Utah State University

DigitalCommons@USU

5-1999

Predictively Mapping the Plant Associations of the North Fork John Day Wilderness in Northeastern Oregon Using Classification Tree Modeling

Alison M. Kelly
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Part of the Biology Commons

Recommended Citation

UtahState University
MERRILL-CAZIER LIBRARY

PREDICTIVELY MAPPING THE PLANT ASSOCIATIONS OF THE NORTH FORK

JOHN DAY WILDERNESS IN NORTHEASTERN OREGON USING

CLASSIFICATION TREE MODELING

by

Alison M. Kelly

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Biology

Approved:

Dr. Noelle E. Cockett
Interim Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

1999

ABSTRACT

Predictively Mapping the Plant Associations of the North Fork John Day Wilderness

in Northeastern Oregon Using Classification Tree Modeling

by

Alison M. Kelly, Master of Science

Utah State University, 1999

Major Professor: Dr. Mary E. Barkworth
Department: Biology

Shifting perspectives on restoration and management of public lands in the inland West have resulted in an increased need for maps of potential natural vegetation which cover large areas at sufficient scale to delineate individual stands. In this study, classification tree modeling was used to predictively model and map the plant association types of a relatively undisturbed wilderness area in the Blue Mountains of northeastern Oregon. Models were developed using field data and data derived from a geographic information system database. Elevation, slope, aspect, annual precipitation, solar radiation, soil type, and topographic position were important predictor variables. The model predicted plant association types with a relatively high degree of accuracy for most plant association types, with the lowest accuracy for the types within the grand fir series. Fuzzy confusion analysis was used to analyze model performance, and indicated the overall model accuracy was 72%.

(69 pages)

ACKNOWLEDGMENTS

This thesis could not have been completed without full support from the Umatilla National Forest. Karl Urban supported this work from its inception and provided anything and everything that was needed. Kari Yanskey greatly assisted me by identifying plant specimens and by helping me negotiate the forest GIS database. The friendship and support of all the botanists of the Umatilla are greatly appreciated. Additionally, I would like to thank Charles Johnson for sharing plot data.

I would like to thank my committee, Drs. Mary Barkworth, Ted Evans, Janis Boettinger, and David Roberts, for all their support throughout this process. I would especially like to thank Dave Roberts for his patience, his assistance with all manner of software problems, and his infinite interest in the forests of eastern Oregon.

Warm thanks goes to the Kelly family, who supported and encouraged me continually. I would also like to thank Olivia Messinger, Kim Pierson, Joe Mendelson, and all my friends who kept me laughing, never lost patience, and helped in every way imaginable.

This work is dedicated to the late Karl Urban.

Alison Kelly

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Current concerns over declines in forest health and the anticipated effects of

global climate change have led to an increased interest in the spatial representation of

vegetation at the landscape, regional, and global scale. Unfortunately, traditional

methods of vegetation mapping are too expensive and time-consuming to implement over

large areas. The development of geographic information systems (GIS) and other tools

for spatial analysis and visualization and the recent combination of these tools with

predictive models have allowed for maps of predicted vegetation to be created over large

areas (e.g., Moore et al. 1991; Brzeziecki et al. 1993; Franklin 1998).

A major focus of some recent mapping efforts is in representing the climax or

potential natural vegetation (PNV) at a landscape scale for the purposes of environmental

management, restoration, or investigating vegetation-environment relationships (Moore

et al. 1991; Lees & Ritman 1991; Palmer 1991; Lynn et al. 1995). On a landscape level,

the factors affecting the distribution of PNV are those which influence the amount of

water, solar radiation, and nutrients available to plants. Because of this, the PNV

supported by a site is an indication of the natural conditions operating there (Pfister &

Arno 1980; Kalkhoven & van der Werf 1988; Brzeziecki et al. 1993) and can be thought

of as an index of site potential (Layser 1974). Representing PNV spatially in a GIS is

valuable for land management decisions in complex or mountainous terrain, allows for

manipulation of the data, and permits the combining of these data with other spatially

explicit models.

In order to predictively model and map the PNV for an area, there must be

sufficient representation of mature or climax vegetation from which to draw information about correlated environmental variables. There must also be spatially interpolated values for these correlated environmental variables for the area, preferably stored in a GIS. If both of these conditions are met, then it is possible to extrapolate small, point-based field observations of mature vegetation types within a study area to predict the PNV over the entire study area.

The objectives of this research were to develop, analyze, and spatially represent a predictive model for the potential natural vegetation of a small, relatively undisturbed wilderness area in northeastern Oregon. This area has experienced a minimum of disturbance due to humans, has many representative stands of mature vegetation, and is included in a fairly extensive GIS database managed by the Umatilla National Forest. Little is known about the relationships between the vegetation of the Blue Mountains and the controlling climatic and edaphic variables, and there are current concerns of the health of forests in the region. Therefore, this is an ideal site for such a mapping effort.

OVERVIEW

The goal of predictively mapping potential natural vegetation is relatively new and is related to the increasingly widespread use of GIS and other tools for spatial analysis. However, many of the concepts underlying the idea are derived from a more established, yet dynamic, theoretical base. This section begins with a short definition and background of the notion of potential natural vegetation. Next, because predicting the distribution of vegetation requires an understanding of its correlated or controlling environmental factors, an overview of vegetation response analysis is given. Then, a brief summary of predictive vegetation modeling concepts and methods is presented, with comments on the current trends. Finally, a comment on the selection of predictor variables used in predictive modeling is given.

Potential Natural Vegetation

The potential natural vegetation of an area can be defined as the vegetation that would develop at a site if left without human disturbance for a sufficiently long period of time. Reaching a dynamic equilibrium with the current climatic, hydrological, and edaphic conditions, the site will eventually consist of those species that are able to establish, persist, and reproduce. This probable assemblage of species, projected from existing conditions, is referred to as the potential natural vegetation (Tüxen 1956 in Kalkhoven & van der Werf 1988). Thus, while the existing plant community can be indicative of a site's PNV, the two often differ (Kalkhoven & van der Werf 1988). The basic unit of classification of the PNV in the Rocky Mountains is traditionally termed the "habitat type," which "is a collective term for those physical environments capable of

supporting a particular climax plant association" (Layser 1974, p. 354). In the Pacific Northwest Region, the Forest Service uses the term "plant association" instead of habitat type, and this convention has been adopted for this study.

Classifications of plant associations were first employed in the western U.S. by Daubenmire (1952), and have since been used commonly to categorize the environment across much of the inland West (Pfister & Arno 1980). The advantages of PNV classifications over earlier and more generalized methods, evidence of the increased acceptance of the method by the early 1970s, and applications of these classifications are reviewed by Layser (1974). Because the PNV is an indication of the natural conditions functioning at a site, mapping PNV over landscapes can be useful for ecological research and land management (Layser 1974; Pfister & Arno 1980; Kalkhoven & van der Werf 1988; Brzeziecki et al. 1993). Producing maps of PNV or existing vegetation over large areas in the traditional manner (Deitschman 1973) is costly and time-consuming, and recent efforts have focused on developing predictive models which can then be linked to a geographic information system (GIS) for spatial representation.

Vegetation Response Analysis

The basis in ecological theory for predictive vegetation modeling is the concept of the niche, defined by Hutchinson (1957) as the hypervolume, determined by environmental factors, in which a species can survive and reproduce. In nature, biotic forces such as competition reduce this hypervolume, which is then referred to as the "realized niche" or, more commonly in vegetation response analysis, "species response volume" (Franklin 1995). Drawing from Gleason's (1926) individualistic concept of a

plant species' unique response to the environment, work in the 1950s and 1960s focused on community-level, continuous distributions of plant species along environmental gradients. This continuum concept has come to be the generally accepted model of plant community structure, and communities are seen as continuously varying along the continuum. Until recently, the shape of a species response along a gradient was considered to be bell-shaped and unimodal (Whittaker 1956; Gauch & Whittaker 1972). However, recent empirical evidence shows that skewed and bimodal responses occur more commonly than normal responses (Austin et al. 1990; Collins & Glenn 1990). These findings mean that the statistical assumptions of many tools used in vegetation response analysis have been violated, and consequently nonparametric approaches to modeling plant response surfaces have been employed with increasing frequency (Bio et al. 1998).

The major tools for analyzing vegetation response to the environment fall into three main categories: ordination, regression techniques, and machine-learning methods. Ordination methods used in plant community analysis include principal components analysis (PCA) (Bradfield & Scagel 1984), detrended correspondence analysis (DCA) (Hill & Gauch 1980), nonmetric multidimensional scaling (NMDS) (Whittaker 1987), and canonical correspondence analysis (CCA) (ter Braak 1987). These represent techniques which allow for both linear (PCA) and nonlinear (DCA, NMDS, CCA) responses to the environment. Regression techniques include general linear models, generalized linear models (GLMs) introduced by Austin et al. (1984), and the non-parametric extension of these, the generalized additive models (GAMs). Of these, the GAM is the least restrictive, allowing for both linear and complex vegetation responses,

and it includes smoothers that can fit any shape of response curve (Bio et al. 1998). For these reasons, GAMs have been employed in several recent studies to determine species response surfaces (Austin et al. 1990; Bio et al. 1998) and to predict plant species distribution by linking these models with a GIS (Austin et al. 1994c; Brown 1994; Franklin 1998).

The machine-learning methods used in exploring vegetation-environment relationships are neural networks and classification and regression trees. Presumably because they require skill and experience to use successfully (Austin et al. 1994b), neural network techniques are not used widely in vegetation modeling. Much more frequently used in predictive vegetation modeling, classification and regression trees (CART) were developed by Breiman (1984). Suggested for use in vegetation analysis by Verbyla (1987), CART was previously applied in optimization and prediction in medicine and industry. Although the conditional rules of CART models make it difficult to examine species response surfaces (Austin et al. 1994a), they do provide reasonable curves that are similar, and in some cases, "at least as good" as those produced by GLM and GAM models (Austin et al. 1994a). Of the methods of vegetation response analysis mentioned above, the GLM, GAM, and CART models are currently the most commonly employed in predictive vegetation modeling and mapping.

Predictive Modeling: Concepts and Methods

The concept of combining vegetation analysis methods with a GIS to represent predictions spatially was introduced by Kessell (1979; in Franklin 1995). Since that time, a set of different techniques has been used predictively with many different mapping

goals, including representing the existing vegetation, PNV, plant species richness, diversity, and vegetation structure. A full review of this topic can be found in Franklin (1995); presented here is a brief outline of some of the trends in predictive vegetation modeling and mapping and some comments on more recent work.

Some of the earliest predictive vegetation models were based on Boolean logic. These predict one class per location based on ranges of explanatory data variables, and these data ranges can be modeled or actually observed. A number of statistical models have also been used. Earlier examples of parametric models would include maximum likelihood models, linear regression, discriminant analysis, logistic regression models, and GLMs. Because vegetation responses to the environment have been found to be nonlinear and nonnormal, the assumptions of these parametric models are violated. Recently, the nonparametric GAMs have been used with increasing frequency, and these models, like classification trees, predict the probability of class membership. Lastly, machine-learning type models such as expert systems and CART have been employed to predict vegetation distribution. Expert systems, where someone very familiar with the vegetation and ecology of an area subjectively makes predictive rules to determine the distribution of vegetation, are often time-consuming to develop and not repeatable.

The CART method is advantageous because it is nonparametric, tends to be robust to outliers (Verbyla 1987), captures hierarchical relationships well (Michaelson et al. 1994), and provides output in the form of a dichotomous key that is easy to implement in the field. Because it is difficult to extend GAMs beyond a binary response such as presence or absence of a species, CART is advantageous in certain modeling efforts, as it handles multiple possible classes directly. Classification trees differ from regression

trees in that the response variable is categorical, such as a plant association, and the predictions are class membership probabilities for these categories; regression trees predict average values and the response variables are continuous. This study focuses primarily on classification trees.

The first use of classification trees in predictive vegetation modeling and mapping was by Moore et al. (1991), where a tree classifier was linked with a GIS and subsequently used to predictively model and map 30 forest types in southeastern Australia. Lynn et al. (1995) used classification tree modeling to predict PNV in central New York State, and Skidmore et al. (1996) employed this method in predicting kangaroo habitat across Australia. Recently, comparisons of results from CART, GLM, and GAM models have been undertaken. Austin et al. (1994 a,b) found that, overall, GAMs performed best in predicting several *Eucalyptus* species distributions, but the tree models in this study were not pruned as suggested by Clark and Pregibon (1992). Franklin (1998) compared the predictions of these models for 20 species of shrubs in southern California. For each species, the three models generated all had similar levels of accuracy, but classification trees "yielded the lowest prediction errors (lower by 3-5%)" (Franklin 1998, p. 733).

Selection of Predictor Variables

The major factors affecting the distribution of potential natural vegetation are those which influence the amount of water, solar radiation, and nutrients available to plants. Over regional scales, climate exerts the greatest control over the pattern of vegetation (Covington et al. 1994). On smaller scales where climatic conditions do not

vary widely, such as the landscape level, topography and parent material most greatly affect site moisture, radiation, and nutrient levels and therefore plant distribution. Austin et al. (1984) define two major types of gradients based on these ideas. "Direct gradients" are those which have a direct physiological effect on a plant, such as amount of solar radiation or available water. "Indirect gradients," the second type of gradient, are those that represent a number of interrelated factors that act together to affect a resource needed by a plant and are often location-specific. Examples would include aspect, elevation, and topographic position. As the direct gradients of moisture and nutrient availability are often difficult to measure in the field, surrogate factors which affect these gradients indirectly are used to predict PNV (Roberts & Cooper 1987; Moore et al. 1991). For example, the moisture budget of a site can be influenced by a number of topographic and edaphic factors, including slope curvature, topographic position, surface and subsurface soil texture, and soil depth. Moore et al. (1991) comment that the use of indirect gradients in predicting vegetation can result in a complex model, but further note that exclusive use of direct gradients may not be feasible "as it would require prior specification of the relationships between the direct gradients and the many topographic and edaphic variables that determine these" (Moore et al. 1991, p. 60).

In this study, both direct and indirect factors were used to predict PNV over the study area. Recent radiation modeling efforts (Dubayah 1994) have resulted in good estimates of solar radiation over landscapes, and these data were used in this study. Improved spatial hydrologic models are also being developed (e.g., Tarboton 1997), which aid in determining moisture gains and losses due to the flow of water over uneven terrain. Future work could be done based on Moore et al. (1993a) in predicting soil

attributes over landscapes based on digital elevation models. Interpolating direct soil-related gradients could be very useful in predictive vegetation mapping, as it may overcome some of the current limitations of traditional soil maps, which can have large variation within mapping units for many properties which affect plants.

MATERIALS AND METHODS

Study Site

The North Fork John Day Wilderness (Figure 1) was established by the Oregon Wilderness Act of 1984. About 344 km^2 in size, it includes parts of both the Greenhorn and Elkhorn Ranges of the Blue Mountains in northeastern Oregon. The area is highly variable with respect to physiography and parent material. The Wilderness is dissected by the generally east-west flowing North Fork John Day River, along which the lowest elevations (about 1050 m) can be found. Surrounding the steep river canyon are rolling benches, mountain meadows, rocky side canyons, and abrupt ridges. The highest point in the study site is Desolation Butte, which rises to 2135 m.

The Blue Mountains were formed by the collision of oceanic and continental plates during the early Triassic period, about 200 million years ago. The subducting ocean plate added material to the then coastal mountains, which were composed of sedimentary rocks of the coastal plain and from the ocean floor. Extensive volcanism followed, covering large areas with basalt material until around 35 million years ago. At this time, granitic intrusions formed further inland in smaller areas. Another period of volcanism occurred in the Miocene, about 20 to 25 million years ago, which added more basaltic rock to northeastern Oregon (Alt & Hyndman 1978). Major volcanic activity in the Blue Mountains ended 12 million years ago. During the Pleistocene, two million years ago, alpine glaciation occurred in the highest peaks of the Greenhorn, Elkhorn, and Strawberry ranges and resulted in redistribution of upper elevation geologic materials, including intrusive granite. The major parent materials evident in the Blue Mountains
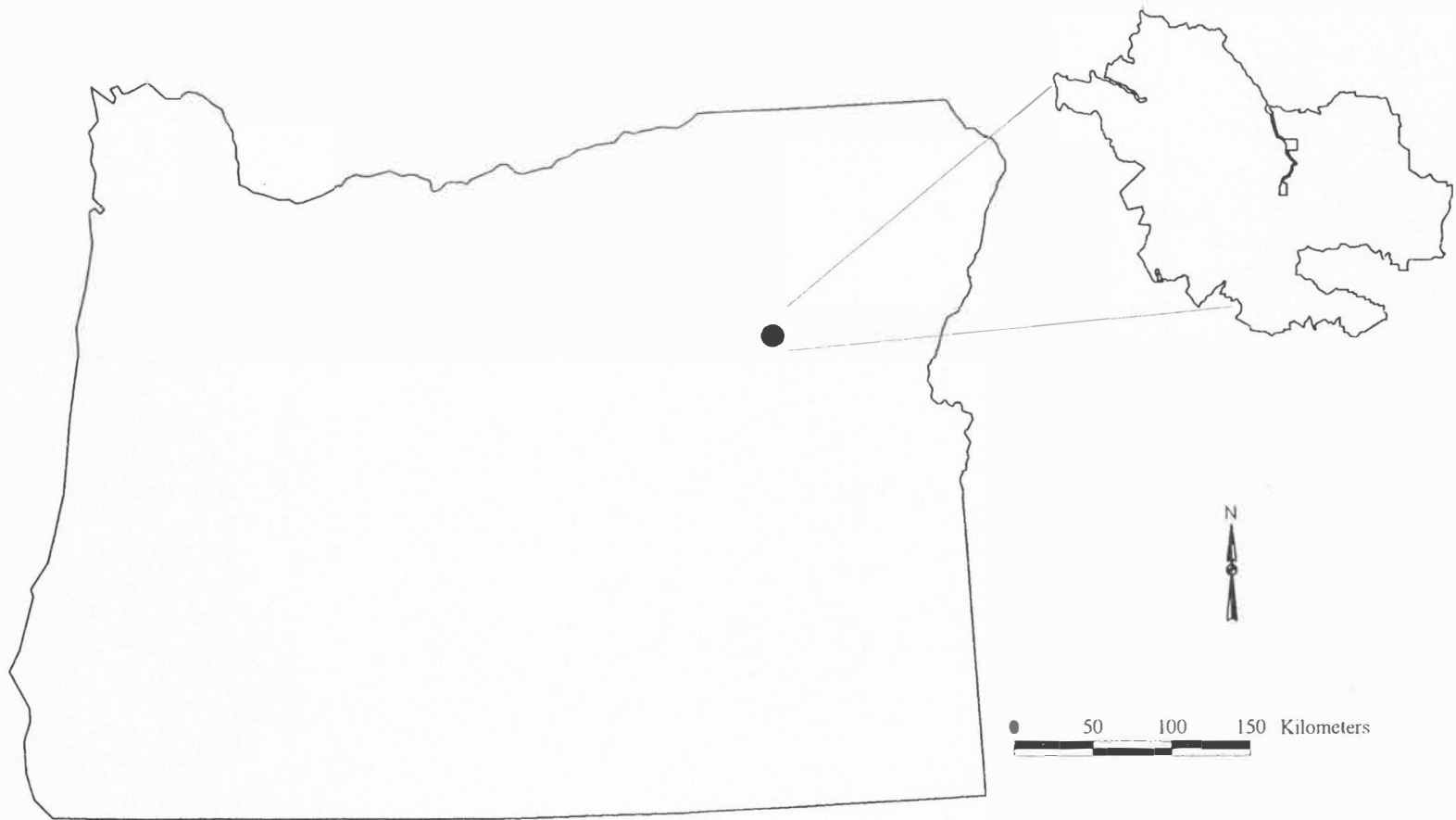
Fig. 1. Location of the North Fork John Day Wilderness in Oregon (projection in Universal Transverse Mercator).

today include marine sedimentary rock, meta-volcanics, granite, and basalt (Ehmer 1978).

The diversity of parent materials contributes to the variety of soils found in the North Fork John Day Wilderness. Additionally, the eruptions of both Glacier Peak (12,000 years ago) and Mount Mazama (6,800 years ago) resulted in the deposition of ash over most of the Blue Mountains. This ash was later redeposited by wind and water (Johnson & Clausnitzer 1992). The silt-sized ash is a major influence on many of the soils of the region. In contrast to the droughty basalt-derived soils, ash-influenced soils tend to have higher water-holding capacities and fewer coarse fragments in the rooting zone. Thus, these soils tend to hold more water during the dry summer months characteristic of this region (Geist & Strickler 1978). Across the Wilderness, soils vary from shallow, coarse soil mapping units with basaltic or granitic parent materials to units with deep clays or clay loams derived from tuffs and pyroclastics (Ehmer 1978). Specific soil properties for each mapping unit found within the Wilderness are listed in Appendix A.

Most of the precipitation in this part of the Blue Mountains occurs during the winter months. Local relief influences both precipitation and temperatures, with higher elevations being both colder and wetter than the values reported here, which were recorded in Ukiah, Oregon, at approximately 1100 m. The 30-year (1966 to 1996) average annual snowfall is 97.8 cm and less than one-fifth of the mean annual precipitation of 41.6 cm falls in June, July, and August. Temperatures vary widely, with annual average highs of 14.8 °C and lows of -2.3 °C (Anon. 1999).

The potential natural vegetation in the Wilderness includes plant associations

within the subalpine fir (*Abies lasiocarpa* (Hook.) Nutt.), grand fir (*A. grandis* (Dougl. ex D. Don) Lindl.), Douglas-fir (*Pseudotsuga menziesii var. glauca* (Beissn.) Franco), lodgepole pine (*Pinus contorta var. latifolia* Engelm. ex S. Wats.), and ponderosa pine (*Pinus ponderosa* P.& C. Lawson) series. Subalpine fir associations occur only at fairly high elevations within the study area, generally above 1800 m. Stands within this series at lower elevations tend to occur in frost pockets (Franklin & Dyrness 1988). In general, these stands are indicative of the coolest and wettest sites in the Wilderness. The desiccation of sites by wind has been noted as one factor limiting the distribution of subalpine fir plant associations (Johnson et al. 1994).

Occurring at lower elevations than subalpine fir, the grand fir series is widespread throughout the study site and is especially common on northerly aspects and in sheltered canyons. Grand fir plant associations comprise the most extensive midslope forest zone in the Blue Mountains. They typically occur on volcanic ash soils (Franklin & Dyrness 1988). The common plant associations within the series range from characteristically mesic types with twinflower (*Linnaea borealis* L.) and queen's cup beadlily (*Clintonia uniflora* (Schult.) Kunth.) in the understory to drier types with pinegrass (*Calamagrostis rubescens* Buckl.) present in the herbaceous layer. Intense fires at intervals of less than 150 years tend to result in dense, seral stands dominated by lodgepole pine on sites supporting this species (Agee 1994).

Douglas-fir plant associations are also widespread but tend to occupy slightly drier and warmer sites than the grand fir. Several of the associations within this series commonly occur on steep canyon slope positions (Agee 1994). While not always locally

abundant in the Blue Mountain region (Franklin & Dyrness 1988), many stands within the Douglas-fir series can be found within the Wilderness.

Found mainly on southerly aspects or rolling topography at elevations below 1600 m, ponderosa pine plant associations are open, parklike stands. These climax communities tend to occur on "coarse, sandy soils and where fissures in the underlying bedrock permit the tree to tap deep moisture sources" (Agee 1994). Franklin and Dyrness (1988) report that the soil moisture regime is the most important factor in determining the plant association within this series. This vegetation type is the most stable with regard to disturbance in the Blue Mountain landscape (Agee 1994, p. 30).

The lodgepole pine series is represented in the Blue Mountains by a single plant association, the lodgepole pine/pinegrass association. This type is thought to be a topographic and edaphic climax, following Tansley's (1935) polyclimax concept. It occurs only in frost pockets where cold air tends to accumulate. These sites are characterized by frosty, wet soils that are unsuitable for survival and reproduction of other tree species (Johnson & Clausnitzer 1992). Because soil moisture is important for lodgepole pine seeding establishment, this plant association is never found on steep south-facing slopes (Agee 1994). While seral lodgepole stands occur over many parts of the Wilderness, the climax plant association is uncommon. Also uncommon, wet meadows and scattered, dry grasslands dominated by Sandberg bluegrass (*Poa secunda* J. Presl), occur only over small areas of the study site.

Database

Both field observations and data from GIS layers were used to develop the database for input into the classification tree model. Field observations were collected in a stratified random fashion during the summers of 1996 and 1997. Plant association point data were recorded when representative stands were located and were found to correspond to the plant associations described in Johnson and Clausnitzer (1992). For each of the 200 known data points across the Wilderness (Figure 2), the plant association type was determined according to Johnson and Clausnitzer (1992). Meadow plant associations, not addressed in Johnson and Clausnitzer (1992), were recorded as "meadow" types. In the field, the points were georeferenced using a hand-held GPS, and the slope, aspect, and elevation were found using a clinometer, a compass, and a 7.5" topographic map, respectively. In order to minimize errors associated with inaccurate GPS readings, location coordinates were always determined at least 100 m away from any transition in vegetation type.

At representative plant associations, a total of 26 inventory plots were designated to further characterize plant species composition and soil attributes of the major plant associations of the Wilderness. Each type was represented by a least three inventory plots. Circular plots with a radius of 11.32 m (0.1 acre in area) were established, and coverage percentage of all vascular plant species was determined visually according to Johnson and Clausnitzer (1992). Voucher specimens of plant species were collected, dried in a standard plant press, and deposited in the Intermountain Herbarium at Utah State University. Species identifications were made using Hitchcock and Cronquist (1973), and specimens of difficult taxa were verified by two botanists of the Umatilla
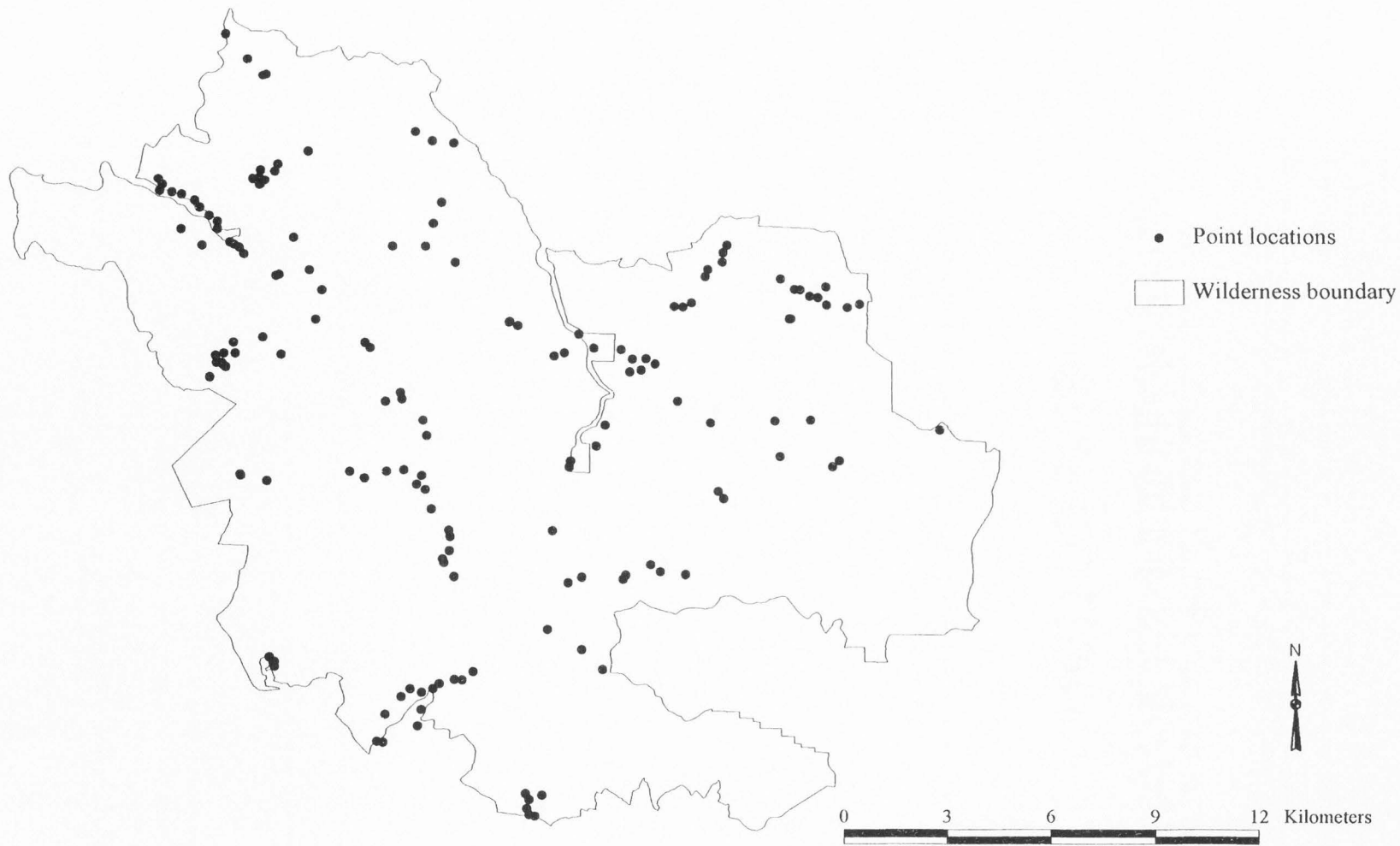
Fig. 2. Map of the 200 known plant association data points within the study area.

National Forest, Karl Urban and Kari Yanskey. Taxonomic circumscriptions and name changes were carefully considered and nomenclature follows the USDA PLANTS database. At each plot site, soils were characterized using abbreviated pit descriptions. Soil pits were excavated to bedrock or to a depth of 120 cm, whichever was shallower. For each horizon, moist and dry colors were determined using the Munsell notation, and the texture and clay content were found using the texture-by-feel method. Also, dry, moist, and wet consistence was recorded, the pH for each horizon was determined using colorimetry, and the volume of coarse fragments was visually estimated. Samples of soils from each horizon for each site were also air-dried at room temperature and stored for future analysis. These soil data were intended to supplement data for mapping units in the Umatilla National Forest Soil Resource Inventory, and to aid in the determination of soil characteristics with predictive value for plant associations.

The GIS layers used in the development of the database were provided by the Umatilla National Forest and manipulated using ARC/Info Version 7.1.2 on a PC with the Windows NT operating system. The digital elevation models (DEMs) for each quadrangle representing part of the Wilderness were merged to create a single elevation coverage (Figure 3). The DEMs and all other GIS layers utilized have a resolution of 30 m. From the DEMs, slope and aspect were calculated using ARC/Info. Aspect in degrees is not handled well by classification tree model, as values such as 1° and 359° both represent north-facing slopes but are very different numerically. Because of this, these data were transformed into aspect values using the equation

$$\text{aspect value} = (\cos(\text{aspect}° - 30) + 1) / 2$$
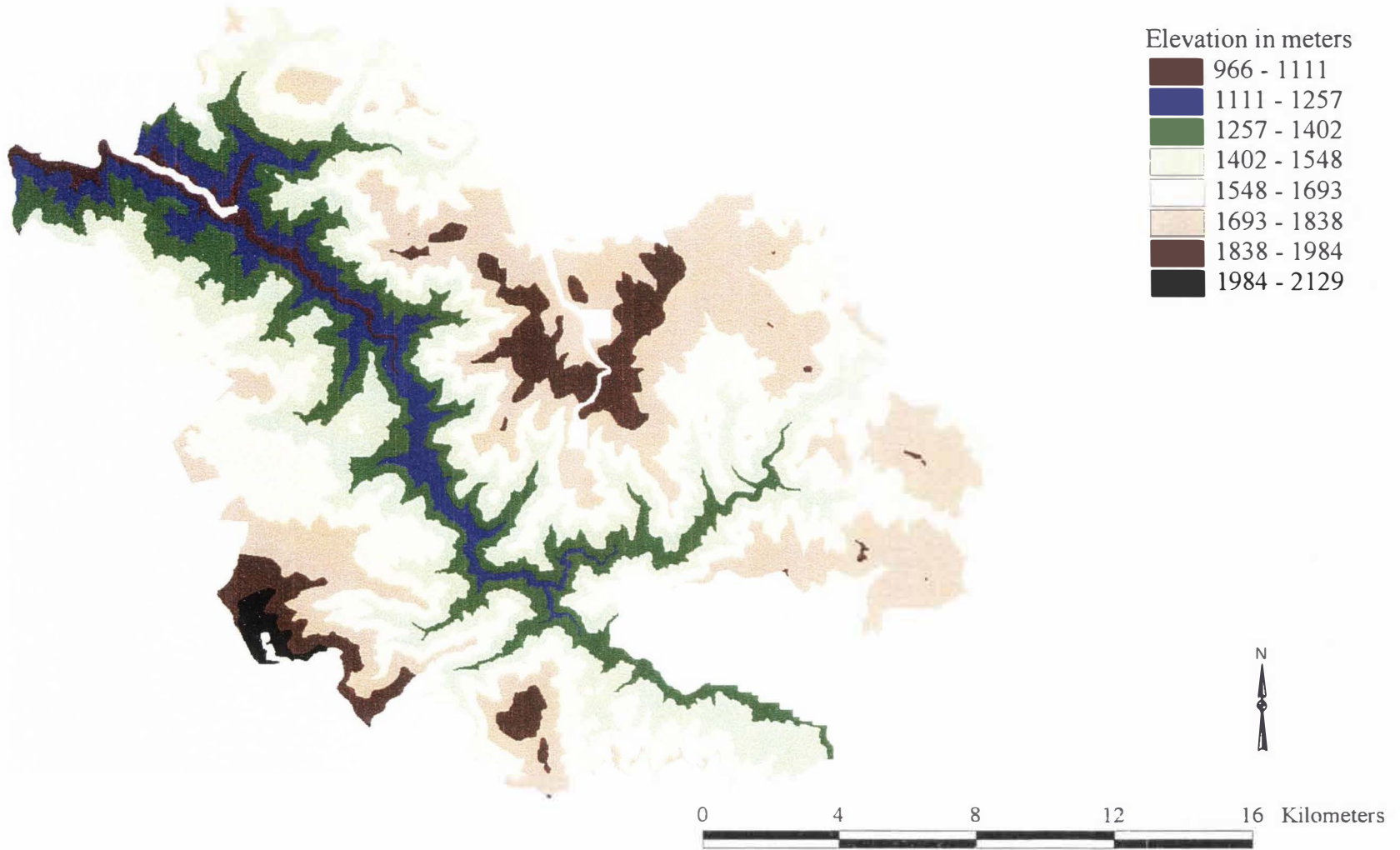
Fig. 3. Elevation in meters for the North Fork John Day Wilderness.

from Roberts and Cooper (1987).

Topographic position was also derived from the elevation layer and was obtained using the program Toposcale, written by Niklaus Zimmerman (unpubl.) to run in the GRID module of ARC/Info. This program calculates topographic position values by calculating the mean elevation value for square cell neighborhoods and subtracting the mean value for the center cell. The program operates on multiple scales by iterating with different numbers of cells per set. For example, values would first be calculated using a 3 x 3 cell neighborhood, then again using a 5 x 5 cell set. For the final value of each cell, the program compares the standardized values of each iteration from the largest cell set to smallest, and the smallest values are chosen. The final layer has a range of values from negative to positive, with highly negative values indicating lowest slope positions such as stream channels, zero indicating flat areas, and positive values representing higher positions such as ridgelines.

The geology layer contained categorical data relating to parent material. This layer was originally digitized by the Umatilla National Forest from a 1:500,000 map drawn by the U. S. Geological Survey. Also consisting of categorical data, the soil type layer (Figure 4) contained mapping units that are described in detail in Ehmer (1978) and summarized briefly in Appendix A. These data and also data calculated from this source by Busskohl (unpubl.) were used to create separate layers describing aspects of the mapping units thought to influence the distribution of plant associations. These continuous data layers were created in ARC/Info and included total soil depth, the estimated water holding capacity of the rooting zone, the percentage of coarse fragments in the rooting zone, the thickness of the ash layer in the soil profile, soil taxonomic
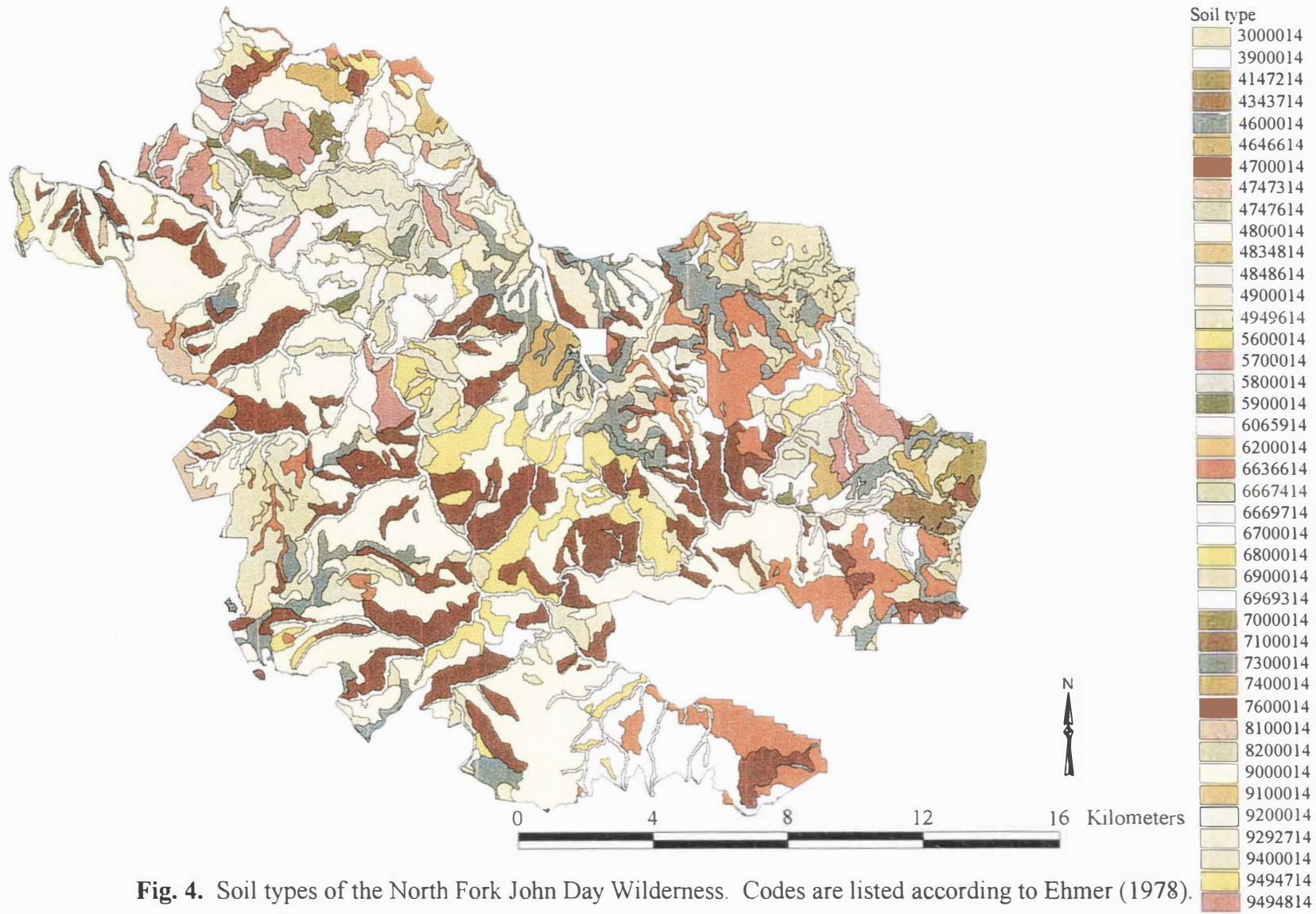
Fig. 4. Soil types of the North Fork John Day Wilderness. Codes are listed according to Ehmer (1978).

Soil type
3000014
3900014
4147214
4343714
4600014
4646614
4700014
4747314
4747614
4800014
4834814
4848614
4900014
4949614
5600014
5700014
5800014
5900014
6065914
6200014
6636614
6667414
6669714
6700014
6800014
6900014
6969314
7000014
7100014
7300014
7400014
7600014
8100014
8200014
9000014
9100014
9200014
9292714
9400014
9494714
9494814

moisture and temperature regimes, and mineralogy class. All of these data layers

represent average values for each soil attribute for numerous observations within each

soil mapping unit.

The precipitation layer was produced for the Umatilla Forest by the Oregon State

Climate Service Center and originally contained only contours representing annual

precipitation in 5-inch intervals. These data were manipulated in ARC/Info so that each

cell would contain a numerical value. Values for the entire map were generated

through the use of the ARC/Info Triangulated Irregular Network algorithm and these

values were then stored in a new precipitation layer.

A layer approximating average annual direct solar radiation was developed using

the Solarflux Version 2.1 program (Hetrick 1996). A full description of this program can

be found in Dubayah (1994). Briefly, this program uses as input a DEM, start and end

times, the Julian day, the latitude and longitude of the site, and the local time meridian. It

also includes a feature that calculates the shading of areas over the landscape produced by

local topography. Calculations for daily total solar radiation were made for every ten

days of a calendar year, and these values were then added and expressed in

kilojoules/m^2/year.

In order to estimate soil moisture due to runoff across the study area, a simple

topographic wetness index was calculated as:

$$W_i = \ln(A/\tan \beta)$$

where W_i is the wetness index, A is contributing area, and β is the slope angle in degrees

(Moore et al. 1993b). The contributing area was calculated using the Dinf method, which

first calculates flow direction and then upslope catchment area by proportioning flow between downslope pixels (Tarboton 1997). The data layer was created in ARC/Info Grid.

Other layers relating to the hydrology and therefore soil moisture of the study site were created using algorithms in ARC/Info Grid. A grid with values approximating surface curvature based on a 3 x 3 cell neighborhood was created using the CURVE function in Grid. Also, the distance downslope to a stream was calculated from the Dinf grid representing contributing area described above.

After all of the data layers were prepared, the field data point positions were used to sample each layer in order to extract the data corresponding to that location. The resulting text file was then compared with the field database to check for agreement between field-measured and ARC/Info-generated values for elevation, slope, aspect, and topographic position. With the exception of aspect, the data values for the two sources were similar. Brzeziecki et al. (1993) also noted this discrepancy. There was a mean difference of 40.3 degrees between the aspect values from the two sources. Because the accuracy of the GIS-derived aspect values depended upon the resolution of the DEM and was therefore known to be lower than the field values, the actual measured aspect values were used in the final database in their raw form and were also used as the basis for transformed aspect values for input into the classification tree model. A summary of the data layers and their sources is included in Table 1.

Predictive Modeling

The classification tree was developed using the "tree" function in S-Plus 4.0 for

Table 1. Summary of environmental predictor variables and their sources.

Variable	Source
elevation	DEM
slope	derived from DEM
slope aspect	measured in field
transformed aspect value	calculated as indicated from field values for aspect
precipitation	Oregon Climate Service
geology type	Umatilla National Forest
soil type	Umatilla National Forest, from Ehmer (1978)
-soil depth	-derived from soil type from above
-rooting zone water capacity	-derived from soil type from above
-percent coarse fragments	-derived from soil type from above
-ash thickness in soil profile	-derived from soil type from above
-soil moisture regime	-derived from soil type from above
-soil temperature regime	-derived from soil type from above
-mineralogy class	-derived from soil type from above
solar radiation	Solarflux program (Hetrick 1996)
topographic position	Toposcale program
wetness index	based on Dinf method and Moore et al. (1993b)
surface curvature	CURVE algorithm in ARC/Info
distance to stream	FLOWLENGTH algorithm in ARC/Info

Windows. The input is a text file with environmental variables as labeled columns and sample plots as rows. The classification tree model is then used to model plant association type as a function of these variables. In particular, the model selects in stepwise fashion those environmental variables that best distinguish successive subgroups of the specified association types from each other. In producing the tree, the classification model repeatedly selects the predictor variable that splits the data points into the two purest, or most homogeneous, classes. In essence, the model minimizes the number of misclassified data points at each split based on values for predictor variables. The splitting rules are made using a likelihood ratio statistic called deviance, and this determines the purest or best split at each node using the data. It is important to note that as the tree continues to grow, there are less data at any given node, so confidence in the splitting rules decreases (Aitken 1998). Any predictor variable may be used once, more than once, or not at all in the development of the tree. Also, predictor variables can be both categorical and continuous in nature. Partitioning continues until less than five class members remain (Clark & Pregibon 1992).

This resulting tree is always overfitted, so that some of the splits are not actually informative (Breiman et al. 1984). To correct for this problem, the tree was pruned to an optimal size, which was determined through 10-fold cross-validation. In cross-validation, the original data set is divided into 10 subsamples without replacement. Withholding one of the subsamples, a tree is developed using the remaining 90% of the original data set. The excluded data points are then classified using this tree and used to estimate the misclassification error rate. This process is repeated until all subsamples have been withheld sequentially, and the mean misclassification rate from the 10 trials is

taken as the cross-validated estimation of accuracy (Verbyla 1987). From a plot of the cross-validated deviance versus the number of terminal nodes, the tree of optimal size, with the lowest deviance value, can be determined. Because it has been noted that the optimal tree size varies slightly each time a cross-validation is performed (Roberts pers. comm.), 10 cross-validations were performed and the optimal tree size was inferred from an overlay of the results. The original tree was then reduced to this size using the cost-complexity pruning method (Clark & Pregibon 1992).

The preliminary trees developed using this method always used the categorical soil and geology types as predictors. Because not all of the soil and geology types (or combinations of the types) present in the study site were represented in the data set, it was necessary to develop a second tree using only continuous data as predictors. This second tree would be used only in those instances where the primary tree did not adequately describe the conditions of the site. Further information on this method can be found in Aitken (1998).

To further analyze the model developed for the Wilderness, a program based on fuzzy set theory (Roberts 1999) was used. This program, fuzzy confusion analysis, uses as input a similarity matrix and the tree model. The similarity matrix was constructed using data from the 26 plant association plots and additional plot data provided by C. Johnson (unpubl.). Similarity values were calculated using the natural log of the coverage values. Overall, the effect of using fuzzy class membership is that "near misses" in prediction are taken into account in analyzing the tree. If the tree predicts a plant association which is incorrect, but very similar to the actual plant association, then the error is regarded as less severe. The program calculates weighted errors, based on the

ratio of the mean similarity among different plots of the same plant association to the

similarity between the two plant associations. These weighted errors are used to calculate

significant fuzzy errors of commission, omission, bias, and the estimated fuzzy accuracy

of prediction for each type. The program also produces a confusion matrix, a table of the

actual type versus the tree-predicted type. The level of confusion in the error matrix is

given by K_{HAT} (Cohen 1960) and ranges from -4 to 1, with values closer to one indicating

higher levels of agreement. Lastly, the fuzzy analysis includes a fuzzy probability value

for each terminal node in the tree. This value gives an indication of the probability that

the prediction at that node is correct or similar to the correct value. This program was

used to analyze both the main model and the secondary tree constructed using only

continuous predictor variables.

The process of predictive modeling is an iterative one, so that a model is made,

examined for weaknesses, and the model database adjusted or expanded in the hope of

improving predictions in the next model generated. For this study, preliminary models

did not predict four of the five plant associations within the grand fir series. Because of

this, these four types were aggregated into two classes as indicated in Table 2. The

ABGR/VAME and ABGR/LIBO associations were combined to form an "wet ABGR"

type, and the ABGR/CARU and ABGR/VASC types formed the "dry ABGR" type.

Additionally, few stands within the subalpine fir series were encountered, and no single

association had more than five sample points. Because of this, all plant association

points for this series were aggregated into a single subalpine fir type.

The final map of the plant associations of the North Fork John Day Wilderness

was created in the GRID module of ARC/Info using the final, pruned tree. First, the tree

was rewritten as a series of conditional statements. Then these statements were used to classify each cell of a grid using the DOCELL command of GRID. Each model parameter existed as a grid, which GRID called up as needed in order to determine the value for that parameter. GRID then proceeded through the predictive key to arrive at a designation of a plant association for each cell. This process continued on a cell-by-cell basis until a complete map of the plant associations of the study area was produced. As noted above, a second model, created from only continuous data, was used to classify those pixels that could not be described by the original tree.

RESULTS

Plant Association Data

Two hundred known plant association points were recorded in the North Fork John Day Wilderness. These represented a total of 14 different plant association types, 12 within five forested series and two within non-forested types. Table 2 summarizes the plant association names, codes, the number of data points, and the codes used in the final model for each type.

The vascular plant species that were encountered in the 26 plant association plots are reported in Appendix B. These data were combined with data provided by C. Johnson (unpubl.) to create a matrix of similarity values between plant associations. This similarity matrix, used in the fuzzy uncertainty analysis, appears in Appendix C.

Predictive Modeling

The initial, unpruned model for the Wilderness had 30 terminal nodes and used nearly every environmental variable as a predictor of plant association types. This tree was pruned to the optimal size of eight terminal nodes, and the predictors used in this smaller tree were soil type, slope, precipitation, elevation, and aspect. The final classification tree developed for the Wilderness predicted eight different plant associations with an initial estimated accuracy of 40%. The tree is presented in Figure 5. The types predicted by this main tree were dry ABGR, ABGR/CLUN, ABLA, MEADOW, PICO/CARU, PIPO/CAGE, PIPO/CARU, and PSME/PHMA.

The model developed with only continuous environmental variables as predictors also initially had 30 terminal nodes, and used nearly every continuous predictor in the

Table 2. Summary of plant associations, codes, number of sample points, and codes in the final model.

| Plant Association	Code	Number of points	Code in model
Abies grandis/Calamagrostis rubescens Buckl.	ABGR/CARU	19	dry ABGR
A. grandis/Clintonia uniflora	ABGR/CLUN	11	ABGR/CLUN
A. grandis/Linnea borealis	ABGR/LIBO	11	wet ABGR
A. grandis/Vaccinium membranaceum Dougl.	ABGR/VAME	12	wet ABGR
A. grandis/Vaccinium scoparium Leiberg	ABGR/VASC	12	dry ABGR
Abies lasiocarpa type	ABLA	21	ABLA
MEADOW type	MEADOW	21	MEADOW
Pinus contorta/Calamagrostis rubescens	PICO/CARU	10	PICO/CARU
P. ponderosa/Carex geyeri Boott	PIPO/CAGE	22	PIPO/CAGE
P. ponderosa/Calamagrostis rubescens	PIPO/CARU	16	PIPO/CARU
Poa secunda J. Presl/*Danthonia unispicata* (Thurb.) Munro	POSE/DAUN	8	POSE/DAUN
Pseudotsuga menziesii/Carex geyeri	PSME/CAGE	7	PSME/CAGE
Pseudotsuga menziesii/ Calamagrostis rubescens	PSME/CARU	14	PSME/CARU
Pseudotsuga menziesii/Physocarpus malvaceous (Greene) Kuntze	PSME/PHMA	16	PSME/P
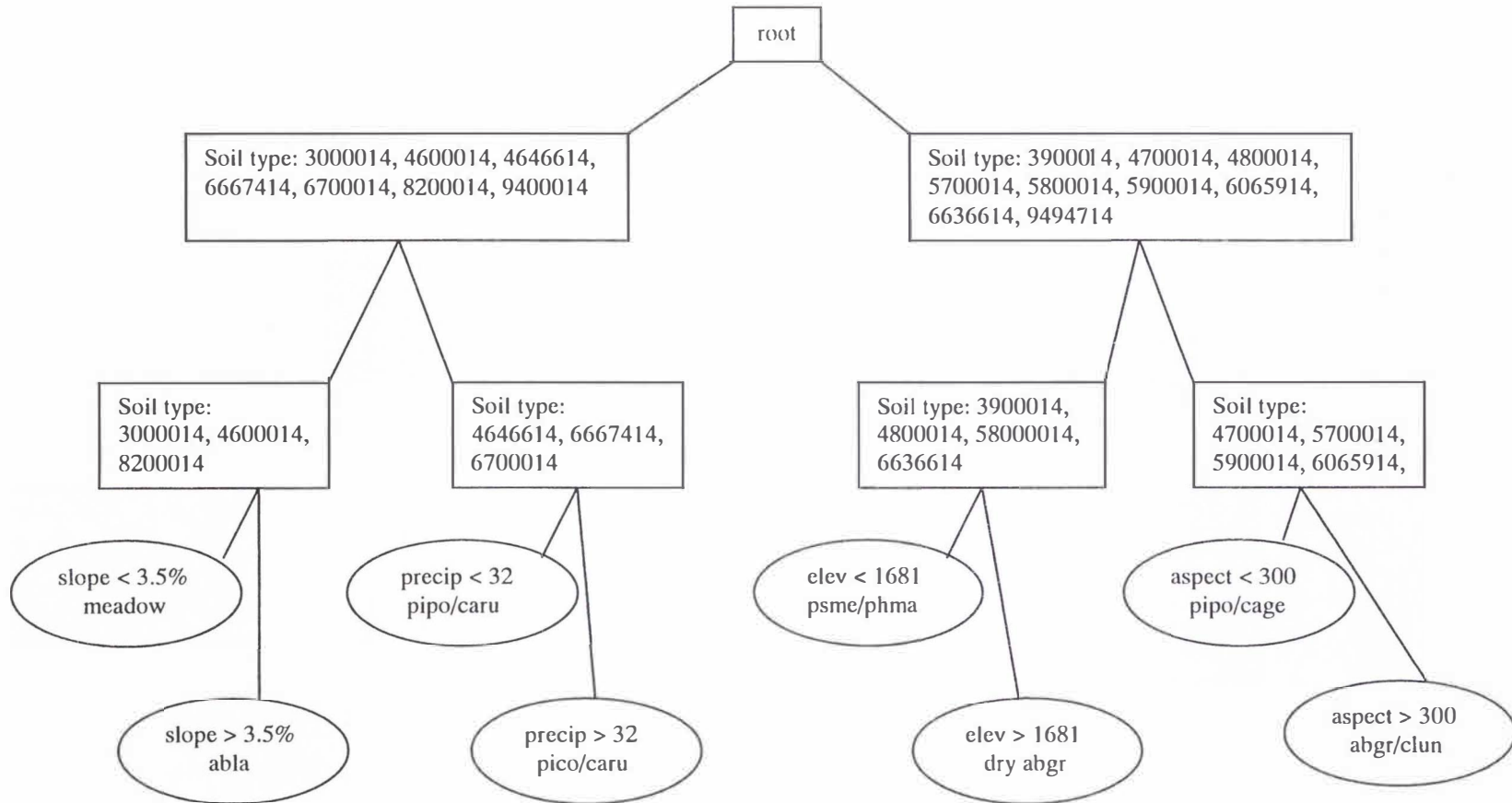
Fig. 5. Diagram of the tree model produced to predict the plant associations of the North Fork John Day Wilderness. Codes for soil types correspond to Ehmer (1978).

dataset. This tree was pruned to six terminal nodes, and used slope, precipitation, topographic position, elevation, and solar radiation as predictors. This secondary tree predicted six different plant associations. Only one of these, the wet ABGR type, was not also predicted by the main model. The initial estimated accuracy of the secondary tree was 33%, and it is presented in Figure 6.

The results of the fuzzy confusion analysis of the main tree, including the fuzzy confusion matrix, are shown in Appendix D. The overall fuzzy accuracy of the main tree, which as stated earlier takes into account near-misses in prediction, is 72%, with slightly under 145 correct predictions out of 200. There was significant bias (overestimation) in the prediction of the ABLA, MEADOW, PIPO/CARU, and PSME/PHMA types. For the main tree, the fuzzy K_{HAT} value was 0.6781. In examining the fuzzy confusion matrix for the main tree, several trends are apparent. With the exception of the ABGR/CLUN and MEADOW types with low (53%) fuzzy accuracy values, the plant associations which were predicted had fairly high accuracies. Two types, the PIPO/CAGE and PSME/PHMA associations were correctly predicted 94% and nearly 100% of the time, respectively. The other ponderosa pine plant association, PIPO/CARU, was predicted accurately 81% of the time. And finally, the dry ABGR, ABLA, and PICO/CARU types had fuzzy accuracy values of 78%, 75%, and 71%, respectively.

For the secondary tree, the overall fuzzy accuracy was somewhat lower at 67%. The results of the fuzzy confusion analysis for this model are summarized in Appendix E. This tree had significant fuzzy bias in all six types that were predicted. The fuzzy K_{HAT} value for this tree, 0.6151, indicated lower agreement between observations and predictions than in the main tree. The fuzzy confusion matrix for the secondary tree
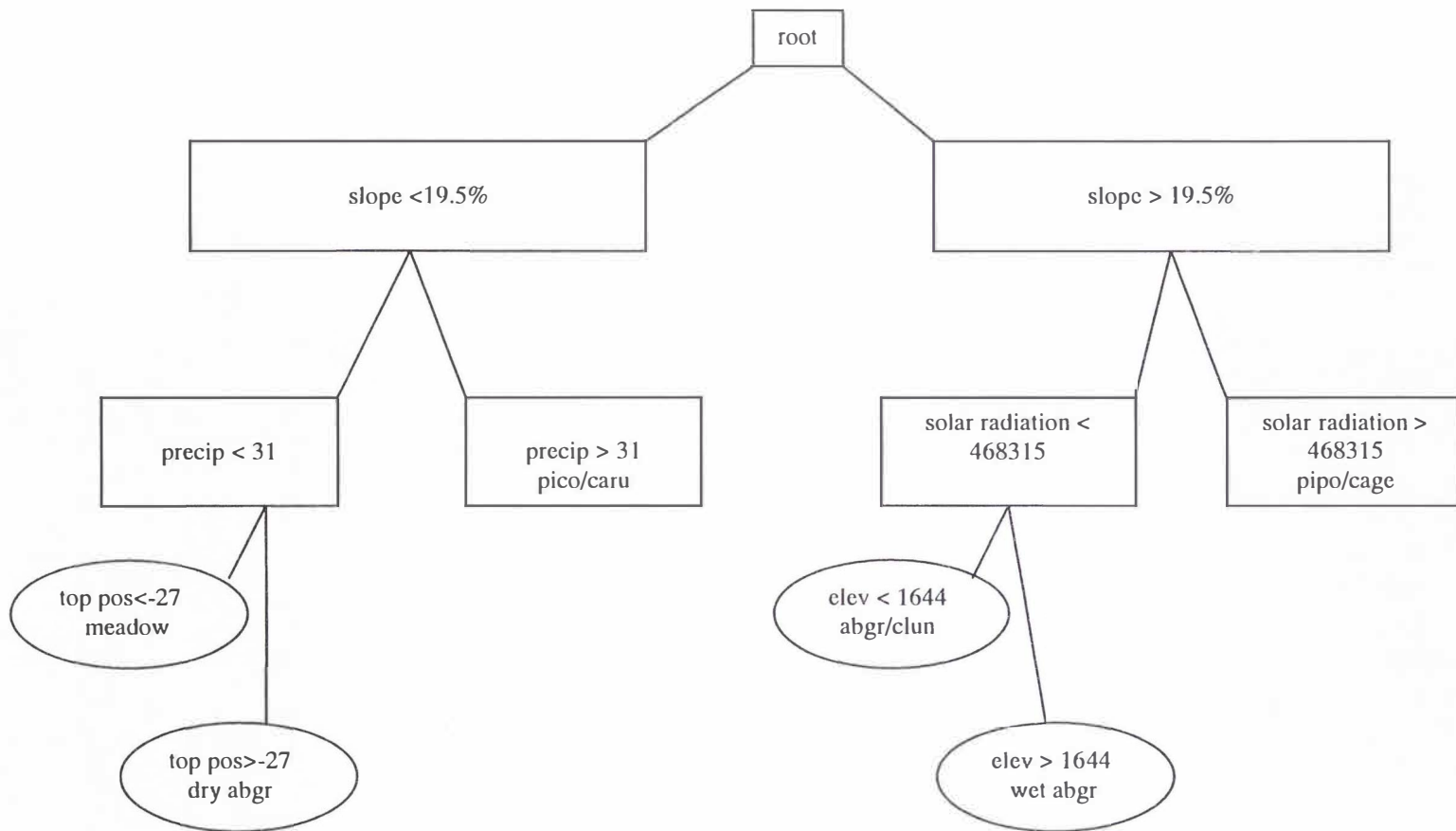
Fig. 6. Diagram of the secondary tree model produced to predict the plant associations of the North Fork John Day Wilderness.

indicated that the ABGR/CLUN type was actually more accurately predicted by this model, which lacked soil and geology type as possible predictor variables, than by the main tree. This type was predicted accurately over 96% of the time, as compared to the 53% presented above. A very high accuracy value of nearly 100% was found for the PICO/CARU type. This model predicted the dry ABGR, PIPO/CAGE, and MEADOW types with the same or slightly lower levels of accuracy than the main model. Finally, the dry ABGR type, which was not predicted by the main tree, was correctly predicted 84% of the time. In comparison with the main model, this tree did not predict the ABLA, PIPO/CARU, or PSME/PHMA types.

Using both of the trees, the final map of the plant associations predicted in the North Fork John Day Wilderness is presented in Figure 7. A map of the fuzzy probability of correct prediction for each terminal node of the model is shown in Figure 8. The main tree was used to classify 83% of the study area, and the second tree was used to predict plant associations over the remaining 17% of the Wilderness. Plant associations observed but not predicted in the study area were POSE/DAUN, PSME/CAGE, and PSME/CARU.

Predicted plant associations
- dry abgr
- abgr/clun
- wet abgr
- abla
- meadow
- pico/caru
- pipo/cage
- pipo/caru
- psme/phma

N

0 4 8 12 16 Kilometers
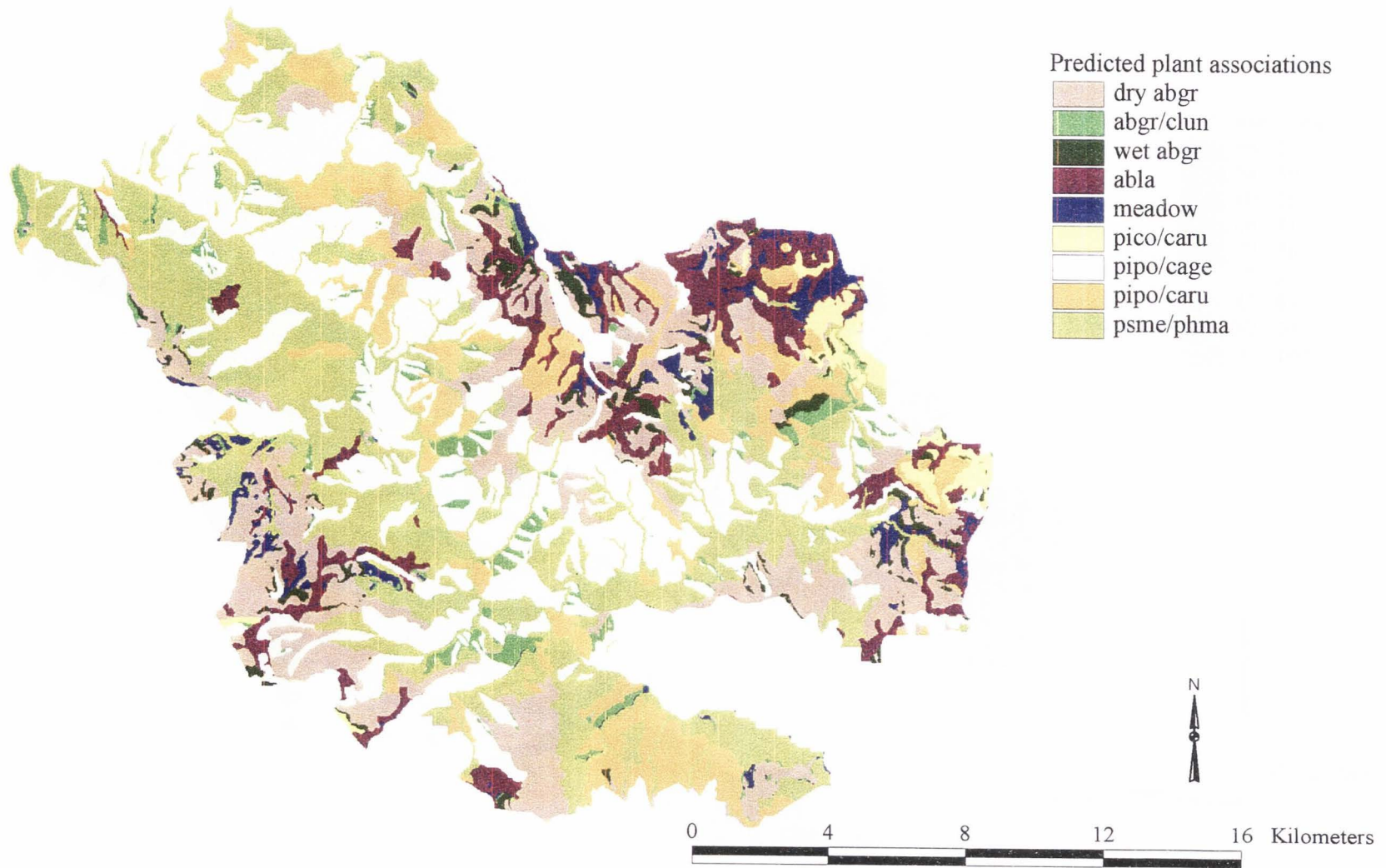
Fig. 7. Predicted plant associations of the North Fork John Day Wilderness.
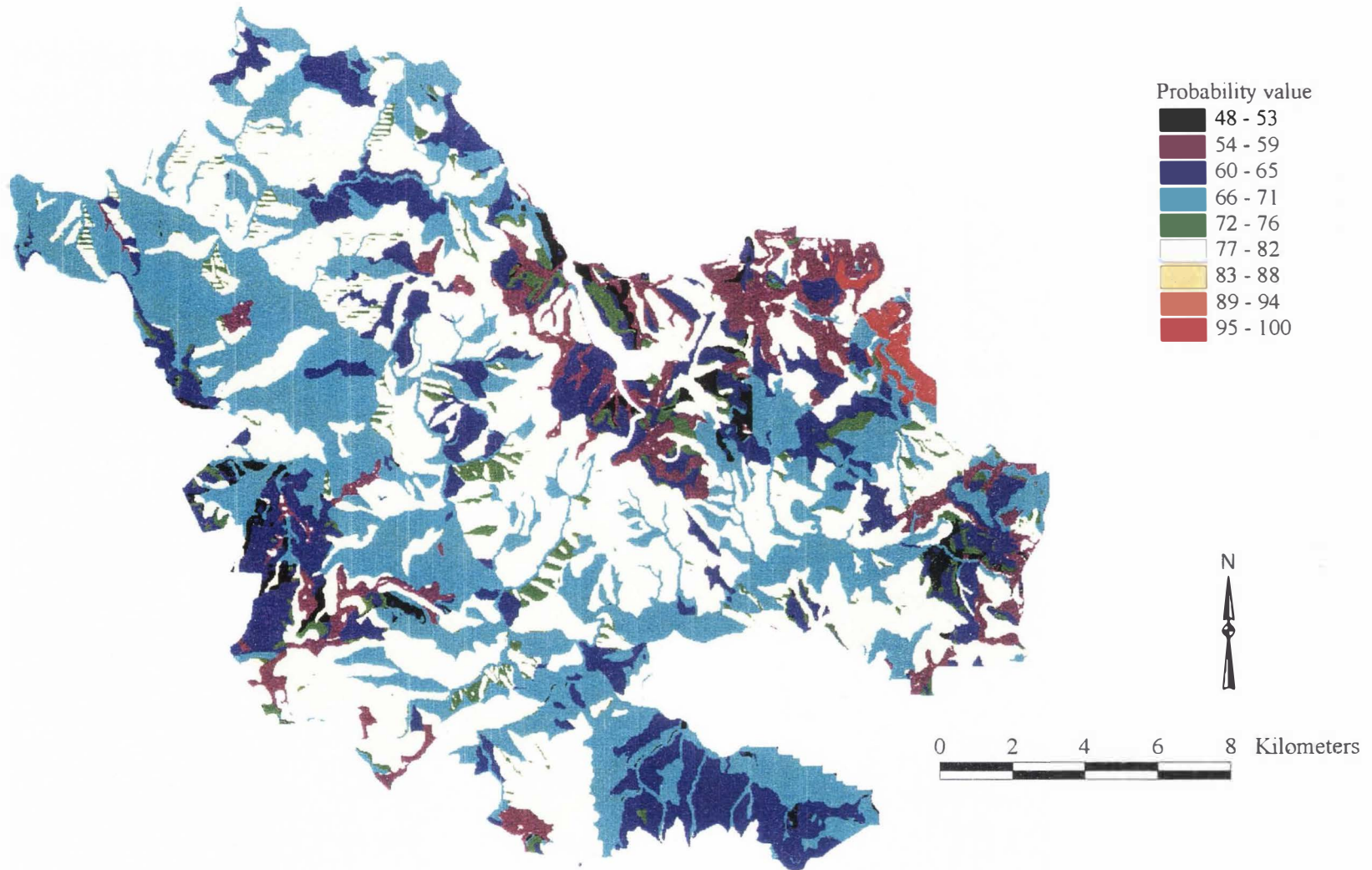
Fig. 8. Map of the fuzzy probability for each terminal node.

Probability value
- 48 - 53
- 54 - 59
- 60 - 65
- 66 - 71
- 72 - 76
- 77 - 82
- 83 - 88
- 89 - 94
- 95 - 100

0   2   4   6   8   Kilometers

DISCUSSION

The study met its objective of predictively modeling and mapping the potential natural vegetation of the North Fork John Day Wilderness. The main tree model predicted a total of eight different plant associations, and, in combination with the secondary tree, allowed for the spatial representation of nine different plant associations across the study site. The fact that the tree model uses soil type, slope, precipitation, elevation, and aspect as predictors is not surprising. These are factors which are very commonly used by predictive vegetation models, and together they represent major influences on gradients of nutrients, light, and moisture.

Examining the Tree Model

In examining the main tree itself, the first splits are made using soil types, which was somewhat unexpected. Other studies (Moore et al. 1991; McCullough 1995) have noted that variables which act broadly over the landscape, such as elevation and geology type, tend to be selected as predictors high up in the tree. These same authors also comment that the reverse is true, that predictors which tend to affect the environment more locally are selected lower in the tree models. At first, the results from this study seem to contradict these findings. Soil type, which is a more locally acting predictor than elevation, slope, aspect, or precipitation is selected twice as a predictor before these more broadly acting variables. However, to examine the soil splits further, I created maps that showed the areas classified by each leaf in the tree model. From the maps, the soil type split generally divided the Wilderness into steeply sloping areas along the main river canyon and areas of more rolling relief in the uplands surrounding the canyon. From this

and subsequent examination of the soil type descriptions, I believe that for the most part, the model is initially dividing the dataset into wetter sites characterized by deeper soil mapping units found on rolling topography and drier sites characterized by more shallow soil units on steeper slopes.

Classifying the study area in this manner makes sense ecologically, and a close relationship between vegetation and soil types in the region has been shown. Geist and Strickler (1978) noted strong agreement between parent material (ash or basalt) and overstory dominants in the Blue Mountains. Lentz and Simonson (1987) found that soil types closely corresponded with plant community types in southeastern Oregon. However, it is not clear why the topographic wetness index or any of the specific soil properties, especially water-holding capacity in the rooting zone and soil depth, were not selected as predictors along with or instead of soil type categories. Even the secondary model, developed without soil type and geology as predictors, failed to use any of these variables and instead used slope at the first split. While the soil types were mapped with considerable detail, the variation in soil properties within mapping units must still have been considerable. This variation may not have been adequately addressed by the data derived from the soil type coverage, and may partially explain why the indirect and complex gradient of soil type was preferred as a predictor. Alternatively, soil type may represent multiple correlated factors affecting plants in a single value and may therefore be the single best predictor of plant association type in the dataset.

Within the wetter soil types, slope and precipitation are the next predictors that are used. Slope is used to separate meadow types from subalpine fir, and this may reflect different drainage patterns. Slopes less than 3.5%, essentially flat, are used to predict the

meadow type, which are certainly flatter areas with poor drainage of water. It is interesting to note here that one plant association generally found on dry sites, the PIPO/CARU type, is predicted here by relatively low precipitation instead of being included with the dry soil type group. The precipitation variable is used to separate this type from the PICO/CARU type, which is a topographic and edaphic climax in frost pockets. A layer representing the flow and pooling of cold air over the landscape would probably serve as a better predictor for this type. The extremely limited distribution of this association in the Wilderness has led to a somewhat artificial prediction of PICO/CARU. This plant association occurs only in one small cold-air depression area, which also happens to have a relatively high amount of precipitation as compared to the rest of the study site. Most likely, this explains why the precipitation variable was used as a predictor, not because the lodgepole association is actually controlled by precipitation. This observation is an important one. The predictions of the model are based on the dataset and can reveal its limitations in splits that are ecologically unsound.

Within the drier soils group, elevation and untransformed aspect are selected as the next best predictors of plant association type. Higher elevations are used to differentiate the dry grand fir types from the PSME/PHMA plant association. Based on field observations, this splitting rule is fairly good for the study site but would not apply outside the Wilderness, since both plant associations can be found outside of these elevational limits. The untransformed aspect variable separates the PIPO/CAGE plant association from the more mesic ABGR/CLUN type, with northwest- to north-facing aspects predicting the grand fir association. This is a rational split, since northerly aspects tend to be wetter sites and thus more suited to the grand fir.

In comparing the main and secondary trees, similar trends are seen in both models. Instead of soil type, the secondary tree selected slope as the best initial split. This seems to confirm the observed differences in the two main soil groups described above. Within the less steeply sloped group, precipitation is the next best predictor. Next, low topographic position values are used to separate meadows from the dry ABGR types. This split is consistent with ecological understanding, since meadows typically occupy lower slope positions which accumulate water. Again, high precipitation values are used to predict the PICO/CARU type. As discussed earlier, this is an artifact of the dataset and does not represent an ecologically based splitting rule.

Within the group on steep slopes, solar radiation and elevation are the next predictors selected. Lower values for solar radiation separate two grand fir types, the ABGR/CLUN and the wet ABGR, from the PIPO/CAGE type. Recall that in the main tree, a similar split was made using aspect as a predictor. Finally, high elevations distinguish the wet ABGR type from the ABGR/CLUN plant association. This is consistent with field observations but is again not applicable to areas outside the Wilderness. The ABGR/CLUN type is usually found at low slope positions along drainages, which corresponds to low elevations within the study site. Overall, the two models are fairly similar in the manner that they predict plant associations for the study area.

Fuzzy Confusion Analysis

In examining the fuzzy confusion matrix for the main tree, most of the predicted plant associations were predicted with a high accuracy, with six of the eight associations

having accuracy values of 72% or higher. Interestingly, the most distinct type of all the plant associations, the meadow, was predicted with only 53% accuracy. While this type is correlated with only a few soil types, the soil type alone does not produce accurate predictions. It is surprising that the topographic moisture index was not used as a predictor for meadows, since this index gives high values for areas in which water tends to accumulate.

The inability of the model to predict five of the plant association types was a major limitation. This is a result of several factors. Plant associations occurring infrequently in the Wilderness and having few observations, such as the POSE/DAUN and PSME/CAGE types, were probably not predicted because of a lack of data. Were this study to continue, the addition of more known points for these associations would increase the likelihood of prediction. However, in initial models, four of the five grand fir plant associations, each with greater than 10 observations, were not predicted. This included the ABGR/CARU type, which had a relatively high number of observations (19). The combining of the grand fir plant associations as described in Table 2 resulted in the final model, which predicts all of the grand fir types. Whereas this combination was helpful in predicting the very common grand fir plant associations across the Wilderness, it is not entirely clear why regrouping the plant associations was necessary.

Several comments have been made about the pruning process in the refining of classification trees. Austin et al. (1994a) report that accuracy is lost by pruning the tree models and did not prune them at all. Franklin (1998) uses the same method of pruning as that presented here, but chooses what appears to be a large tree size (22 nodes) based on the range of optimal tree sizes reported (10 to 24 nodes) for pruning of all 20 models

developed. In this study, pruning resulted in the inability of the model to predict five of the 12, or nearly half of the plant association types observed in the Wilderness. This is an important consideration, because the utility of mapped predictions in certain land management applications is related to the number of associations predicted by the model.

In comparison with the initial accuracy estimates provided by the S-Plus software, the overall fuzzy accuracy estimates were much higher. The increase in accuracy in the main tree from 40% to 72% indicates that a large proportion of the misclassification errors were not severe, that is, the plant association that was predicted was similar to the actual plant association. This is consistent with other findings. Lynn et al. (1995) comment that although they found a high rate of misclassification using classification tree modeling to predict PNV in central New York, in many cases the actual type was similar to the predicted type. Roberts (1999) reports that accuracy values for forested types predicted in western Wyoming increased from 49% to 83% when analyzed with the fuzzy confusion method. The results of the fuzzy confusion analysis are especially useful in light of the practical applications of the predictive model developed for the Wilderness. Since the model may be used in restoration and other management practices in which relatively general plant community information is needed, the knowledge that most plant associations are not seriously misclassified is important.

Considerations and Conclusions

Based on field observations and the results discussed above, most of the plant associations in the study area are predicted satisfactorily using this method. The problem of misclassification and lack of prediction within the grand fir series, especially before

the combining of plant associations into two major types, is difficult to explain. The series as a whole has wide ecological amplitude within the study site, and occupies a wide variety of sites. Based on the results of the secondary tree, it seems that soil type is not the best predictor variable within this series.

Michaelson et al. (1994) suggest that the classification tree method works best with at least 300 to 400 observations, but is still possible with datasets as small as 100. This study employed 200 observations in training the model. To improve the accuracy of the model and the number of plant association types predicted, further data collection would be necessary and should focus on plant associations within the grand fir series.

Along with expanding the number of observations, future work in this area should focus on expanding the data layers used as environmental predictors. Most importantly, both the spatial resolution and the detail of the soil database should be expanded. At present, no soil nutrient characteristic data are available for the study area. Ash-influenced soils are known to have a different nutrient status than soils derived from other materials, and this could be affecting plant distribution. Also, very recently, GIS layers representing precipitation by month have become available, and these could be used to model differences in growing season precipitation in the study area. Lastly, a spatially explicit model for cold air drainage across a landscape would aid in an ecologically based prediction of the lodgepole pine plant association.

LITERATURE CITED

Agee, J.K. 1994. *Fire and weather disturbances in terrestrial ecosystems of the eastern Cascades.* Gen. Tech. Rep. PNW-GTR-320. USDA Forest Service, Pacific Northwest Research Station, Portland, OR.

Aitken, M.J. 1998. *Predictive modeling of rare plant habitat in the eastern Great Basin.* M.S. Thesis, Utah State University, Logan, UT.

Alt, D.D & Hyndman, D.W. 1978. *Roadside geology of Oregon.* Montana Press, Missoula, MT.

Anonymous, 1999. Climate data for Ukiah, OR. [On-line]. Oregon Climate Service, http://www.ocs.orst.edu.

Austin, M.P., Cunningham, R.B. & Fleming, P.M. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* 55: 11-27.

Austin, M.P., Meyers, J.A., Belbin, L. & Doherty, M.D. 1994a. *Modelling of landscape patterns and processes using biological data, Sub-project 5: Simulated data case study.* Division of Wildlife and Ecology, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra.

Austin, M.P., Meyers, J.A. & Doherty, M.D. 1994b. *Modelling of landscape patterns and processes using biological data, Sub-project 2: Predictive models for landscape patterns and processes.* Division of Wildlife and Ecology, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra.

Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. 1994c. Determining species response functions to an environmental gradient by means of a β-function. *J. Veg. Sci* 5: 215-228.

Austin, M.P., Nicholls, A.O. & Margules C.R. 1990. Measurement of the realized qualitative niche: environmental niches of five Eucalyptus species. *Ecol. Monogr.* 60(2): 161-177.

Bio, A.M.F., Alkemade, R. & Barendregt, A. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. *J. Veg. Sci.* 9: 5-16.

Bradfield, G.E. & Scagel, A. 1984. Correlations among vegetation strata and environmental variables in subalpine spruce-fir forests, southeast British Columbia. *Vegetatio* 55: 105-114.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. 1984. *Classification and regression trees*. Wadsworth, Belmont, CA.

Brown, D.G. 1994. Predicting vegetation types at treeline using topography and biophysical disturbance variables. *J. Veg. Sci* 5: 642-656.

Brzeziecki, B., Kienast, F. & Wildi, O. 1993. A simulated map of the potential natural forest vegetation of Switzerland. *J. Veg. Sci.* 4: 499-508.

Clark, L.A. & Pregibon, D. 1992. Tree-based models. In: Chambers, J. & Hastie, T.J. (eds.) *Statistical models in S*, pp. 377-419. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, CA.

Cohen, J. 1960. A coefficient of agreement of nominal scales. *Educ. Psych. Meas.* 20: 37-46.

Collins, S.L. & Glenn, S.M. 1990. Importance of spatial and temporal dynamics in species regional abundance and distribution. *Ecology* 72: 654-664.

Covington, W.W., Everett, R.L., Steele, R., Irwin, L.L., Daer, T.A. & Auclair, A.N.D. 1994. Historical and anticipated changes in forest ecosystems of the Inland West of the United States. *J. Sust. For.* 2: 13-63.

Daubenmire, R. 1952. Forest vegetation of northern Idaho and adjacent Washington, and its bearing on concepts of vegetation classification. *Ecol. Monogr.* 22: 301-330.

Deitschman, G.H. 1973. *Mapping of habitat types throughout a national forest*. Gen. Tech. Report INT-11. USDA Forest Service Intermountain Forest and Range Experiment Station, Ogden, UT.

Dubayah, R.C. 1994. Modeling a solar radiation topoclimatology for the Rio Grande River Basin. *J. Veg. Sci.* 5: 627-640.

Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19: 474-499.

Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *J. Veg. Sci.* 9: 733-748.

Franklin, J.G. & Dyrness, C.T. 1988. *Natural vegetation of Oregon and Washington*. Oregon State University Press, Corvallis, OR.

Gauch, H.G. Jr. & Whittaker, R.H. 1972. Coenocline simulation. *Ecology* 53: 446-451.

Geist, J.M. & Strickler, G.S. 1978. *Physical and chemical properties of some Blue Mountain soils in northeast Oregon.* USDA Forest Service Res. Pap. PNW-236. Pacific Northwest Forest and Range Experiment Station, Portland, OR.

Gleason, H.A. 1926. The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club* 53: 7-26.

Hetrick, W. 1996. *Solarflux Version 2.1.* University of Kansas, Lawrence, KA.

Hill, M.O. & Gauch, H.G. 1980. Detrended correspondence analysis, an improved ordination technique. *Vegetatio* 42: 47-58.

Hitchcock , C.L. & Cronquist, A. 1973. *Flora of the Pacific Northwest.* University of Washington Press, Seattle, WA.

Hutchinson, G.E. 1957. Concluding remarks. *Cold Springs Harbour Symposia on Quantitative Biology* 22: 415-27.

Johnson, C.G. Jr. & Clausnitzer, R.R. 1992. *Plant associations of the Blue and Ochoco Mountains.* USDA Forest Service, Pacific Northwest Region, Portland, OR.

Johnson C.G. Jr., Clausnitzer, R.R., Mehringer, P.J. & Oliver, C.D. 1994. *Biotic and abiotic processes of eastside ecosystems: the effects of management on plant and community ecology, and on stand and landscape vegetation dynamics.* Gen. Tech. Rep. PNW-GTR-322. USDA Forest Service, Pacific Northwest Research Station, Portland, OR.

Kalkhoven, J.T.R. & van der Werf, S. 1988. Mapping the potential natural vegetation. In: Küchler, A. W. and I. S. Zonneveld (eds.). *Vegetation mapping*, pp. 375-385. Kluwer Academic, Dordrecht.

Layser, E.F. 1974. Vegetative classification: its application to forestry in the northern Rocky Mountains. *J. For.* 72: 354-357.

Lees, B.G. & Ritman, K. 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Env. Management* 15: 823-831.

Lentz, R.D. & Simonson, G.H. 1987. Correspondence of soil properties and classification units with sagebrush communities of southeastern Oregon: I. Comparisons between mono-taxa soil-vegetation units. *Soil Sci. Soc. Am. J.* 51: 1263-1271.

Lynn, H., Mohler, C.L., DeGloria, S.D. & McCulloch, C.E. 1995. Error assessment in decision-tree models applied to vegetation analysis. *Landscape Ecol.* 10: 323-335.

McCullough, P.D. 1995. *Predictive vegetation mapping using classification tree analysis*. M.A. Thesis, San Diego State University, San Diego, CA.

Michaelson, J., Schimel, D.S., Friedl, M.A., Davis, F.W. & Dubayah, R.C. 1994. Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys. *J. Veg. Sci.* 5: 673-686.

Moore, D.M., Lees, B.G. & Davey, S.M. 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. *Env. Management* 15: 59-71.

Moore, I.D., Gessler, P.E., Nielsen, G.A. & Peterson, G.A. 1993a. Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Am. J.* 57: 443-452.

Moore, I.D., Norton, T.W. & Williams, J.E. 1993b. Modelling environmental heterogeneity in forested landscapes. *J. Hydrology* 150: 717:747.

Palmer, A.R. 1991. The potential vegetation of the upper Orange River, South Africa: concentration analysis and its application to rangeland assessment. *Ceonoses* 6: 131-138.

Pfister, R.D. & Arno, S.F. 1980. Classifying forest habitat types based on potential climax vegetation. *For. Sci.* 26: 52-70.

Roberts, D.W. 1999. Fuzzy confusion analysis of statistical predictive models of vegetation distribution. *Ecol. Soc. Amer. Bull.* (in press).

Roberts, D.W. & Cooper, S.V. 1987. Concepts and techniques of vegetation mapping. In: Ferguson, D., Morgan, P., & Johnson, F. D. (eds.). *Proceedings–Land classifications based on vegetation: applications for resource management, pp. 1-12.* GTR-257. USDA Forest Service Intermountain Research Station. Ogden, UT.

Skidmore, A.K., Gauld, A. & Walker, P. 1996. Classification of kangaroo habitat distribution using three GIS models. *Int. J. Geographical Information Systems* 10: 441-454.

Tansley, A.G. 1935. The use and abuse of vegetation concepts and terms. *Ecology* 16: 284-307.

Tarboton, D.G. 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Res. Research* 33: 309-319.

ter Braak, C.J.F. 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio* 69: 69-77.

Verbyla, D.L. 1987. Classification trees: a new discrimination tool. *Can. J. For. Res.* 17: 1150-1152.

Whittaker, R.M. 1956. Vegetation of the Great Smoky Mountains. *Ecol. Monographs* 26: 1-80.

Whittaker, R.M. 1987. An application of detrended correspondence analysis and nonmetric multidimensional scaling to the identification and analysis of environmental factor complexes and vegetation structures. *J. Ecol.* 75: 363-376.

APPENDICES

Appendix A. Soil Data Tables

Table A1. Summary of specific soil properties for each soil mapping unit defined in Ehmer (1978).

Soil Mapping Unit	Total Soil Depth (cm)	Volcanic Ash Thickness (cm)	Root Zone Coarse Fragment %	Root Zone Water Holding Capacity (cm)	Bedrock Type
3000014	152.4	50.8	0	6.90	Tuff
3900014	60.9	0	190.5	1.06	Andesite
4600014	50.8	0	0	15.74	Metavolcanic
4700014	48.2	0	88.9	7.11	Metavolcanic
4800014	76.2	45.7	0	14.17	Metavolcanic
4900014	106.6	45.7	0	14.17	Metavolcanic
5600014	101.6	0	12.7	11.73	Granite
5700014	111.7	0	12.7	5.58	Granite
5800014	203.2	0	12.7	19.73	Granite
5900014	30.4	0	190.5	0.91	Granite
6200014	187.9	55.8	0	17.32	Granite
6700014	127.0	66.0	12.7	19.15	Tuff
6800014	66.0	25.4	0	12.59	Tuff
6900014	91.4	27.9	0	14.96	Tuff
7000014	111.7	22.8	0	8.66	Tuff
7100014	152.4	45.7	0	23.62	Tuff
7300014	40.6	0	165.1	2.03	Tuff
7400014	101.6	63.5	0	19.68	Tuff
7600014	127.0	68.5	0	21.25	Tuff
8100014	66.0	55.8	0	17.32	Sedimentary
8200014	124.4	55.8	0	17.32	Sedimentary
9000014	304.8	0	203.2	0.50	Andesite
9100014	10.1	0	203.2	0.40	Basalt
9200014	7.6	0	203.2	0.22	Tuff
9400014	10.1	0	165.1	0.81	Metavolcanic

Table A2. Complexes which occur in the study site.

Complex	Constituent map units
4147214	4100014, 7200014
4343714	4300014, 4700014, 9400014
4646614	4600014, 4900014
4747314	3000014, 4700014
4747614	4700014, 6800014, 7600014
4834814	3000014, 4700014, 4800014
4848614	4800014, 6700014
4949614	4900014, 6000014
6065914	5900014, 6000014, 9400014
6636614	3000014, 6600014, 6900014
6667414	6600014, 7200014
6669714	6600014, 6900014
6969314	6900014, 7300014
9292714	7000014, 9200014
9494714	4700014, 9400014
9494814	4700014, 4800014, 9400014

Appendix B. Species List

List of the vascular plant species encountered within the 26 inventory plots in the North Fork John Day Wilderness.

Abies grandis (D. Don) Lindl.
Abies lasiocarpa (Hook.) Nutt.
Acer glabrum (Hook.) Dippel var. douglasii
Achillea millefolium L.
Agoseris grandiflora (Nutt.) Greene
Agrostis spp.
Allium tolmiei S. Wats. *var. tolmiei*
Antennaria dimorpha (Nutt.) Torr. & Gray
Antennaria stenophylla (Gray) Gray
Apocynum androsaemifolium L.
Arctostaphylos uva-ursi (L.) Spreng.
Arenaria congesta Nutt.
Arnica cordifolia Hook.
Artemisia rigida (Nutt.) Gray
Aster foliaceus DC.
Astragalus whitneyi Gray
Bromus carinatus Hook. & Arn.
Bromus tectorum L.
Bromus vulgaris (Hook.) Shear
Calamagrostis koelerioides Vasey
Calamagrostis rubescens Buckl.
Calochortus eurycarpus S. Wats.
Carex geyeri Boott
Castilleja hispida Benth.
Castilleja tenuis (Heller) Chuang & Heckard
Ceanothus velutinu Hook.
Cerastium arvense L.
Cercocarpus ledifolius Nutt.
Chimaphila menziesii (D. Don) Spreng.
Chimaphila umbellata (L.) W. Bart.
Cirsium arvense (L.) Scop.
Cirsium neomexicanum (Petrak) Welsh *var. utahense*
Cirsium scariosum Nutt.
Cirsium vulgare (Savi) Ten.
Collinsia parviflora Lindl.
Collomia grandiflora Lindl.
Dactylis glomerata L.
Danthonia unispicata (Thurb.) Macoun
Deschampsia cespitosa (L.) Beauv.
Deschampsia elongata (Hook.) Munro
Elymus elymoides (Raf.) Swezey *ssp. elymoides*
Elymus glaucus Buckl. *ssp. glaucus*

Epilobium angustifolium L.
Epilobium brachycarpum K. Presl
Erigeron chrysopsidis Gray
Eriogonum heracleoides Nutt.
Eriogonum umbellatum Torr.
Erythronium grandiflorum Pursh
Festuca occidentalis Hook.
Festuca rubra L.
Fragaria vesca L.
Fragaria virginiana Duchesne
Galium aparine L.
Geum triflorum (Pursh) Fassett *var. ciliatum*
Goodyera oblongifolia Raf.
Helianthella uniflora (Nutt.) Torr. & Gray
Hieracium albiflorum Hook.
Hieracium cynoglossoides Arv.-Touv.
Hieracium scouleri Hook.
Hypericum anagalloides Cham. & Schlecht.
Juncus ensifolius Wikstr.
Juniperus occidentalis Hook.
Larix occidentalis Nutt.
Ligusticum canbyi Coult. & Rose
Linanthus harknessii (Curran) Greene
Linnaea borealis L.
Listera caurina Piper
Lomatium triternatum (Pursh) Coult. & Rose
Lonicera spp.
Lupinus caudatus Kellogg
Lupinus sericeus Pursh
Luzula campestris (L.) DC.
Mahonia repens (Lindl.) G. Don
Melica bulbosa Porter & Coult.
Moehringia macrophylla (Hook.) Fenzl
Orthilia secunda (L.) House
Osmorhiza occidentalis (Torr. & Gray) Torr.
Paxistima myrsinites (Pursh) Raf.
Pedicularis groenlandica Retz.
Pedicularis racemosa Benth.
Phacelia hastata Lehm.
Phleum pratense L.
Physocarpus malvaceus (Greene) Kuntze
Picea engelmannii Engelm.
Pinus contorta S. Wats. *var. latifolia*
Pinus ponderosa P. & C. Lawson
Plagiobothrys tenellus (Hook.) Gray

Poa nervosa (Hook.) Vasey
Poa pratensis L.
Poa spp.
Polemonium pulcherrimum Hook.
Polygonum douglasii (Meisn.) Hickman *ssp. majus*
Potentilla glandulosa Lindl.*ssp. glandulosa*
Potentilla gracilis Hook.
Pseudoroegneria spicata (Pursh) A. Love *ssp. spicata*
Pseudotsuga menziesii (Beissn.) Franco *var. glauca*
Pyrrocoma carthamoides Hook. *var. carthamoides*
Ranunculus uncinatu G. Don
Rosa nutkana K. Presl
Rosa woodsii Lindl.
Salix spp.
Scutellaria angustifolia Pursh
Sedum stenopetalum Pursh
Sedum stenopetalum Pursh
Spiraea betulifolia Pallas
Symphoricarpos albus (L.) Blake
Taraxacum officinale Wiggers
Tragopogon dubius Scop.
Trifolium longipes Nutt.
Trifolium pratense L.
Trifolium repens L.
Trisetum spp.
Trisetum wolfii Vasey
Vaccinium membranaceum Torr.
Vaccinium scoparium Coville
Vahlodea atropurpurea Hartman
Valeriana sitchensis Bong.
Viola adunca Sm.
Viola orbiculata Holz

Appendix C. Table of Similarity Values

Table C1. Similarity matrix for plant association types based on inventory plot data and data given by Johnson (unpubl.).

Plant association type	Number of plots	Similarity values											
dry abgr	8	0.42	0.28	0.35	0.25	0.00	0.28	0.21	0.33	0.12	0.35	0.36	0.33
abgr/clun	3	0.28	0.56	0.39	0.16	0.00	0.06	0.02	0.11	0.01	0.06	0.16	0.19
wet abgr	8	0.35	0.39	0.39	0.27	0.01	0.20	0.14	0.17	0.08	0.18	0.25	0.24
abla	7	0.25	0.16	0.27	0.47	0.02	0.37	0.11	0.16	0.15	0.22	0.18	0.18
meadow	1	0.00	0.00	0.01	0.02	1.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00
pico/caru	3	0.28	0.06	0.20	0.37	0.04	0.52	0.19	0.25	0.27	0.28	0.22	0.21
pipo/cage	3	0.21	0.02	0.14	0.11	0.00	0.19	0.52	0.35	0.16	0.32	0.37	0.35
pipo/caru	4	0.33	0.11	0.17	0.16	0.00	0.25	0.35	0.45	0.17	0.45	0.44	0.32
posa/daun	2	0.12	0.01	0.08	0.15	0.00	0.27	0.16	0.17	0.57	0.08	0.13	0.07
psme/cage	3	0.35	0.06	0.18	0.22	0.00	0.28	0.32	0.45	0.08	0.62	0.43	0.39
psme/caru	4	0.36	0.16	0.25	0.18	0.00	0.22	0.37	0.44	0.13	0.43	0.49	0.30
psme/phma	2	0.33	0.19	0.24	0.18	0.00	0.21	0.35	0.32	0.07	0.39	0.30	0.37

Appendix D. Fuzzy Analysis of Final Model

Appendix D. Fuzzy confusion matrix and data for the final model:

							Actual							
Predicted	dry abgr	abgr /clun	wet abgr	abla	mead.	pico/ caru	pipo/ cage	pipo/ caru	pose/ daun	psme/ cage	psme /caru	psme/ phma		
dry abgr	23.2	0	0.5	3.7	0	0	0	0	0	0.4	1.1	0	29	80.0
abgr/clun	0	6.1	0	0	0	0	1.9	0	0	0	0	0	8	76.0
wet abgr	0	0	0	0	0	0	0	0	0	0	0	0	0	0
abla	1.6	0	0.6	16.2	7.8	0	0	0	1.5	0	1.3	0	29	55.8
meadow	0	0	0	0	11.2	2.8	0	0	0	0	0	0	14	80.2
pico/caru	0	0	0	0	0	7.0	0	0	0	0	0	0	7	100.0
pipo/cage	3.0	0	0	1.5	0	0	40.0	2.0	0.7	2.9	1.7	0.2	52	76.9
pipo/caru	0.6	0	1.1	0	1.0	0	0.7	11.1	3.5	0	0	0	18	61.5
pose/daun	0	0	0	0	0	0	0	0	0	0	0	0	0	0
psme/cage	0	0	0	0	0	0	0	0	0	0	0	0	0	0
psme/caru	0	0	0	0	0	0	0	0	0	0	0	0	0	0
psme/phma	1.5	5.3	4.2	0	1.0	0	0	0.6	0	0	0.4	30.0	43	69.8
total	30.0	11.4	6.5	21.5	21.1	9.8	42.5	13.6	5.7	3.3	4.4	30.2	200	
corrected	77.5	53.5	0	75.4	53.3	71.7	93.9	81.1	0	0	0	99.5	144.8	72.4

Significant fuzzy bias:

type	predicted	actual	relative fraction
abla	29.0	21.5	.3512
meadow	14.0	21.1	-.3356
pipo/cage	52.0	42.5	.2222
pipo/caru	18.0	13.6	.3192
psme/phma	43.0	30.2	.4248

Number correct = 144.79 of 200.00
K_{HAT} = .6781

Appendix E. Fuzzy Analysis of Secondary Model

Appendix E. Fuzzy confusion matrix and data for the secondary model:

Actual

Predicted	dry abgr	abgr/ clun	wet abgr	abla	mead.	pico/ caru	pipo/ cage	pipo/ caru	pose/ daun	psme/ cage	psme/ caru	psme/ phma		
dry abgr	19.0	0	0	2.3	3.0	0	3.0	0.5	3.2	0	0	0	31	61.3
abgr/clun	1.3	29.2	0	0	0	0	1.9	3	0	2.7	0	4.9	43	67.8
wet abgr	1.2	0	20.6	2.1	1.0	0	0	0	0	0.7	2.4	0	28	73.4
abla	0	0	0	0	0	0	0	0	0	0	0	0	0	0
meadow	3.0	0	3.9	1.0	11.1	0	0	2.0	0	0	1.0	1.0	23	48.5
pico/caru	0	0	0	1.9	5.8	18.2	0.6	0	0.5	0	0	0	27	67.3
pipo/cage	2.5	1	0	0.8	0	0	36.2	1.8	2.2	1.5	2	0.3	48	75.3
pipo/caru	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pose/daun	0	0	0	0	0	0	0	0	0	0	0	0	0	0
psme/cage	0	0	0	0	0	0	0	0	0	0	0	0	0	0
psme/caru	0	0	0	0	0	0	0	0	0	0	0	0	0	0
psme/phma	0	0	0	0	0	0	0	0	0	0	0	0	0	0
total	27.0	30.1	24.5	8.1	20.9	18.2	41.7	7.3	5.8	4.9	5.4	6.1	200	
corrected	70.4	96.8	84.1	0	53.3	100	86.7	0	0	0	0	0	134	67.1

Significant fuzzy bias:

type	predicted	actual	relative fraction
dry abgr	31.0	27.0	.1487
abgr/clun	43.0	30.1	.4280
wet abgr	28.0	24.5	.1450
meadow	23.0	20.9	.1007
pico/caru	27.0	18.2	.4864
pipo/cage	48.0	41.7	.1513

Number correct = 134.17 of 200.01

K_{HAT} = .6151