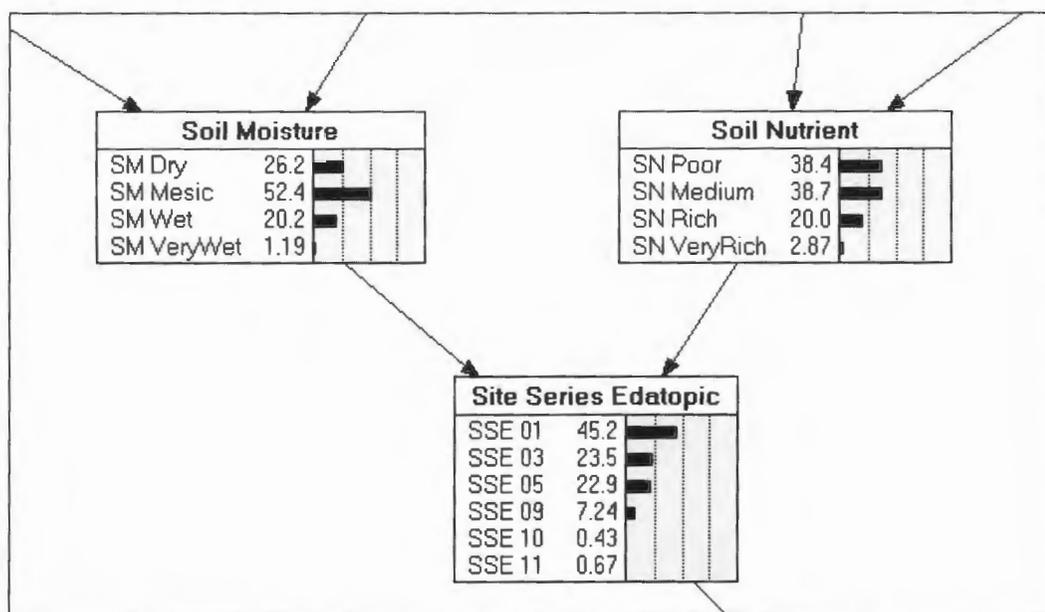


Figure 11 – The portion of the CWHvm2 Bayesian-network, which illustrates the relationship between site-series group (**Site Series Edatopic**), soil-moisture regime (**Soil Moisture**) and soil-nutrient regime (**Soil Nutrient**). A description of the node format and nomenclature can be found in Figure 5.

#### *Adding nodes to adjust the estimation of site-series group*

The BN knowledge base for the CWHvm2 required the fewest post-edatopic adjustments to site-series group. Based on the ecological-domain expert's advice, I incorporated slope gradient (**Slope group2**) (Figure 12) and forest vegetation (**Site Productivity** and **Species Influence**) (Figure 13) directly into the site-series-group prediction. Slope gradient was used to express the ecological-domain expert's belief that the 05 site-series group occurred on sites with rich soil and moderate to high slope gradient, and the 11 site-series group occurred on sites with rich soil and little to no slope gradient. In addition, slope was used to differentiate site-series groups with poor/medium soils, such as 03 that occurred on sites with high slope gradient, from 09 and 10 that occurred on low slope gradients. The influence of forest vegetation on the prediction of site-series group was

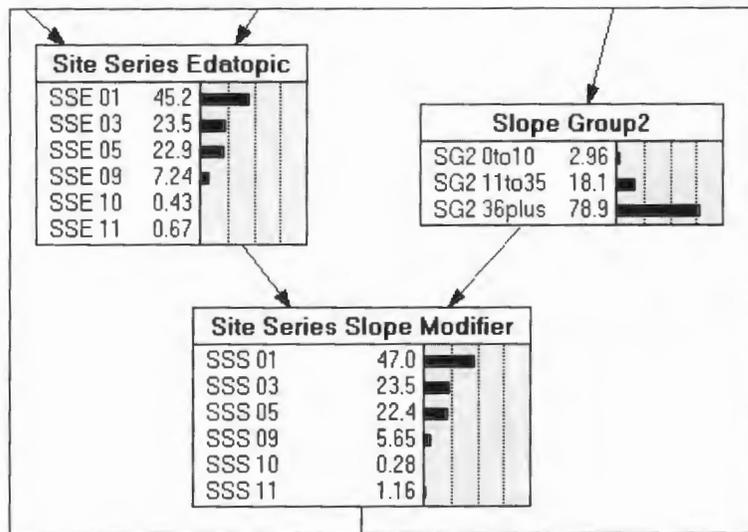


Figure 12 – The portion of the Bayesian network for the CWHvm2, which illustrates the influence of slope (**Slope Group2**) on the previous estimation of site-series-group (**Site Series Edatopic**) to produce a new prediction of site-series group (**Site Series Slope Modifier**). A description of the node format and nomenclature can be found in Figure 5.

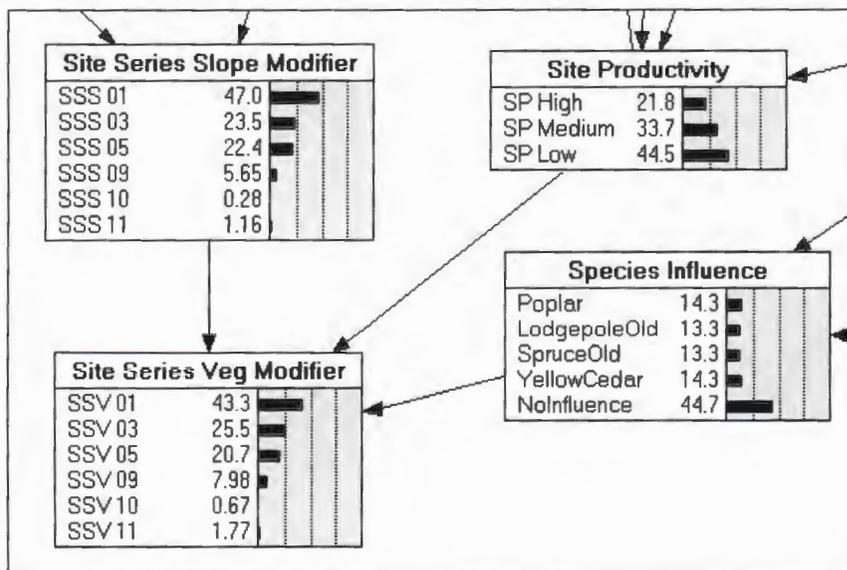


Figure 13 – The portion of the Bayesian network for the CWHvm2, which illustrates the influence of forest-stand productivity (**Site Productivity**) and forest-stand species (**Species Influence**) on the previous estimation of site-series group (**Site Series Slope Modifier**) to produce a final estimation of site-series group (**Site Series Veg Modifier**) for the CWHvm2. A description of the node format and nomenclature can be found in Figure 5.

more complex. In general, the presence of old and low-productivity lodgepole pine (*Pinus contorta* var. *contorta*) was believed by the ecological-domain expert to often occur on 03, 09 and 10 site-series groups. Sites with highly productive lodgepole pine were more often on 01 site-series group. Sites that were highly productive and contained Sitka spruce (*Picea sitchensis*) were more likely to be 05 site-series group, and the low-productivity sites with Sitka spruce were more likely to be 11. The presence of highly productive yellow-cedar usually signified 01 site-series group.

The BN knowledge base for the CWHvm1 incorporated similar beliefs as the CWHvm2 with respect to the probability that certain site-series groups were associated with certain slope gradients and forest vegetation. Tree species was used similarly in the CWHvm1 as in the CWHvm2, but, based on the ecological-domain expert's opinion, I added amabilis fir, poplar (*Populus balsamifera*) and red alder (*Alnus rubra*) to the list of vegetation species. It was the ecological-domain expert's opinion that both poplar and red alder helped to predict the occurrence of 09 site-series group. The 09 site-series group occurred on similar slopes and soil-moisture/nutrient regimes as the 14 site-series group, but was associated with fluvial benches that were commonly vegetated with mature red alder and poplar.

In the BN knowledge base of the CWHvm1, I incorporated river floodplains (**Floodplain**) directly into the estimation of site-series group to further refine the prediction of site-series group 09 (Figure 14). I gave low floodplains (LowerFloodplain) a high probability of being 09 site-series group, high floodplains (HighFloodplain) a medium probability of being 09, and glacial floodplains (GlacialFloodplain) a low probability of being 09.

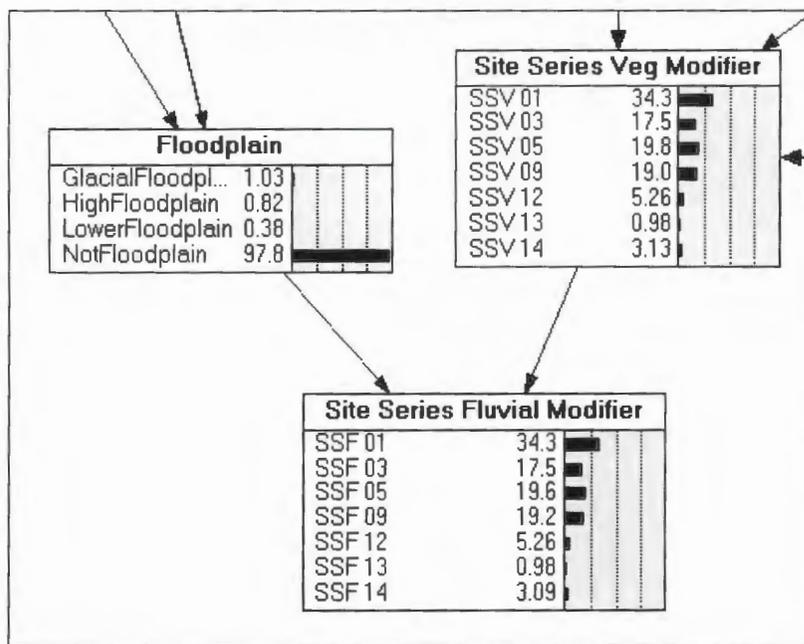


Figure 14 – The portion of the Bayesian network for the CWHvm1, which illustrates the influence of floodplains (**Floodplain**) on the previous estimate of site-series group (**Site Series Veg Modifier**) to produce a final estimate of site-series group (**Site Series Fluvial Modifier**) for the CWHvm1. A description of the node format and nomenclature can be found in Figure 5.

I developed the BN knowledge base of the CWHvh2 to incorporate slope gradient, vegetation and floodplains variables directly into the site-series-group prediction. I defined conditional-probability tables for these relationships using the ecological-domain expert's opinion. I also incorporated shoreline effects directly into the prediction of site-series group (Figure 15). I converted to site-series group 14, all site-series groups occurring in areas affected by the ocean wind input layer (**Salt Spray Shoreline**). In addition, based on the ecological-domain expert's advice and the CWHvh2 edatopic grid (Figure 10), I increased the probability that a site was site-series group 18 if it occurred on an existing very wet site (site-series groups 12 and 13) in a gully adjacent to the ocean (**Gully Adjacent to Shoreline**).

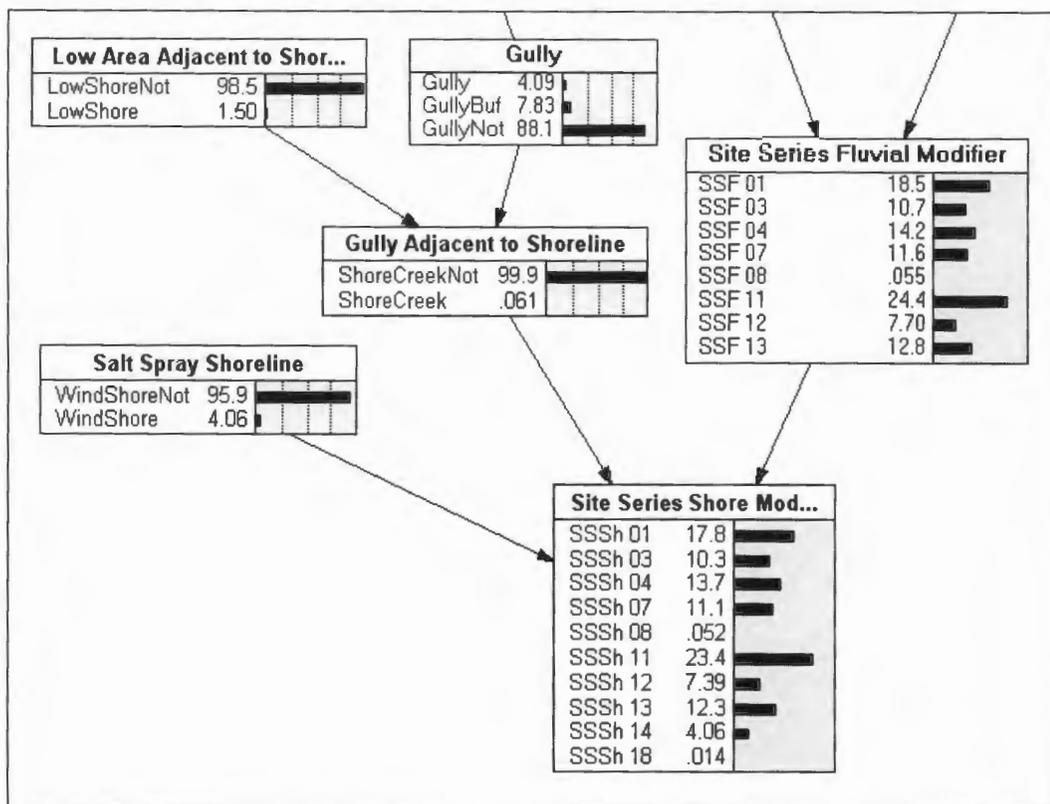


Figure 15 – The Bayesian network for the CWHvh2, which illustrates the influence of ocean wind (**Salt Spray Shoreline**) and ocean estuaries (**Gully Adjacent to Shore**) on the previous estimation of site-series group (**Site Series Fluvial Modifier**) to produce a final estimation of site-series group (**Site Series Shore Modifier**) for the CWHvh2. A description of the node format and nomenclature can be found in Figure 5.

### *Adjusting the knowledge base conditional probabilities*

I populated each input node with an unconditional-probability distribution, which represented the proportion that each input class occurred within the corresponding BEC subzone/variant. These unconditional probabilities aided model adjustment because the posterior distributions generated by each of the intermediate nodes could then reflect the relative abundance of each class within the BEC subzone/variant (see the “probability bars” of each node in Figures 7, 8, 9, 10, 11, 12, and 13). If the posterior distribution of an intermediate node did not reflect the relative distribution of the classes within the study area then this signalled the possible need to revise the node’s conditional-probability table.

For example, in Figure 9, which contains the unconditional probabilities for the CWHvm2 BEC unit, 0.28% of the area has slopes of 0% (S 0), 2.68% of the area has slopes from 1% to 10% (S 1to10), 18.1% of the area has slopes from 11% to 35% (S 11to35), 34.4% of the area has slope from 36% to 60% (S 36to60), 22.5% of the area has slopes from 61% to 80% (S 61to80), and 22% of the area has slopes greater than 80% (S 81plus). Based on the unconditional probabilities of **Slope Class** and the conditional probabilities of **Slope Group**, the posterior probabilities of the **Slope Group** classes were 12% Gentle slope and 88% Very Steep slope. If the posterior-probability values for **Slope Group** did not reflect expert opinion then I would adjust the conditional-probability values for **Slope Group** in consultation with ecological-domain expert.

Although the posterior probabilities that resulted from the BEC subzone/variant's unconditional-probability tables reflected the relative abundance of each class within the BEC subzone/variant, they did not portray the spatial arrangement of each class. To assess the spatial distribution of the intermediate- and leaf-nodes' posterior-probabilities, I routinely generated maps using a sample of the GIS database as input. The ecological-domain expert and I visually interpreted the maps generated from intermediate- and leaf-nodes' posterior-probabilities to assess how well the knowledge bases were predicting the spatial distribution of the intermediate- and leaf-nodes' classes. If the spatial arrangement of a node's classes did not reflect the ecological-domain expert's expectation then I revised the conditional probabilities using the advice of the ecological-domain expert.

### *Testing the knowledge base*

I tested the spatial output from the knowledge bases against spatially explicit reference data once the knowledge base appeared to produce predictions that spatially matched the expectation of the ecological-domain expert. To test the knowledge bases of the CWHvm1 and CWHvm2, I randomly selected 68 raster cells from an area in which I had stereo air-photo coverage. For each randomly located point, I identified the corresponding location on the air photos. Using air-photo interpretation, the ecological-domain expert estimated the site-series group occurring at each randomly selected point. For the CWHvh2, I created 32 randomly distributed points, but I augmented these with 59 site-series field plots gathered for a previous project conducted in the CWHvh2 (Banner *et al.* 2004). These field plots occurred in two spatially distinct clusters, with each cluster employing a grid-sampling plan ranging from 100 m to 200 m spacing.

By comparing the knowledge base estimation of site-series group with the air-photo-interpreted and field-gathered predictions it was possible to assess how well the knowledge base predicted site-series group. I conducted the comparison between the map's site-series group and the site-series group interpreted from air-photos using a simple hard classification. For each sample point, I chose the site-series group with the highest probability from the map and the dominant site-series group from the photo. To rate knowledge base performance I simply tallied the number of times that the highest-probability site-series-group prediction matched the dominant air-photo interpretation. If the knowledge base correctly predicted site-series group greater than 65% of the time then I made no additional adjustments to the knowledge base. If the testing results were low, however, I did make additional adjustments to the conditional-probability table. I based these additional adjustments on both evidence

resulting from the test analysis and further visual analysis. All of the conditional-probability adjustments I made were reviewed by the ecological-domain expert prior to further map generation.

I repeated this process of adjusting the conditional-probability values, making visual assessments, and testing against reference data until the adjustments, believed to be required to increase accuracy, were no longer resulting in increased prediction success. Visual assessment of the preliminary results from the knowledge bases for the CWHvm2 and CWHvm1 revealed that the conditional-probability values required only minor adjustment. After the first round of testing using the testing reference data, the CWHvm2 knowledge base required no adjustments to the conditional-probability tables, however, the CWHvm1 required some adjustments to increase the successful prediction of the 09 site-series group. Final testing of knowledge bases for the CWHvm2 and CHWvm1 indicated that the CWHvm2 knowledge base achieved a 92% successful prediction rate for the 14 air-photo reference points, while the CWHvm1 knowledge base correctly predicted 72% of the 54 air-photo reference points.

The CWHvh2 knowledge base, however, required extensive adjustments due to either visual assessment or the results generated from the testing procedure. In total the CWHvh2 was tested and adjusted three times, but never achieved the 65% successful prediction rate. At best, the CWHvh2 knowledge base was successful in predicting site-series group on 59% of the test sites. At this point, I used all three BN knowledge bases to generate a final site-series-group map for each of the three BEC subzone/variants in the study area.

## 2.5 FINAL MAP GENERATION

I created a soft-classification map of site-series group from the input GIS database by passing each individual cell in the input database through the appropriate CWHvh2, CWHvm1 and CWHvm2 knowledge base. The resulting map of fuzzy site-series groups recorded the probability of site-series-group occurrence within each raster cell. I created a hard classification of site-series group by choosing the site-series group with the highest probability within each raster cell.

## 2.6 MAP ACCURACY ASSESSMENT

To assess map accuracy, I compared reference data, which were comprised of site-series-group information gathered from georeferenced field observations, to geographically equivalent areas on the map of predicted site-series group. In total, I used 286 reference points for accuracy assessment: 92 in the CWHvm2, 161 in the CWHvm1, and 33 in the CWHvh2. These field observations were gathered for the North Coast PEM Project (Jones 2003) using a cluster-sampling plan.

For each point in the reference data, I compared the site-series group in the reference data to the site-series-group prediction at the same location on the map. I used a simple hard classification for this comparison. For each point in the reference data, I chose the dominant site-series group, and for each point on the map, I chose the site-series group with the highest probability. I constructed a contingency table from these map/reference-data comparisons for each BEC subzone/variant. From these contingency tables, I generated several accuracy-assessment measures: percentage correct, producer's accuracy, user's accuracy and Kappa

analysis. I used an  $\alpha = 0.05$  level of significance throughout these test for inferring statistical significance.

In conjunction with each contingency matrix, I constructed a separate matrix that recorded the average posterior-probability from the site-series group with the highest probability. The content of this matrix revealed the knowledge base's confidence when predicting the most-probable site-series group. For example, if the contingency matrix records that five samples were 01 in the reference data and most likely 01 in the map, an average posterior-probability value of 0.7 for this combination would indicate that the knowledge base was 70% confident that these locations were 01.

### 3 RESULTS

#### 3.1 KNOWLEDGE BASE STRUCTURE

The Bayesian-network knowledge base for the CWHvm2 was composed of 46 nodes, 58 links and 2179 conditional probabilities (Figure 16). Of the 46 nodes in the network, 17 nodes were input nodes, which represented the layers in the GIS database, and one node was an output node, which represented the final estimation of site-series group (**Site Series Veg Modifier**). Based on the input nodes' unconditional probability values for the entire study area, the posterior probabilities for the final estimation of site-series groups -01 (SSV 01), -03 (SSV 03), -05 (SSV 05), -09 (SSV 09), -10 (SSV 10) and -11 (SSV 11), were 46.4%, 25.9%, 20.8%, 4.79%, 0.5% and 1.7%, respectively.

The Bayesian-network knowledge base for the CWHvm1 was composed of 47 nodes, 60 links and 2759 conditional probabilities (Figure 17). Of the 47 nodes in the network, 17 nodes were input nodes, which represented the layers in the GIS database, and one node was an output node, which represented the final estimation of site-series group (**Site Series Fluvial Modifier**). Based on the input nodes' unconditional-probability values for the entire study area, the posterior probabilities for the final estimation of site-series groups -01 (SSF 01), -03 (SSF 03), -05 (SSF 05), -09 (SSF 09), -12 (SSF 12), -13 (SSF 13) and -14 (SSF 14), were 42.6%, 16.3%, 25%, 6%, 5.5%, 0.4% and 4.2%, respectively.

The Bayesian-network knowledge base for the CWHvh2 was composed of 50 nodes, 64 links and 3725 conditional probabilities (Figure 18). Of the 50 nodes in the network, 19 nodes were input nodes, which represented the layers in the GIS database, and one node was

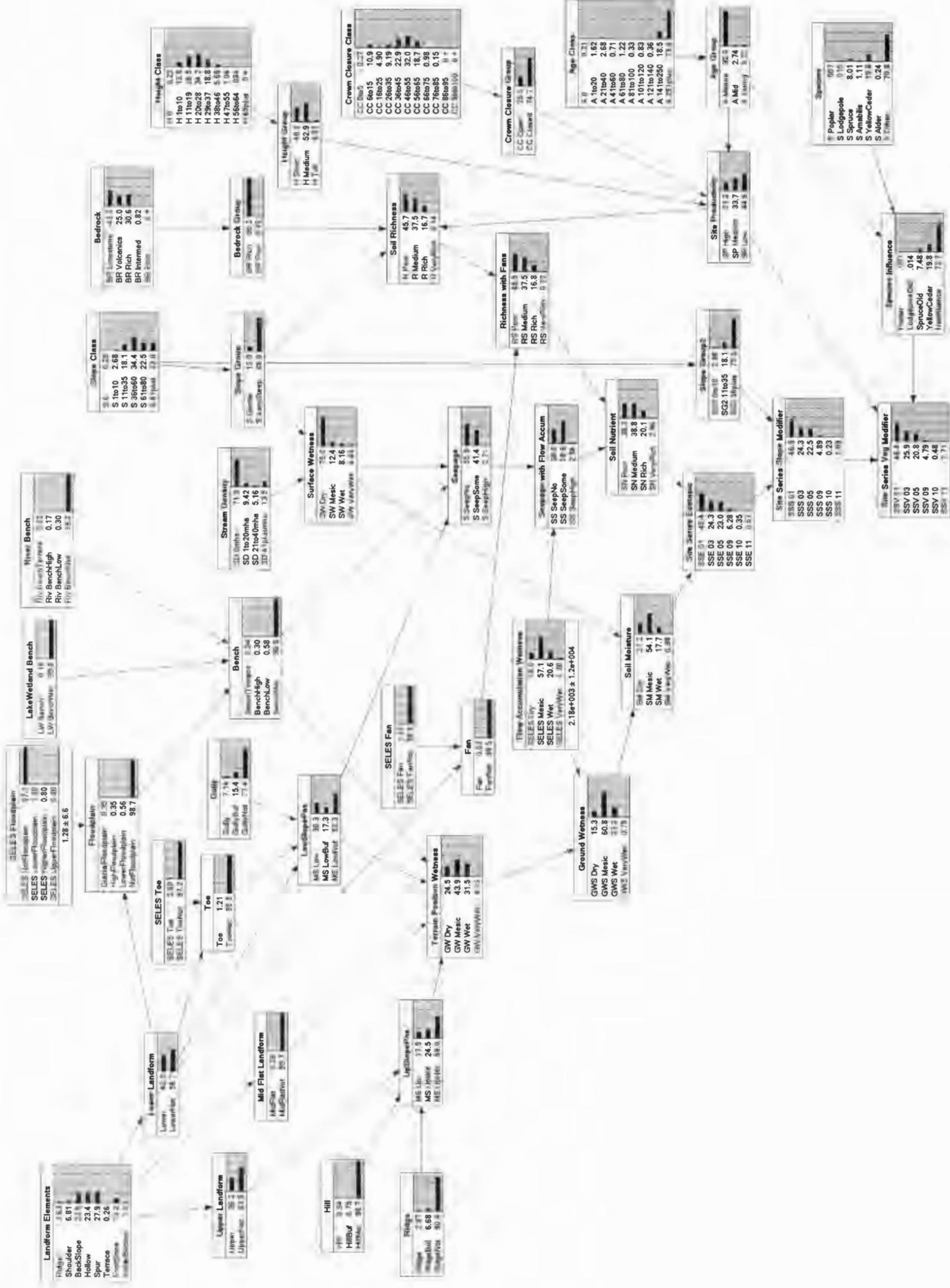


Figure 16 – The complete Bayesian network developed to predict site-series units in the CWHv2m2 BEC group in the CWHv2m2 BEC unit of this study. A description of the node format and nomenclature can be found in Figure 5.





an output node, which represented the final estimation of site-series group (**Site Series Shore Modifier**). Based on the input nodes' unconditional-probability values for the entire study area, the posterior probabilities for site-series groups -01 (SSSh 01), -03 (SSSh 03), -04 (SSSh 04), -07 (SSSh 07), -08 (SSSh 08), -11 (SSSh 11), -12 (SSSh 12), -13 (SSSh 13), -14 (SSSh 14) and -18 (SSSh 18), were 20.2%, 10.6%, 14%, 8.5%, 0.04%, 28%, 9.3%, 5.2%, 4.1% and 0.01%, respectively.

### 3.2 PREDICTIVE ECOSYSTEM MAPS

Figure 19 illustrates the spatial arrangement of those site-series groups with the highest probability value within each raster cell. The spatial arrangement reveals the general location of each of the predicted site-series-groups and their abundance within the landscape. Figure 20 and Figure 21 provide detailed views of a section of the study area, with the intention of demonstrating the relative terrain position of the predicted site series groups.

Within the CWHvm2, 01 was the most abundant site-series group and situated mid-slope. The second most abundant site-series group was 03, which was situated at higher elevations within large valleys. The third most abundant site-series group was 05, which was placed close to rivers and streams. The 09 and 10 site-series groups occurred only in small isolated areas at lower elevations within the valleys.

Within the CWHvm1, like the CWHvm2, 01 was the most abundant site-series group, followed by 03 and 05. With respect to the relative spatial position of these three leading site-series groups, 01 site-series group was placed on mid-slopes, 03 was placed at higher

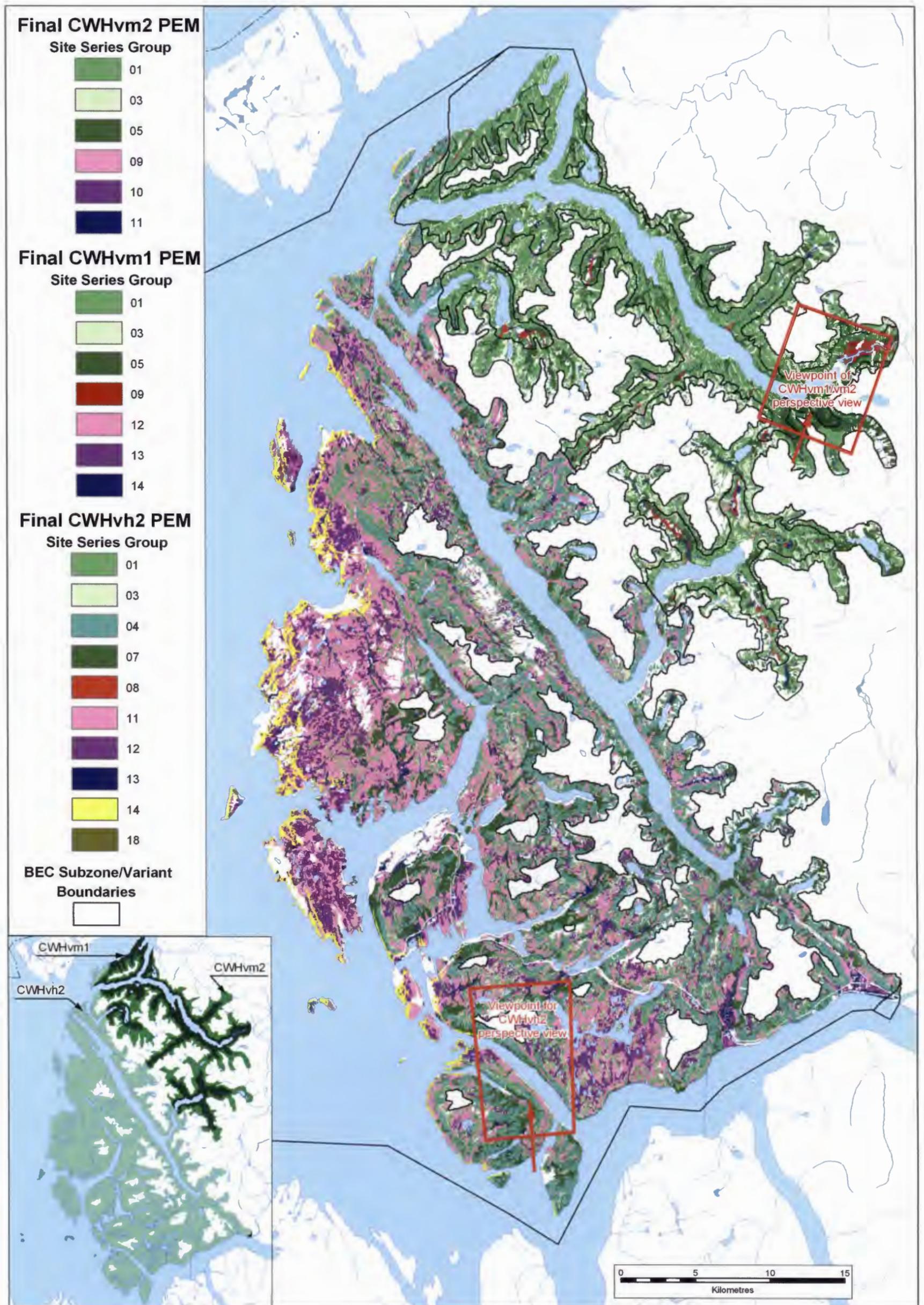


Figure 19 – An overview map of predicted site-series groups in all three BEC subzone/variants in the study area. This map shows the relative position and abundance of each of the predicted site-series groups within each of the three BEC subzone/variants. The red “viewpoint” boxes indicate the spatial extent of the perspective view in Figure 20 and Figure 21.

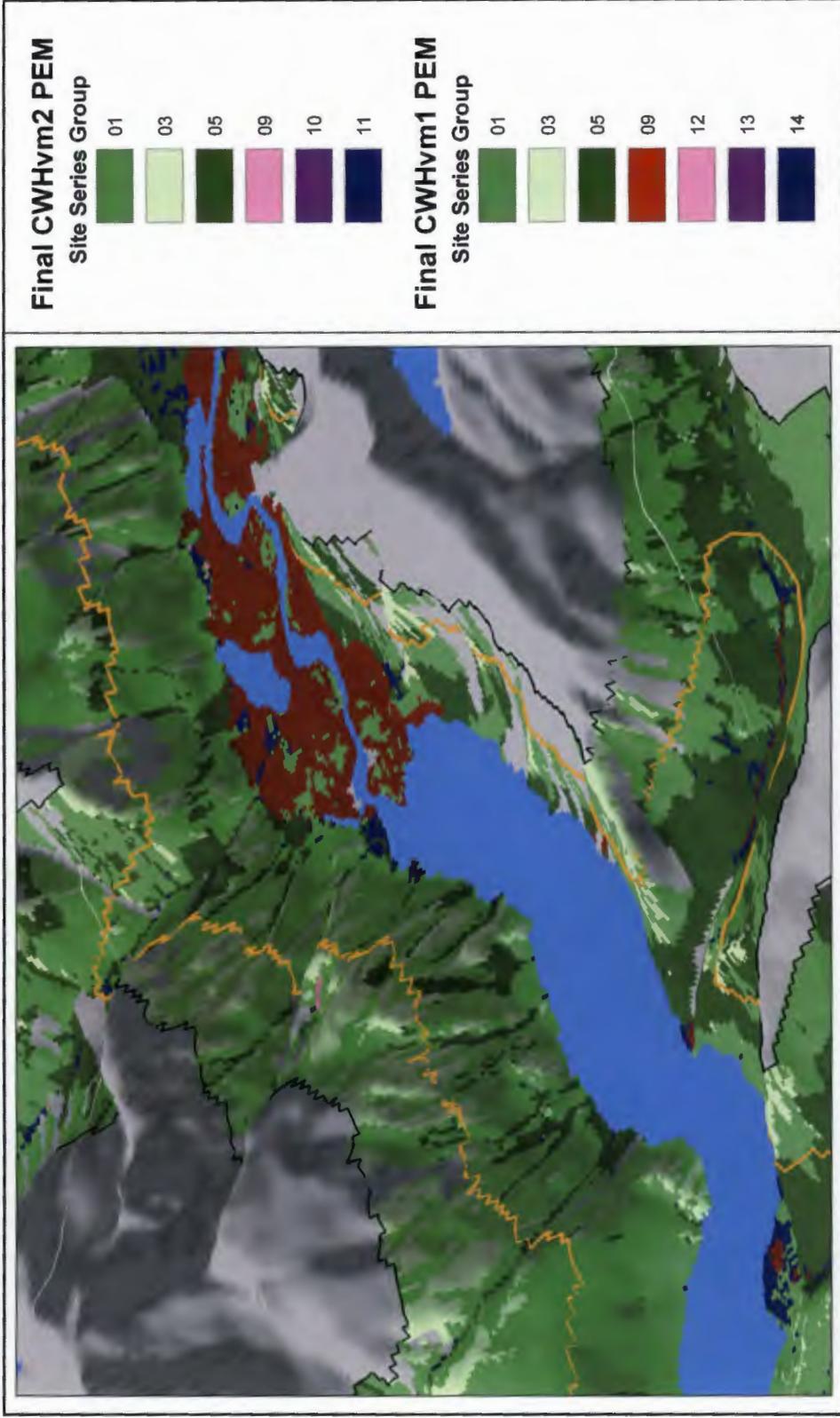


Figure 20 – A perspective view illustrating the relative terrain position of the final site-series-group predictions for the CWHvm2 and CWHvm1 BEC units. The wet site-series groups, such as 05, appear to be correctly placed within the gullies and toe slopes, and the dry site-series groups, such as 03, appear to be correctly placed on ridges and shoulders. In the figure, the CWHvm2 occurs above the orange line, which separates the CWHvm1 from the CWHvm2. Figure 19 shows the location of this perspective view within the study area.



Figure 21 – A perspective view illustrating the relative terrain position of the final site-series-group predictions for the CWHvh2 BEC unit. The wet site-series groups, such as 07 and 14, appear to be correctly placed in the gullies and toe slopes, and the dry site-series groups, such as 03, appear to be correctly placed on ridges and shoulders. Figure 19 shows the spatial extent of this perspective view within the study area.

elevations within the large valleys, and 05 was placed close to small rivers. The 09 site-series group was situated along the bottom of larger valleys while 13 and 14 appeared higher up the valley sides than 09.

Within the CWHvh2, the low relief areas were defined by the dominance of site-series-groups 11, 12 and 13, while the 01, 03, 04 and 07 site-series-groups dominated the mountainous areas. Site-series-group 14, which was related to offshore winds, occurred along the western coastline of the study area.

### 3.3 MAP ACCURACY ASSESSMENT

The map of predicted site-series group for the CWHvm2 obtained an overall percentage correct of 47.8%, with a 95% confidence interval of  $\pm 9.8\%$ . Producer's accuracy varied from 0% to 65.6% and user's accuracy varied from 0% to 59.1% (Table 9). The Kappa analysis produced a KHAT of 0.249 with a variance of 0.006, producing a Z-score of 3.28 ( $p = 0.001$ ). Average posterior probabilities for the predicted site-series groups ranged from 0% to 55% (Table 10). For those map-accuracy points where there was a match between the site-series-group value in the reference data and in the map, the average posterior probabilities ranged from 0% to 59.4%.

The map of predicted site-series group for the CWHvm1 obtained an overall percentage correct of 50.3%, with a 95% confidence interval of  $\pm 7.4\%$ . Producer's accuracy varied from 0% to 73.8% and user's accuracy varied from 0% to 82.8% (Table 11). The Kappa analysis produced a KHAT of 0.37 with a variance of 0.022, producing a Z-score of 2.51 ( $p = 0.012$ ). Average posterior probabilities for the predicted site-series groups ranged from 0% to 53.7% (Table 12). For those map-accuracy points where there was a match

Table 9 – The accuracy-assessment contingency-matrix for the CWHvm2, which includes percentage correct, user’s and producer’s accuracy.

Map Site Series Group	Reference Site Series Group						Total	User's Accuracy
	01	03	05	09	10	11		
01	21	4	18	1	1	2	47	44.7%
03	5	13	1	0	3	0	22	59.1%
05	6	3	10	0	0	3	22	45.4%
09	0	0	0	0	0	0	0	0.0%
10	0	0	0	0	0	0	0	0.0%
11	0	0	0	1	0	0	1	0.0%
<b>Total</b>	32	20	29	2	4	5	92	
<b>Producer's Accuracy</b>	65.6%	65.0%	34.5%	0.0%	0.0%	0.0%		
<b>Percentage Correct</b>	47.8%							

Table 10 – The matrix of average posterior probabilities for the CWHvm2, which includes the average posterior probabilities for each mapped site-series group.

Map Site Series Group	Reference Site Series Group						Average
	01	03	05	09	12	14	
01	<b>0.484</b>	0.511	0.500	0.560	0.468	0.459	0.493
03	0.534	<b>0.594</b>	0.516	0	0.400	0	0.550
05	0.493	0.456	<b>0.436</b>	0	0	0.466	0.458
09	0	0	0	<b>0</b>	0	0	0.000
10	0	0	0	0	<b>0</b>	0	0.000
14	0	0	0	0.290	0	<b>0</b>	0.290

Table 11 – The accuracy-assessment contingency-matrix for the CWHvm1, which includes percentage correct, user’s and producer’s accuracy.

Map Site Series Group	Reference Site Series Group							Total	User's Accuracy
	01	03	05	09	12	13	14		
01	31	12	15	2	8	0	4	72	43.1%
03	6	10	0	0	1	4	0	21	47.6%
05	4	0	24	1	0	0	0	29	82.8%
09	1	0	7	9	0	0	7	24	37.5%
12	0	0	0	0	0	1	0	1	0.0%
13	0	0	0	0	0	0	0	0	0.0%
14	0	0	1	3	3	0	7	14	50.0%
<b>Total</b>	42	22	47	15	12	5	18	161	
<b>Producer's Accuracy</b>	73.8%	45.4%	51.1%	60.0%	0.0%	0.0%	38.9%		
<b>Percentage Correct</b>	50.3%								

Table 12 – The matrix of average posterior probabilities for the CWHvm1, which includes the average posterior probabilities for each mapped site-series group.

Map Site Series Group	Reference Site Series Group							Average
	01	03	05	09	12	13	14	
01	<b>0.435</b>	0.441	0.423	0.368	0.420	0	0.346	0.425
03	0.469	<b>0.548</b>	0	0	0.820	0.543	0	0.537
05	0.488	0	<b>0.423</b>	0.380	0	0	0	0.431
09	0.371	0	0.506	<b>0.451</b>	0	0	0.474	0.470
12	0	0	0	0	<b>0</b>	0.408	0	0.408
13	0	0	0	0	0	<b>0</b>	0	0
14	0	0	0.426	0.306	0.365	0	<b>0.355</b>	0.352

from 0% to 53.7% (Table 12). For those map-accuracy points where there was a match between the site-series-group value in the reference data and in the map, the average posterior probabilities ranged from 0% to 54.8%.

The map of predicted site-series group for the CWHvh2 obtained an overall percentage correct of 33.3%, with a 95% confidence interval of  $\pm 15.2\%$ . Producer's accuracy varied from 0% to 60% and user's accuracy varied from 0% to 100% (Table 13). The Kappa analysis produced a KHAT of 0.199 with a variance of 0.294, producing a Z-score of 0.366 ( $p = 0.714$ ). Average posterior probabilities for the predicted site-series groups ranged from 0% to 57.6% (Table 14). For those map-accuracy points where there was a match between the reference data and the map site-series-group values, the average posterior probabilities ranged from 0% to 54.6%.

Table 13 – The accuracy-assessment contingency-matrix for the CWHvh2, which includes percentage correct, user's and producer's accuracy.

Map Site Series Group	Reference Site Series Group																		Total	User's Accuracy	
	01	03	04	07	08	11	12	13	14	18											
01	3	2	1	1	0	4	0	0	0	0	0	0	0	0	0	0	0	0	11	27.3%	
03	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	100.0%	
04	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	33.3%	
07	0	0	4	6	0	0	0	2	0	0	0	0	0	0	0	0	0	0	12	50.0%	
08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0%	
11	2	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	5	0.0%	
12	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0.0%	
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0%	
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0%	
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0%	
<b>Total</b>	5	5	6	7	0	4	1	5	0	0	0	0	0	0	0	0	0	0	33		
<b>Producer's Accuracy</b>	60.0%	20.0%	16.7%	85.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%			
<b>Percentage Correct</b>	33.3%																				

Table 14 – The matrix of average posterior probabilities for the CWHvh2, which includes the average posterior probabilities for each mapped site-series group.

Map Site Series Group	Reference Site Series Group																		Average
	01	03	04	07	08	11	12	13	14	18	18								
01	<b>0.380</b>	0.385	0.396	0.362	0	0.393	0	0	0	0	0	0	0	0	0	0	0	0.385	
03	0	<b>0.374</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.374	
04	0	0.351	<b>0.318</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.340	
07	0	0	0.600	<b>0.546</b>	0	0	0	0.619	0	0	0	0	0	0	0	0	0	0.576	
08	0	0	0	0	<b>0</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	0.522	0	0	0	0	<b>0</b>	0.490	0.556	0	0	0	0	0	0	0	0	0	0.529	
12	0	0	0	0	0	0	<b>0</b>	0.388	0	0	0	0	0	0	0	0	0	0.388	
13	0	0	0	0	0	0	0	<b>0</b>	0	0	0	0	0	0	0	0	0	0	
14	0	0	0	0	0	0	0	0	<b>0</b>	0	0	0	0	0	0	0	0	0	
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

#### 4 DISCUSSION

One of the objectives of this project was to develop an approach to map large-scale ecological units that incorporated expert knowledge. It was anticipated that this objective would be met by using Bayesian networks as the knowledge base of an expert-based system. Bayesian networks also held promise of being a good way to allow an ecological-domain expert to build a model that could capture prediction uncertainty when describing the relationships between environmental variables and site series.

Unfortunately, none of the maps of site-series group produced by this study correctly predicted the spatial locations of even 65% of the known site-series groups. Results from the accuracy assessment procedures performed within the CWHvm2 unit indicated that the predictive-ecosystem map was only fairly representing the spatial distribution of site-series group. The predictive ecosystem map for the CWHvm2 obtained an overall percentage correct of 47.8%, with a 95% confidence interval of  $\pm 9.8$ . This indicated that the map performed fairly well, but fell short of the target of 65% established by Meidinger (2003). This result, however, was higher than the result obtained by the North Coast PEM Project, which used a belief-matrix knowledge base. The North Coast PEM Project, although utilizing slightly different site-series groupings, achieved an overall percentage correct of 21% in the CWHvm2 (Meidinger 2004).

According to the contingency matrix for the CWHvm2 (Table 9), the map successfully identified the 01 and 03 site-series groups, with producer's accuracies of 65.6% and 65% respectively. The map, however, performed poorly at identifying the remaining site-series groups, with 05 being successfully predicted only 34.5% of the time and 09, 10

and 11 not predicted at all relative to the reference data. Results of the CWHvm2 user's accuracy assessment revealed that several of the site-series groups were assigned to the wrong unit. A large number of the samples identified as 01 on the map were identified as 05 in the reference data. This was similar for the 05, where the map falsely predicted the 01 site-series groups as 05. These errors of commission were confirmed by the low user's accuracy values (Table 9). The map was also unable to identify any of the 09 or 10 units within the reference data, and unable to correctly identify any 11 units within the reference data.

Although the KHAT value of 0.249 indicated that the map poorly predicted site-series group, the KHAT variance was 0.006. This meant that, although the map was unable to meet the 65% overall percent correct, the map was significantly better than chance agreement.

Evaluation of the average posterior probabilities generated by the BN for each of the site-series groups mapped in the CWHvm2 (Table 10) revealed that each site-series group was predicted with differing levels of confidence. The 03 site-series group was predicted with the highest confidence, with an average posterior probability of 55%, but none of the other site-series groups were predicted with high confidence. The 01 and 05 site-series groups were mapped with confidence levels of 49.3% and 45.8% respectively, while 14 was predicted with the lowest confidence: 29.2%. Site-series-groups 09 and 10 were not predicted at all in relation to the reference data, resulting in 0% confidence for both.

Results from the accuracy assessment procedures for the CWHvm1 indicated that the predictive ecosystem map only partially represented the spatial distribution of site-series group (Table 11). The overall percentage correct of 50.3%, with a 95% confidence interval

of  $\pm 7.4$ , indicated that the map performed fairly well at predicting site-series group, but, as with the CWHvm2, fell short of the target of 65%. This result, however, was similar to the result obtained by the North Coast PEM Project, which achieved an overall percentage correct of 50% for the CWHvm1 using similar site-series groups (Meidinger 2004).

According to the contingency matrix for the CWHvm1 (Table 11), the map of site-series group did successfully identify the 01, with a high producer's accuracy of 73.8%. The map, however, achieved only moderate success when predicting the 09, 05 and 03 sites-series groups, with producer's accuracy results of 60%, 51.1% and 45.4% respectively. The map poorly predicted the 14 site-series group, with only 38.9% producer's accuracy, and was unable to predict any of the 12 or 13 site-series groups. Results of the user's accuracy assessment revealed that several of the site-series groups were assigned to the wrong unit. The user's accuracy results for both 03 and 14 were moderate, with approximately half the mapped sites actually representing the reference data. Although the producer's accuracy for 01 was high, a large number of the accuracy assessment points identified as 01 on the map were identified as 05 or 03 in the reference data, resulting in a low user's accuracy value for the 01 sites-series group. The same was true for the 09, where the map falsely predicted the 05 and 14 site-series groups to be 09. The map was also unable to identify any 13 site-series groups within the reference data, and unable to correctly identify any 12 site-series groups within the reference data.

The KHAT value of 0.37 suggests that the map poorly/fairly predicted site series group. However, a KHAT variance of 0.022 meant that, like the site-series-group map for the CWHvm2, the map was significantly better than chance agreement, although the map was unable to meet the 65% overall percent correct.

The average posterior probabilities generated by the BN were roughly equivalent for each of the site-series groups that were mapped in the CWHvm1 (Table 12), although none of the site-series groups were predicted with high confidence. Like the CWHvm2, the 03 site-series group was predicted with the highest confidence, with an average posterior-probability of 53.7%. The remaining site-series groups, with the exception of 13, were predicted with moderately low confidence, with values ranging from 35.2% to 47%. Site-series-group 13 was not predicted at all with respect to the reference data, resulting in 0% confidence.

An overall percentage-correct value of 33.3%, with a 95% confidence interval of  $\pm 15.2\%$ , indicated that the CWHvh2 map poorly predicted site-series group, and fell well short of the target of 65%. Again, this result was similar to the result obtained by the North Coast PEM Project, which achieved an overall percentage correct of 38% in the CWHvh2 using similar site-series groups (Meidinger 2004).

The map of site-series group for the CWHvh2 performed well when identifying 07 and moderately well when identifying 01, with producer's accuracies of 85.7% and 60% respectively. The map, however, performed poorly when identifying the remaining site-series groups. The 03 and 04 site-series groups obtained producer's accuracies of only 20% and 16.7%, respectively, and site-series-groups 11, 12 and 13 were never identified correctly with respect to the reference data. The 08, 14 and 18 site-series groups could not generate producer's accuracy values greater than zero because these site-series groups were not present in the reference data. The user's accuracy for the 07 site-series group was fair, with approximately half the mapped 07 sites actually representing 07 in the reference data (Table 13). Several of the other site-series groups, however, were assigned to the incorrect unit.

Although the producer's accuracy for 01 was high, a large number of the accuracy-assessment points identified as 01 on the map were identified as 11 in the reference data. The 04, 11, 12 and 13 site-series groups were overestimated on the map, which resulted in low user's accuracy. The user's accuracy results for site-series groups 08, 14 and 18 were understandably zero because they were not present in the reference data. The KHAT value of 0.199 indicated that the map poorly predicted site-series group. In addition, due to the low number of reference-data points per site-series group, the comparison between map and reference data produced a high KHAT variance value: 0.294. This meant that not only was the site-series-group map not able to meet the 65% overall percent correct target but the map was not significantly better than chance agreement.

The average posterior probabilities for the site-series groups in the CWHvh2 (Table 14) can roughly be broken into three categories: those site-series groups that were not predicted at all, those that were predicted with low confidence, and those that were predicted with moderate confidence. Site-series-groups, 08, 13, 14 and 18, were predicted with 0% confidence because they were never predicted to be the dominant site-series group at the reference-data locations. Site-series-groups, 01, 03, 04 and 12, were all predicted to be the dominant site-series group with a low level of confidence, with confidence values ranging from 34% to 38.8%. The 07 and 11 site-series groups were predicted to be the dominant site-series group with moderate confidence, with confidence values of 57.6% and 52.9%, respectively.

During the accuracy assessment procedures, the georeferenced reference-data were considered correct in both spatial accuracy and thematic content. The impact of this assumption was that, because accuracy-assessment results were relative to the reference data,

the accuracy-assessment results indicated how well the generated map represented the reference data, not how well the map represented what was actually on the ground. The spatial juxtaposition of the predicted-site-series groups for the CWHvm2 and CWHvm1, however, were similar to the landscape profiles presented in the field guide of Banner *et al.* (1993). The knowledge bases for the CWHvm2 and CWHvm1 were considered to be performing well at predicting the most common site-series groups, based on the ecological-domain expert's visual interpretation of relative position and abundance of each site-series group within the final ecosystem maps (Figure 19) and the perspective view (Figure 20).

The relative terrain positions of the predicted site-series groups in the CWHvh2 generally match the landscape profiles that were presented in the Banner *et al.* (1993) field guide for site-interpretation. Based on ecological-domain expert's visual interpretation of relative position and abundance of each site-series group within the final ecosystem maps (Figure 19) and the perspective view (Figure 21), the CWHvh2 knowledge base appeared to perform fairly at predicting the most common site-series groups. The 07 site-series group occurred at lower elevations on moderate slopes while the 03 occurred at higher elevations. Site-series-groups 01 and 04 were between the 07 and 03, with 04 tending towards steeper slopes than 01. Site-series-groups 11 and 12 were found on low-slope areas with 12 being found closer to water bodies than 11. Observation that the 03 site-series group was being placed in the upland areas within complexes of 11 and 12, however, was disconcerting. According to the ecological-domain expert, these areas of 03 within 12/13 should have a higher probability of being 01 rather than 03. This indicated that the knowledge base was underestimating the soil moisture of these upland areas within larger lowland complexes.

Comparison of the row totals to the column totals in all three contingency matrices revealed that 01 site-series group was overestimated in all three BEC subzone/variants. In the CWHvm2 and CWHvm1 maps, this overestimation was to the detriment of the second most common site-series group – 05. While there was overestimation of the 01 on the map, there was also underestimation of some site-series groups. Generally, those site-series groups being underestimated and incorrectly mapped had very wet soil moisture regime and occurred infrequently in the reference data. The underestimation of these very wet site-series groups may in part be due to the scale of the GIS data.

If these very wet units occurred in small isolated areas then the input GIS data may not have contained adequate spatial resolution to detect the environmental conditions required to delineate these units. For example, site-series-group 13 in the CWHvh2 usually covers no more than a couple of hectares and is commonly associated with depressions with low forest productivity. Although the minimum spacing between TRIM elevation points was 25 m, the average was roughly 75 m, making it difficult to detect small one- or two-hectare depressions. In addition, forest-cover polygons had an average polygon area of 12 ha in the study area and did not delineate small one- or two-hectare features. A common technique used in past PEM projects was to indicate that some small-unmapped sites were present within some larger ecological units (Jones 2003). For example, units mapped as site-series 01 in the CWHvh2 of Jones (2003) actually contained approximately 90% –01 and 10% –11.

Evaluation of the off-diagonal elements in all three contingency matrices revealed that generally, errors occurred as a result of a cell being incorrectly assigned to the adjacent unit in the edatopic grid. For example, 01 site-series group, which was adjacent to 03, 05 and 09 groups in the edatopic grid of CWHvm2, was commonly misidentified as either 03 or 05

site-series group. Although this was an encouraging sign that the knowledge base was producing results close to the correct location on the edatopic grid, the level of misclassification was considered high. The solution employed by previous predictive-mapping projects that map site-series was to improve classification accuracy by grouping site-series units into larger site-series groups (Jones 2003). A further grouping of site-series groups was not considered a reasonable solution in the current project because the knowledge bases already incorporated groups of site series.

In the CWHvh2, the original 01 and the 04 site series of Banner *et al.* (1993) occupy adjacent areas on the edatopic grid and exist within similar environmental conditions. During development of the CWHvh2 knowledge base, it was recognised that the GIS database may not be detailed enough to separate the 01 site series from the 04. Although these two similar site series were not grouped during CWHvh2 model development, they could have been grouped within the Bayesian network through the creation of an alternate output node.

In retrospect, it may have been best to create the knowledge bases with the intent to predict site series, not grouped site-series. As a last step in the Bayesian network development, an alternate output node could have been added, which grouped together those ecologically similar site series that the knowledge base was having difficulty differentiating. This alternate approach could also ease future knowledge base updating. For example, with the current knowledge bases, which group site-series from the onset, if new data become available that can differentiate previously grouped site-series then many of the conditional probability tables involving site-series group would need to be altered. If the Bayesian knowledge base was created with the intention to predict site series then the final site-series-

grouping node would be the only node requiring an alteration to the number of site-series classes.

Although the BN knowledge bases were constructed solely from expert opinion, the results of the accuracy assessment procedures were lower than anticipated, especially, given the moderately-high agreement between the CWHvm2 and CWHvm1 testing-data set and the test maps. This disparity between the results of knowledge base testing and the accuracy assessment results could be partially explained by the different methodologies used to collect the two sets of reference data. The reference data used for accuracy assessment was comprised of site-series observations collected through ground surveys, while the set of reference data used for testing was collected using air-photo interpretation. Although both sets of data were gathered using sampling plans that intended to capture the proportional occurrence of each site series within the study area, the testing set contained a higher proportion of points in the more common site-series groups than did the accuracy assessment set. This increase in the more common site-series groups in the set of testing reference data may have contributed to the high percent-correct scores during knowledge base testing. Disagreement between the results of knowledge base testing and the results of map-accuracy assessment could have been reduced using the same sampling plan to gather the reference data for knowledge base testing and accuracy assessment of the map.

A review of the matrices that recorded the average posterior-probabilities for each of the site-series groups revealed that none of the most likely site-series-groups were predicted with high confidence. In all three BEC subzone/variants, the average confidence values for the most-likely site-series group were around 0.4 or 0.5. These values indicated that rarely was there a clear majority in the soft-classification results at each location. Although

disappointingly low, these confidence values reflected how difficult it was to determine a location's site-series group, given the information provided in the input GIS database. The ecological-domain expert often indicated that conditional-probability values should be lowered to compensate for a lack of confidence in either the input GIS database or subsequent intermediate nodes. Thus, lowering the conditional-probability values leads to a reduction in prediction confidence.

The accuracy assessment results and confidence values revealed the knowledge base and GIS database performed poorly together when making an estimate of site-series group. The overall percentage-correct results, which were produced using the site-series group with the highest probability at each location, indicated that the dominant site-series group was correctly identified by the knowledge base at less-than 50% of the reference data locations. The low posterior-probability values further highlighted the poor performance of the knowledge base. Rarely did the knowledge base confidently predict the correct site-series group from the input GIS database. In addition to the low percent-correct and confidence values, the user's accuracy results were poor for almost each site-series group.

The methods employed in this project did not produce a large-scale ecosystem classification map that met the needs of forest-related management activities. With low percent-correct results and low prediction confidence, the maps produced by this study would not be suitable for forest management activities that rely on a minimum 65% correct estimation of either site-series or the site-series groups created in this study. For example, the resulting maps would not be suitable for timber supply analysis, identifying the location of rare ecosystems, or silviculture prescriptions.

As with all expert systems, the results and knowledge base were closely linked to the opinion of the expert. This could make it difficult to replicate the results without using the same experts involved in the initial model development. It also makes it difficult to assess the model's sensitivities, especially if the model is not explicitly and transparently presenting the assumptions and decisions made by the domain expert. The Bayesian-networks developed for this study, however, explicitly and transparently captured and presented the assumptions and decisions made by the ecological-domain expert. The chances of successfully replicating the process of estimating site-series group from the GIS input layers are greatly improved by the Bayesian network's transparency.

The BN was successful at capturing and expressing ecological knowledge with respect to the relationships between environmental variables and site-series group. The opportunity to create intermediate nodes, which were ecologically meaningful to the ecological-domain expert, allowed the ecological-domain expert to focus on only a few ecological interactions at a time. The task of generating conditional probabilities for intermediate nodes that were derived from either multiple parent nodes or parent nodes with multiple classes, however, was sometimes daunting. For example, the number of conditional probabilities required for a node with two classes and two parent nodes, each with two classes, equals eight ( $2 \times 2 \times 2$ ). A node with six classes and two parents, each with four classes, however, required the definition of 96 ( $6 \times 4 \times 4$ ) conditional probabilities.

Although nodes with a large number of conditional probabilities existed within the three Bayesian networks, no nodes had more than three parent nodes. By keeping the number of parent nodes to a minimum, the ecological-domain expert could focus on a few environmental factors at a time. This was an improvement over the belief-matrix method of

large-scale ecosystem mapping, where the knowledge base could be comprised of a single table requiring 2000 or more defined weights (e.g., Timberline Forest Inventory Consultants Ltd. 2000; Rosen *et al.* 2001; Atticus Resource Consulting Ltd. 2001; Jones 2003; Timberline Forest Inventory Consultants Ltd. 2004). Although the total number of conditional probabilities within the three BN knowledge bases ranged from 2179 to 3725, partitioning the ecological influences into smaller, independent interactions made model parameterization manageable for the ecological-domain expert.

One difficulty with regards to the parameterization of the Bayesian network's conditional-probability tables, however, was the requirement to enter conditional probabilities for all combinations of the parent-node classes. This meant that conditional probabilities would be required for environmental conditions that seem mutually exclusive. For example, the **Ground Wetness** node was dependent on parent nodes that defined terrain position: upper-terrain position (**UpSlopePos**), mid-terrain position (**Mid Flat Landform**), lower-terrain position (**LowSlopePos**), and fluvial benches (**Bench**). Sometimes it was easy entering the conditional probabilities for **Ground Wetness**, such as when the input combination was: upper-terrain position (MS Up), not mid-terrain position (MidFlatNot), not lower-terrain position (MS LowNot) and not fluvial bench (BenchNot). With this input combination, the **Ground Wetness** was given 100% probability of being dry. It was more difficult, however, to enter conditional-probabilities values for **Ground Wetness** when the combination of input conditions was: upper-terrain position (MS Up), mid-terrain position (MidFlat), lower-terrain position (MS Low) and high bench (BenchHigh). Each one these input conditions is associated with a different level of ground wetness. As a result, with this combination of input conditions, there was less certainty that wetness could be ascribed to

only one **Ground Wetness** class. To capture this uncertainty, **Ground Wetness** was given a 20% probability of being wet and an 80% probability of being very wet for this combination of input conditions.

In previous projects conducted in BC involving large-scale predictive ecosystem mapping, each GIS database layer could only contain a single value for each location (Ketcheson *et al.* 2001a; Ketcheson *et al.* 2001b; Ketcheson *et al.* 2001c; Jones & McGregor 2002; Jones 2003). The ability of the Bayesian network knowledge base to accept fuzzy classified GIS data inputs was seen as an advantage over previously employed methods for predictive-site-series mapping. The BN was able to capture the continuous nature of the **Landform Elements** input node, which was generated by the continuous-landform-elements BN. The BN was also able to capture the composition of tree species within a forest-vegetation stand through the forest-species input node (**Species**). Other input nodes, such as **Age Class**, **Crown Closure Class** and **Height Class** could be extended to use this technique of accepting data as a soft classification rather than a hard classification.

In all three of the BN knowledge bases created to estimate site-series group, some input and intermediate nodes were parents to more than one child node. For example, in all three models the **Lower Landform** intermediate node was used to help define floodplains (**Floodplain**), toe slopes (**Toe**), low-terrain position (**LowSlopePos**) and alluvial fans (**Fan**). This produced a messy and confusing looking network with links crossing each other and pointing different directions, rather than a simple dendritic structure, which progresses from input to output with all links pointing in the same general direction. The use of input and intermediate nodes for multiple conditional-probability tables, however, reduced variable redundancy within the network.

Previous projects of large-scale predictive ecosystem mapping conducted in BC, which employed the belief-matrix knowledge base, were unable to use automated procedures to update the knowledge base to reflect field observations (Jones & McGregor 2002; Jones 2003). Bayesian networks, however, are able to incorporate relationships derived either statistically and from an expert. At the outset of this project, however, it was recognized that model development would rely primarily on expert opinion because the availability of useful reference data was low.

Although the study area was partially chosen because of reference-data availability, examination of the reference data during model development revealed that not all site-series groups were adequately sampled. Future projects of large-scale predictive mapping should ensure that an adequate representative sample of reference data be acquired, preferably, specifically for the predictive-mapping project.

Additional field gathered reference-data would have proved useful at many stages during model development, not necessarily to substitute expert opinion with statistical models, but to aid the expert's decision making process. Field data, which recorded site series and were georeferenced, would have proved useful for both model testing and increasing the number of reference points available for accuracy assessment. For example, the disparity between the results of the knowledge base testing, which ranged from 59% to 92% correct, and the results of the map-accuracy assessment, which ranged from 33% to 48% correct, highlighted the difficulty of accurately interpreting site series from air photos. The results of knowledge base testing may have matched the results of map-accuracy assessment if knowledge base testing was conducted using reference points that were field gathered rather than air photo interpreted.

The lack of reference data for map-accuracy assessment hindered adequate map-accuracy assessment, especially when assessing those site-series groups occurring infrequently on the landscape. Although the CWHvm1 and CWHvm2 had a large number of reference plots in the common site-series groups, there were a low number of reference points in the uncommon sites. There were very few reference points in each site-series groups of the CWHvh2, due to the low number of CWHvh2 field plots gathered by Jones (2003) that overlapped with the study area. This low number of CWHvh2 field plots contributed to the high kappa variance. In all three of the developed Bayesian networks, some site-series groups were not present in the reference data, preventing accuracy assessment of these site-series groups.

Field data, which recorded the environmental conditions estimated by some or all of the BN nodes, would have allowed additional knowledge base adjustments. It would have been possible to perform some exploratory analysis into the relationships between environmental factors and site-series groups using this kind of reference data. Using such techniques as cluster analysis and regression it may have been possible to see how relevant each node was to the estimation of site-series group, and remove those nodes with little or no significance prior to model construction.

Reference data would have possibly been useful during conditional-probability adjustment and construction of the input GIS database. All of the input and intermediate nodes in the Bayesian networks were meant to represent measurable environmental factors (e.g., **Soil Moisture** and **Gully**). Although the thematic and spatial accuracy of some input nodes were known (i.e. those nodes derived from the Ministry of Sustainable Resource Management's vegetation resource inventory), this information was unknown both for the

GIS modelled input nodes and the intermediate nodes. Thematic and spatial accuracy assessment of these nodes was based solely on visual comparison to other informative data sources, such as air photos, water body/course mapping or digital elevation models. Consequently, little was known about the input or intermediate nodes' ability to accurately portray their respective themes. Reference data could have identified either those input GIS models needing adjustment or the conditional-probability tables needing adjustment.

One of the objectives of this project was to create a model that could be updated automatically using field data. This objective was achieved by using Bayesian networks, which allow the conditional probabilities to be automatically adjusted using expectation-maximization algorithm. Unfortunately, this procedure was not possible to conduct without reference data, which recorded the environmental conditions estimated by the intermediate nodes. However, by directing future field-sampling plans to collect those environmental conditions modelled in the knowledge bases, it would be possible to continuously update the conditional probabilities, as new sampling data becomes available.

## 5 CONCLUSION AND RECOMMENDATIONS

The overall goal of this project was to develop and test a large-scale ecosystem classification mapping method that met the needs of forest-related management activities in British Columbia. The approach employed was an expert-based system built around a Bayesian-network knowledge base. This approach, however, did not meet the project's overall goal because not all three of the objectives were accomplished successfully. None of the large-scale ecosystem maps produced for the study area's biogeoclimatic variants achieved the accuracy assessment result of 65% correct. The approach, however, did successfully achieve the other two study objectives.

Bayesian networks, with their graphical components of nodes and links, offered to the ecological-domain expert an easy to use interface to express the interaction between ecological factors. The graphical nature of Bayesian networks also improved the interpretability of the ecological interactions. The ability to create intermediate, or summary nodes, between the model's inputs and outputs allowed the domain expert to reduce the complexity of the problem and allowed the ecologist to focus on only a few interactions at a time.

Bayesian networks have the capacity to be updated and adjusted automatically using data gathered from field observations or air-photo interpretation. Due to a lack of supplementary field data, however, the Bayesian-network knowledge bases developed for each of the three BEC subzone/variants within the study area were defined solely using expert opinion. Similar site series were grouped to compensate for the lack of spatial resolution in the input GIS database required to discern individual site series. Even with this

site-series grouping, however, the accuracy assessment results indicated that the resulting predictive ecosystem maps poorly predicted the dominant site-series group.

Despite the lower than expected results from the accuracy assessment, the Bayesian-network approach could still be considered a viable approach to predict large-scale ecological units given some improvements to the procedure. Future attempts to apply Bayesian networks to predictive mapping of site series should consider applying some of the suggestions summarized below. Changes to each of the modelling stages, GIS database construction, knowledge base construction, and accuracy assessment, could improve site series prediction.

To potentially improve the successful prediction of site series, all GIS database layers should be verified using georeferenced points that are gathered through field observations or air-photo interpretation. Performing exploratory statistical analysis on the relationships between the GIS layers and site series could help reduce knowledge base complexity by identifying those GIS layers exhibiting a relationship with site series.

Model updates could be easier if the knowledge base was designed to predict site series, with site-series groups added as an alternate output node. To aid knowledge base development there should be a representative georeferenced sample of site-series points. Additionally, a representative georeferenced sample of each of the intermediate nodes would be beneficial. These samples could be used to inform the ecologist of the need to change the conditional-probability tables, or to automatically update the conditional-probability tables through the expectation-maximization algorithm.

To ensure that the data set used to test the knowledge base adequately represents the data set used to assess map accuracy, both sets of data should come from the same sampling plan. The data set used to test the knowledge base should be a subset of all the samples, while the data set used to assess map accuracy should be comprised of the remaining samples.

Although the focus of the PEM work conducted in BC over the past six years has involved the use of belief matrices, Bayesian networks can offer a viable and advantageous alternative approach. Further pilot studies incorporating some or all of the recommendations suggested herein should result in large-scale ecological-unit maps that meet or exceed map accuracies obtained through the belief-matrix approach. Exploration of the Bayesian-network approach should continue with an overall goal of producing the most accurate large-scale predictive-ecosystem maps possible for managing forest-related activities.

## 6 REFERENCES

- Atkinson, P. M., and Tatnall, A. R. L. 1997. Neural networks in remote sensing. *International Journal of Remote Sensing* 18:699-709.
- Atticus Resource Consulting Ltd. 2001. Predictive Ecosystem Mapping (PEM) Operational Pilot Project for the Upper Wood River. Unpublished report prepared for Wood River Forest Inc., Revelstoke, BC.
- Bailey, R. G. 1989a. *Ecoregions of the Continents (map)*. U.S. Department of Agriculture, Forest Service, Washinton, DC. Scale 1:30,000,000.
- Bailey, R. G. 1989b. Explanitory supplement to the ecoregions map of the continents. *Environmental Conservation* 15:307-309.
- Banner, A., lePage, P., Shaw, J. & deGroot, A. J. (compilers and editors). 2004. Pattern, Process and Productivity in Hypermaritime Forests of North Coastal British Columbia: The HyP<sup>3</sup> Project. Background and Six Year Results. Unpublished Draft Report, BC Ministry of Forests, Smithers, British Columbia.
- Banner, A., Mackenzie, W., Haeussler, S., Thomson, S., Pojar, J. & Trowbridge, R. 1993. A Field Guide to Site Identification and Interpretation for the Prince Rupert Forest Region. Land Management Handbook Number 26. Government of British Columbia, Ministry of Forests, Victoria, BC.
- Barnes, B. V., Pregitzer, K. S., Spies, K. S., and Spooner, T. A. 1982. Ecological forest site classification. *Journal of Forestry* 80:493-498.
- Beckingham, J. D. & Archibald, J. H. 1996. Field Guide to Ecosites of Northern Alberta. Special Report 5. Natural Resources Canada, Canadian Forest Service Northwest Region, Northern Forestry Centre, Edmonton, Alberta.
- Beckingham, J. D., Corns, I. G. W. & Archibald, J. H. 1996a. Field Guide to Ecosites of West-Central Alberta. Special Report 9. Natural Resources Canada, Canadian Forest Service Northwest Region, Northern Forestry Centre, Edmonton, Alberta.
- Beckingham, J. D., Desilets, M., Nielsen, D., and Johns, F. 1999. Multi-layered and statistically based ecosystem mapping: the de facto standard for land resource planning in the 21st century. *Earth Observation Magazine* .
- Beckingham, J. D., Nielsen, D. & Futoransky, V. A. 1996b. Field Guide to Ecosites of the Mid-Boreal Ecoregions of Saskatchewan. Special Report 6. Natural Resources Canada, Canadian Forest Service Northwest Region, Northern Forestry Centre, Edmonton, Alberta.

- Bell, J. C., Grigal, D. F., and Bates, P. C. 2000. A soil-terrain model for estimating spatial patterns of soil organic carbon. Pages 295-310 in Wilson, J., and J. Gallant, editors. *Terrain analysis: principles and application*. John Wiley and Sons, New York City NY.
- Biggs, W., Eligh, P., Sims, R. & Wiart, R. 1997. A Business Approach to Terrestrial Ecosystem Mapping (TEM) and TEM Alternatives. Government of British Columbia, Resource Inventory Committee, Victoria, B.C.
- Burgess, T. M., and Webster, R. 1980. Optimal interpolation and isarithmic mapping of soil properties; the semi-variogram and punctual kriging. *Journal of Soil Science* 31:315-331.
- Burrough, P. A. 1989. Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science* 40:477-492.
- Campbell, R. W., Dawe, N. K., McTaggart-Cowan, I., Cooper, J. M., Kaiser, G. W., and McNall, M. C. E. 1990. *Birds of British Columbia. Volume 1 Nonpasserines: introduction and loons through waterfowl*. Government of British Columbia, Victoria, BC
- Carpenter, G. A., Gjaja, M. N., Gopal, S., and Woodcock, C. E. 1997. ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data. *IEEE Transactions on Geoscience and Remote Sensing* 35:308-325.
- Carpenter, G. A., Gopal, S., Macomber, S., Martens, S., Woodcock, C. E., and Franklin, J. 1999. A neural network method for efficient vegetation mapping. *Remote Sensing of Environment* 70:326-338.
- Chambers, B. A., Naylor, B. J., Nieppola, J., Merchant, B., and Uhlig, P. 1997. *Field Guide to Forest Ecosystems of Central Ontario*. Ontario Ministry of Natural Resources, Toronto, Ontario
- Ciallela, A. T., Dubayah, R., and Lawrence, W. 1997. Predicting soil drainage class using remotely sensed and digital elevation data. *Journal of Soil Science* 62:171-178.
- Cleland, D. T., Avers, P. E., McNab, H., Jensen, M. E., Bailey, R. G., King, T., and Russell, W. 1997. National Hierarchical Framework of Ecological Units. Pages 181-200 in M. S. Boyce, and A. Haney editors. *Ecosystem Management Applications for Sustainable Forest and Wildlife Resources*; Yale University Press, New Haven, CT.
- Congalton, R. G., and Green, K. 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Lewis Publishers, Boca Raton, Florida

- Congalton, R. G., and Mead, A. 1983. A quantitative method to test for consistency and correctness in photo-interpretation. *Photogrammetric Engineering and Remote Sensing* 49:69-74.
- Congalton, R. G., and Story, M. 1986. Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing* 52:397-399.
- D'Ercole, C., Groves, D. I., and Knox-Robinson, C. M. 2000. Using fuzzy logic in a geographic information system environment to enhance conceptually based prospectivity analysis of Mississippi Valley-type mineralisation. *Australian Journal of Earth Sciences* 47:913-927.
- Davis, J. R. 1993. Experts systems and environmental modelling. Pages 3-35 in Jakeman, A. J., M. B. Beck, and M. J. McAleer, editors. *Modelling change in environmental systems*. John Wiley and Sons Ltd., New York NY.
- Davis, J. R., Hoare, J. R. L., and Nanninga, P. M. 1986. Developing a fires management expert system for Kakadu National Park, Australia. *Journal of Environmental Management* 22:215-227.
- Demarchi, D A. 1995. *Ecoregions of British Columbia*. 4th Edition. Ministry of Environment Lands and Parks, Wildlife Branch, Victoria, B.C.
- Dempster, A., Laird, N., and Rubin, D. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39:1-38.
- Devore, J. L. 2000. *Probability and Statistics for Engineering and the Sciences.*, 5th edition. Duxbury, Pacific Grove, CA
- Dikau, R. 1989. The application of a digital relief model to landform analysis in geomorphology. Pages 51-77 in J. Raper editor. *Three Dimensional Applications in Geographic Information Systems*; Taylor & Francis, London.
- Driscoll, R. S. 1984. An Ecological Land Classification Framework for the United States. Miscellaneous Publication 1439. U.S. Department of Agriculture, Forest Service, Washington DC.
- Ecological Stratification Working Group 1996. A National Ecological Framework for Canada. Environment Canada and Agriculture and Agri-Food Canada, Hull, Quebec.
- Eng, M. & Meidinger, D. V. 1999. A method for large-scale biogeoclimatic mapping in British Columbia. *Web report at: <http://www.for.gov.bc.ca/hre/pubs/pubs/1032.htm>*.

- Fall, A. 2002. SELES Predictive Ecosystem Mapping Support Models. Unpublished.
- Foody, G. M. 1996. Fuzzy modelling of vegetation from remotely sensed imagery. *Ecological Modelling* 85:3-12.
- Foody, G. M. 1999. Applications of the self-organising feature map neural network in community data analysis. *Ecological Modelling* 120:97-107.
- Foody, G. M. 2002a. Hard and soft classifications by neural network with a non-exhaustively defined set of classes. *International Journal of Remote Sensing* 23:3853-3864.
- Foody, G. M. 2002b. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* 80:185-201.
- Freidman, N., Geiger, D., and Goldszmidt, M. 1997. Bayesian network classifiers. *Machine Learning* 29:131-163.
- Hailu, G. and Sommer, G. 1999. On amount and quality of bias in reinforcement learning. Pages 1491-1495 in *IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC'99)*. Tokyo, Japan.
- Huser, B. 1999. Integrated Monitoring: Final Project Report. Environment Waikato, Hamilton East, New Zealand.
- Jenny, H. 1941. *Factors of Soil Formation*. McGraw-Hill, New York
- Jensen, F. V. 1996. Bayesian network basics. *AISB Quarterly* 94:9-22.
- Jones, C. 1999. Virtual Ecosystem Mapping Prototype Accuracy Analysis. Unpublished report prepared for Government of British Columbia Ministry of Forests, Victoria, BC.
- Jones, C. 2003. North Coast Predictive Ecosystem Mapping Report 2002-2003. Unpublished consultant report prepared for Allen Banner, Regional Research Ecologist, Ministry of Forests, Smithers, B.C.
- Jones, C. 2004. Patry Lake Predictive Ecosystem Mapping Final Report. Unpublished report prepared for Slocan Group, Nelson, BC.
- Jones, C. & McGregor, G. 2002. EcoGen Predictive Ecosystem Mapping Report for the Lillooet Forest District. Unpublished.

- Jones, R. K. 1983. Field Guide to Forest Ecosystem Classification for the Clay Belt, Site Region 3e. Ministry of Natural Resources, Toronto, Canada.
- Jones, R. K., Meidinger, D. V., Clark, D., Shultz, F. & Moon, D. 1999. Towards the Establishment of Predictive Ecosystem Mapping Standards: A White Paper. Government of British Columbia, Resource Inventory Committee, TEM Alternatives Task Force, Victoria, B.C.
- Kanellopoulos, I., and Wilkinson, G. G. 1997. Strategies and best practice for neural network image classification. *International Journal of Remote Sensing* 18:711-725.
- Ketcheson, M. V., Dool, T., Kernaghan, G., Lessard, K., Burns, G., Wilson, S. & Smith, G. 2001a. TEMBEC Industries Inc.: Canal Plats Operating Area PEM (Predictive Ecosystem Mapping). Unpublished.
- Ketcheson, M. V., Dool, T. & Wilson, S. 2001b. Arrow TSA Site Index Adjustment Project, Predictive Ecosystem Mapping (PEM), Year Two Final Project Report. Unpublished.
- Ketcheson, M. V., Dool, T. & Wilson, S. 2001c. TFL 3 Predictive Ecosystem Mapping (PEM): Year Two Final Report. Unpublished.
- Kimes, D. S., Holben, B. N., Nickeson, J. E., and McKee, A. 1996. Extracting forest age in a Pacific Northwest forest from thematic mapper and topographic data. *Remote Sensing of Environment* 56:133-140.
- King, D., Bourennane, H., Isambert, M., and Macaire, J. J. 1999. Relationships of the presence of a noncalcareous clay-loam horizon to DEM attributes in a gently sloping area. *Geoderma* 89:95-111.
- Klappstein, G. D. & Corns, I. G. W. 1996. Field Guide to Ecosites of Southwestern Alberta. Special Report 8. Natural Resources Canada, Canadian Forest Service Northwest Region, Northern Forestry Centre, Edmonton, Alberta.
- Klinka, K., and Skoda, L. 1977. Synecological Map of the University of British Columbia Research Forest. *The Forestry Chronicle* 53:348-352.
- Krige, D. G. 1963. Two dimensional weighted moving average trend surfaces for ore-evaluation. *Journal of the South African Institution of Mining and Metallurgy* 66:13-38.
- Linderman, M., Liu, J., Qi, J., An, L., Ouyang, Z., Yang, J., and Tan, Y. 2004. Using artificial neural networks to map the spatial distribution of understory bamboo from remote sensing data. *International Journal of Remote Sensing* 25:1685-1700.

- MacMillan, R. A., Pettapiece, W. W., Nolan, S. C., and Goddard, T. W. 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy Sets and Systems* 113:81-109.
- Major, J. 1951. A functional, factorial approach to plant ecology. *Ecology* 32:392-412.
- Marcot, B. G., Holthausen, M. G., Raphael, M. G., Rowland, M., and Wisdom, M. 2001. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest Ecology and Management* 153:29-42.
- Marshall, I. B. & Schuts, P. H. 1999. A National Ecological Framework for Canada: Overview. Environment Canada and Agriculture and Agri-Food Canada, Hull, Quebec.
- McBratney, A. B., Hart, G. A., and McGarry, D. 1991. The use of region partitioning to improve the representation of geostatistically mapped soil attributes. *Journal of Soil Science* 42:513-532.
- McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S., and Shatar, T. M. 2000. An overview of pedometric techniques for use in soil surveys. *Geoderma* 97:293-327.
- McCracken, R. J., and Cate, R. B. 1986. Artificial intelligence, cognitive science, and measurement theory applied in soil classification. *Soil Society of America Journal* 50:557-561.
- Meidinger, D. V. 2003. Ecosystem Mapping Accuracy and Timber Supply Applications. BC Ministry of Forests, Research Branch, Victoria, BC.
- Meidinger, D. V. 2004. North Coast PEM Project Quality Assurance. Unpublished Draft Report, BC Ministry of Forests, Victoria, BC.
- Meidinger, D. V. & Pojar, J. 1991. Ecosystems of British Columbia. Special Report Series No.6. B.C. Ministry of Forests, Victoria, B.C.
- Miller, J., and Franklin, J. 2002. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependencies. *Ecological Modelling* 157:227-247.
- Mitchell, W. R. & Eremko, R. 1987. Ecosystem Mapping of the Truax Creek Basin in the Kamloops Forest Region. Research Report RR87007-KL. BC Ministry of Forests and Lands, Victoria, BC.

- Moon, D. 1999. Situation Analysis for Knowledge-Based Systems. Government of British Columbia, Resource Inventory Committee, TEM Alternatives Task Force, Victoria, B.C.
- Moore, D. M., Lees, B. G., and Davey, S. M. 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. *Environmental Management* 15:59-71.
- Moore, I. D., Gessler, P. E., Nielsen, G. A., and Peterson, G. A. 1993. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* 57:443-520.
- Mulder, J. A. and Corns, I. G. W. 1994. Knowledge Based Ecosystem Prediction: Field Testing and Validation. Pages 829-835 in *GIS '94 Symposium*. Vancouver, BC.
- Norsys. 1997. *Netica Application: User's Guide, version 1.05*. Norsys Software Corp., Vancouver, B.C.
- Omernik, J. M. 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geographers* 77:118-125.
- Pack, R. T., Tarboton, D. G., and Goodwin, C. N. 1998. The SINMAP approach to terrain stability mapping. in *8th Congress of the International Association of Engineering Geology, Vancouver, British Columbia, Canada, 21-25 September 1998*.
- Pal, M., and Mather, P. M. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment* 86:554-565.
- Park, S. J., and Vlek, P. L. G. 2002. Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques. *Geoderma* 109:117-140.
- Pearl, J. 1988. Probabilistic reasoning in intelligent systems: networks of plausible inference. in *Series in Representation and Reasoning*; Morgan Kaufmann Publishers, San Francisco, CA.
- Plach, M. 1999. Bayesian networks as models of human judgement under uncertainty: the role of causal assumptions on belief updating. *Kognitionswissenschaft* 8:30-39.
- Pojar, J., Klinka, K., and Meidinger, D. V. 1987. Biogeoclimatic ecosystem classification in British Columbia. *Forest Ecology and Management* 22:119-154.

- Pojar, J., Meidinger, D. V., and Klinka, K. 1991. Chapter 2: Concepts. Pages 9-38 in Meidinger, D. V., and J. Pojar, editors. *Ecosystems of British Columbia. Special Report Series 6*. BC Ministry of Forests, Victoria, BC.
- Qi, F., and Zhu, A.-X. 2003. Knowledge discovery from soil maps using inductive learning. *International Journal of Geographical Information Science* 17:77-795.
- Renooij, S. 2001. Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review* 16:255-269.
- RIC 1998. Standard for Terrestrial Ecosystem Mapping in British Columbia. Government of British Columbia, Resource Inventory Committee, Terrestrial Ecosystem Task Force, Victoria.
- RIC 1999. Standards for Predictive Ecosystem Mapping - Inventory Standards. Government of British Columbia, Resource Inventory Committee, Terrestrial Ecosystems Task Force, Victoria, B.C.
- Rodriguez, F., Maire, E., Courjault-Radé, P., and Darrozes, J. 2002. The Black Top Hat function applied to a DEM: a tool to estimate recent incision in a mountainous watershed (Estibère Watershed, Central Pyrenees). *Geophysical Research Letters* 29:9-1-9-4.
- Rosen, D., Blashill, W. & Coderre, M. 2001. Predictive Ecosystem Mapping for Canfor's TFL48 (Chetwynd). Unpublished.
- Schmidt, J., and Hewitt, A. 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. *Geoderma* in press.
- Scull, P., Franklin, J., Chadwick, O. A., and McArthur, D. 2003. Predictive soil mapping: a review. *Progress in Physical Geography* 27:171-197.
- Sims, R., Towill, W. D., Baldwin, K. A., and Wickware, G. M. 1989. *Field Guide to the Forest Ecosystem Classification for Northwestern Ontario*. Ontario Ministry of Natural Resources, Toronto, Ontario
- Skidmore, A. K. 1989. An expert system classifies eucalypt forest types using Landsat Thematic mapper data and digital terrain model. *Photogrammetric Engineering and Remote Sensing* 55:1449-1464.
- Skidmore, A. K., Ryan, P. J., Dawes, W., Short, D., and O'Loughlin, E. 1991. Use of an expert system to map forest soils from a geographical information system. *International Journal of Geographical Information Science* 5:431-445.

- Skidmore, A. K., Watford, F., Luckananurug, P., and Ryan, P. J. 1996. An operational GIS expert system for mapping forest soils. *Photogrammetric Engineering and Remote Sensing* 62:501-511.
- Stock, M. 1987. AI and expert systems: an overview. *AI Applications in Natural Resources, Agriculture and Environmental Science* 1:9-17.
- Sulyma, R. & Alward, B. 2004. Blackwater PEM Retrofit Accuracy Assessment. Forest Floor Contracting Ltd., Fort St. James, BC.
- Teng, C. H., and Fairbairn, D. 2000. Comparing expert systems and neural fuzzy systems for object recognition in map object revision. *International Journal of Remote Sensing* 23:555-567.
- Timberline Forest Inventory Consultants Ltd. 2000. An Application of the Standards for Predictive Ecosystem Mapping, Version 1.0: A Report on the Targe Creek - Sutherland PEM Project. Unpublished report prepared for the BC Ministry of Environment, Lands and Parks, Victoria, BC.
- Timberline Forest Inventory Consultants Ltd. 2004. Blackwater Predictive Ecosystem Mapping Retrofitting. Unpublished report prepared for Slocan Forest Products Ltd., Mackenzie, BC.
- Udvardy, M. D. F. 1975. A Classification of the Biogeographical Provinces of the World. IUCN Occasional Paper No. 18. International Union for Conservation of Nature and Natural Resources, Morges, Switzerland.
- Varis, O., and Kuikka, S. 1997. BENE-EIA: A Bayesian approach to expert judgement elicitation with case studies on climate change impacts on surface waters. *Climate Change* 37:539-563.
- Walter, H., and Box, E. 1976. Global classification of natural terrestrial ecosystems. *Vegetatio* 32:75-81.
- Webster, R. 1994. The development of pedometrics. *Geoderma* 62:1-15.
- Wiken, E. B. 1986. Terrestrial Ecozones of Canada. Ecological Land Classification Series No. 19. Environment Canada, Hull, Quebec.
- Wood, J. 1996. *The geomorphological Characterization of Digital Elevation Models*. Ph.D. Thesis. Department of Geography, University of Lancaster, UK.
- Zadeh, L. A. 1965. Fuzzy sets. *Information and Control* 8:338-353.

- Zelazny, V. F., Ng, T. T. M., Hayter, M. G., Bowling, C. L., and Bewick, D. A. 1989. *Field Guide to Forest Site Classification in New Brunswick: Harvey - Harcourt and Fundy Site Regions*. New Brunswick Ministry of Natural Resources and Energy, Fredericton, N.B.
- Zhang, J., and Foody, G. M. 2001. Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches. *International Journal of Remote Sensing* 22:615-628.
- Zhu, A.-X. 1997. Measuring uncertainty in class assignment for natural resource maps under fuzzy logic. *Photogrammetric Engineering and Remote Sensing* 63:1195-1202.
- Zhu, A.-X., Hudson, B., Burt, J., Lubich, K., and Simonson, D. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal* 65:1463-1472.
- Zhu, X., Healey, R. G., and Aspinall, R. J. 1998. A knowledge-based system approach to design of spatial decision support systems for environmental management. *Environmental Management* 22:35-48.
- Zolaneski, C. A., Wickwaer, G. M., Delorme, R. J., Sims, R. & Corns, I. G. W. 1995. *Forest Ecosystem Classification for Manitoba: Field Guide*. Special Report 2. Natural Resources Canada, Canadian Forest Service Northwest Region, Northern Forestry Centre, Edmonton, Alberta.