North American Ecological Zone Classification for the UN Food and Agriculture Organization's Forest Resource Assessment 2000 Project: Map Compilation and Validation

Michelle Douville
Geography Department
McGill University, Montreal

November, 1999

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Master's of Science in Geography

0-612-64345-X

Acknowledgements

I would like to thank all of the people at the EROS Data Center and McGill University for giving me an invaluable opportunity to learn from them. Special thanks go to Jeff Danielson. Alisa Gallant. Tom Loveland. Bhaskar Ramachandran. Brad Reed. Eric Wood. and Bruce Wylie. for their resourcefulness. enthusiasm. and ideas. Thanks are also due to K.D. Singh and the United Nations Foods and Agricultural Organization for providing a global framework from which to gaze through. I would also like to express my gratitude to Ghislain Rompre for illuminating my French ancestry and for graciously translating my abstract. In addition. for their support and guidance. I would like to specially thank my advisors. John Lewis and Zhi-Liang Zhu.

Abstract

Classification and mapping of ecological zones on a global scale has been a topic of research for many years. This research looks at the development of a global spatial database of ecological zones for the FRA 2000 Report of the United Nations FAO. Besides evaluating the most appropriate type of classification scheme for this purpose, it explores and demonstrates how existing data, for the United States and Canada, can be reclassified to match the FAO classification scheme. Accuracy of mapping is a synergistic function of error, uncertainty, and quality. An assessment of the draft FAO Level II Ecological Zone map was performed which classifies 10-year average, bi-monthly, smoothed AVHRR-NDVI composites of the conterminous United States by applying linear discriminant and decision tree analyses. The results of the linear discriminant analysis were more significantly correlated to the FAO classes, although both approaches suggest that the classification scheme does maximize between-class variance of the NDVI temporal series.

Resumé

Le present ouvrage a pour but de decrire les moyens de bien representer les zones ecologiques a l'echelle globale. Cette etude fait partie integrante d'un projet a plus long terme de developpement d'une base de donnees spatiale des differentes zones ecologiques appropriees pour l'Organisation des Ressources Alimentaire et Agricoles des Nations Unies (FAO) a l'interieur du rapport sur l'Evaluation des Resources Forestieres 2000 (Forest Resource Assessment 2000 Report, FRA2000). Pour ce faire, nous avons evalue la methode la plus appropriee de schema classificatif utilise pour delimiter les zones ecologiques a l'echelle globale. Un exemple est fournit demontrant de facon graphique comment peuvent etre combinees et reclassifiees des donnees existantes des Etats-Unis et du Canada. Enfin, nous presentons les resultats de comparaisons de differentes methodes pour etablir la precision des cartes eco-regionales a l'echelle globale. Pour ce faire, nous avons utilise l'ebauche d'une carte eco-regionale de la FAO, representant dix ans de donnees NDVI unifiees des contours des etats americains, la moyenne valeur prises deux fois par mois, en y appliquant une analyse discriminatoire lineaire et un arbre decisionnel. Dans ce cas precis, l'analyse discriminatoire lineaire semble etre plus appropriee, meme si les deux approches suggerent que le schema classificatif maximise la variance entre-classe de la serie temporelle du NDVI.

Table of Contents

Acknowledgements
Abstract
List of Tables
List of Figures

1. **Introduction**

   1.1) **Aim and Objectives**
   1.2) **Structure of Thesis**
   1.3) **Background**

2. **Context: A Global Ecological Zone (EZ) map for the UN FAO FRA 2000**

   2.1) **Global Forest Inventories: new demands, current problems and potential improvements**
   2.2) **Lessons from the past: FRA 1990.**

3. **The Types of Classification for Vegetation Mapping**

   3.1) **Dividing up the Earth**
   3.2) **Floristic Classification Schemes**
   3.3) **Physiognomic Classification Schemes**
   3.4) **Ecological or Biogeocenose Classification Schemes**

4. **Conceptual Frameworks for Classifying Vegetation at the Global Scale**

   4.1) **Source map compilation**
   4.2) **Actual vs. Potential Vegetation**
   4.3) ***A Priori* and *A Posteriori* Schemes**
   4.4) **Hierarchical Classification Schemes**
   4.5) **Future Directions for Vegetation Mapping**
   4.6) **The FAO EZ Classification Scheme**
   4.7) **Existing North American Ecological Zone Maps**

5. **Methodology: Mapping and Assessing Ecological Zones**

   5.1) **Compiling Ecological Zone Maps: The USA and Canada Case Study**
   5.2) **Statistical Multivariate Pattern Recognition and Assessing Accuracy in the North American Ecological Zone Map**

6) Results and Discussion

6.1) A compiled North American Ecological Zone Map
6.2) Statistical Multivariate Pattern Recognition of NDVI composites of North America
6.2.1) The Stratified Random Sample
6.2.2) Stepwise Discriminant Analysis Results
6.2.3) Linear Discriminant Analysis Results
6.2.4) Decision Tree Analysis Results
6.2.5) Assessment of the compiled North American Ecological Zone Map based on Discriminant and Decision Tree Analyses Results

7) Conclusion

7.1) Summary
7.2) Future Research

Bibliography

Appendix A: Equations used by SAS for the calculating linear discriminant functions
Appendix B: Decision Tree Generated from 984-point stratified random sample using C5.0
Appendix C: Map of Predicted Group Membership from Linear Discriminant Analysis
Appendix D: Map of Predicted Group Membership from Decision Tree Analysis
Appendix E: Difference Map of Predicted Group Membership from Linear Discriminant Analysis
Appendix F: Difference Map of Predicted Group Membership from Decision Tree Analysis

List of Tables

Table 1: The proposed FAO Ecological Zone Classification Scheme

Table 2: Regional climates based on Köppen's system of classification with Bailey's Ecoregion equivalence

Table 3: Equivalence table for the CEC's Ecological Regions of Canada and the FAO Ecological Zones

Table 4: Equivalence table for Bailey's Ecoregion Divisions of the United States and the FAO Ecological Zones Level III

Table 5: Breakdown of the 984-point stratified random sample used as training data for the discriminant function and decision tree analysis

Table 6: Stepwise discriminant analysis results: Total amount of variance explained by each predictor variable

Table 7: Matrix of Mahalanobis (D^2) values for comparison between the 9 predicted and a prior ecological zones

Table 8: Linear discriminant constants and coefficients

Table 9: Resubstitution summary on the training data using linear discriminant functions

Table 10: Resubstitution summary on the entire study area (conterminous United States) using the linear discriminant functions

Table 11: Resubstitution summary on the training data using the decision tree analysis

Table 12: Resubstitution summary on the entire study area (conterminous United States) using the decision tree analysis

List of Figures

Figure 1: The Forest Resource Assessment 1990 Ecofloristic Zone Map

Figure 2: Bailey's Ecoregion Divisions of the United States

Figure 3: The Commission for Environmental Cooperation's Ecological Regions of North America Level II

Figure 4: The Draft FAO Ecological Zones of North America

Figure 5: Alaska and the Yukon Ecological Zone discrepancies: before changes were made

Figure 6: Alaska and the Yukon Ecological Zone discrepancies: after changes were made

Figure 7: Eastern Canada and United States Ecological Zone discrepancies: before changes were made

Figure 8: Eastern Canada and United States Ecological Zone discrepancies: after changes were made

Figure 9: Map of the 984-point stratified random sample and FAO Ecological Zones of the United States

Figure 10: Graph of total variance explained by predictor variables

Figure 11: Partial Flowchart of the decision tree trained from the 984-point stratified random sample

CHAPTER 1

1.) Introduction

Over the past thirty years, there has been an increasing emphasis on reporting various environmental statistics by 'ecological zone'. This emphasis is related to the perceived need that modeling natural processes, which are spatially defined by ecologically functional units, are required for making informed, environmentally appropriate, management decisions. Although much recognition has been given to the effects of global processes on regional conditions, few attempts have been made to delineate a global ecological zone map that is both consistent and accurate across the globe. The reasons for this are related to the fact that resource management has traditionally been conducted at regional to local scales, and that few globally-oriented organizational structures exist to conduct such a study. Nevertheless, one of the current mandates of the United Nation's Food and Agriculture Organization (FAO) is to produce a global ecological zone map. The purpose of this mandate, as will be described later, is related to the importance placed on reporting forest statistics by ecological zone within a global perspective.

1.1) Aim and objectives

The primary aim of this research is to develop a methodology for producing an accurate global spatial database of ecological zones for the Forest Resource Assessment 2000 Report (FRA2000) of the FAO. The objectives of this study are to:

Objective 1) Identify and evaluate the most appropriate type of classification scheme for a global ecological zone map.

This objective addresses the characteristics and constraints of a global ecological classification scheme. The high level of generalization at the global scale as well as the compilation-based approach necessitates the classification scheme be systematic, easily understood and reproducible. A better understanding of the constraints and characteristics helps to justify and explain the approach used.

Objective 2) Explore how existing data can best be combined and reclassified to achieve a coherent and consistent global ecological zone map

A USA/Canada case study is used to address this objective by illustrating how regional ecological maps differ in categorical definitions and underlying philosophies. Existing regional ecological zone systems are reclassified to fit the global ecological classification scheme, with the results reported using both maps and tables.

Objective 3) Investigate methodologies through which error, uncertainty, and quality factors can be assessed for a 'compiled' global ecological zone map.

An assessment of accuracy and error is essential for building confidence in the methods and results of a study. With this knowledge, users can judge the most appropriate application of the methods and results. An accuracy assessment is performed using 10-year average, bi-monthly, smoothed 1-km AVHRR-NDVI composites of the conterminous United States by applying discriminant and decision tree analyses, and compares them to the FAO Ecological Zone map.

1.2) Structure of Thesis

Chapter One establishes a basis for this thesis within the field of ecosystem geography, and introduces the research questions that will be addressed. Chapter Two provides the contextual basis for developing a global ecological zone map for the FRA 2000. Within a global perspective, the types of classification for regional vegetation and ecological zone mapping are discussed in Chapter Three. Chapter Four explains the conceptual frameworks and elements of a global classification system and introduces the FAO Ecological Zone classification scheme. Chapter Five describes the methodology used to reclassify and compile regional maps, and to assess accuracy and error in the resulting Ecological Zone map. The United States and Canada are used as a case study to illustrate the reclassification and compilation process. Only the area within the conterminous U.S. is evaluated for the accuracy assessment because of data availability. Chapter Six reports and discusses the results generated from Chapters Five and Six. Chapter Seven, the concluding chapter, discusses the results within the context of the FRA2000 as outlined in Chapters Two, Three and Four, and also revisits the stated research objectives.

1.3) Background

Since its inception in 1945, the FAO has offered direct development assistance by collecting, analyzing and disseminating information, by providing policy and planning advice to governments and by acting as an international forum for debate on food and agriculture issues (including forestry). A specific priority of the Organization is to encourage sustainable agriculture and rural development: a long-term strategy for the conservation and management of natural resources. Throughout the years of statistical reporting on global forests by the FAO, the aim has remained the same: to support policy formulation and investment decision-making by Member Governments of the UN. One of the objectives of this aim is to provide, on a regular basis, information regarding the world's forest resources. In relation to this, the underlying purpose of the various reports published through the FAO Forestry Department is to "contribute to the knowledge base on which reasonable decisions regarding sustainable forest management can be taken at the global level" (FAO, 1998). The Forest Resource Assessment 2000 (FRA 2000), FAO's next periodic evaluation of the world's forest resources, will provide a wide range of information on the state of forest cover for the year 2000. Information in the report will include indicators such as the extent and rates of change in deforestation, biomass, biological diversity, availability of forest area as a source of both timber and non-wood products, and the protective role that forests play in the landscape. Information will be reported at national, regional, and global levels, and will also be based on ecological zones. Information collection will be carried out by national and regional experts in specific disciplines, and complemented by GIS analyses of multiple data sources and remote sensing programs. The effort requires a high degree of collaboration and coordination between many national and international agencies with the results scheduled for release in January 2000. This approach, which requires the participation of many regional experts, promotes the use of regionally developed ecoregion boundaries for global modeling, while also providing a global perspective to countries that utilize the dataset for regional projects.

For the purpose of developing a global ecological zone map, the globe is divided into regions. The Forest Resource Assessment 1990 developed a tropical ecofloristic zone map that includes Africa, India, Southeast Asia and South America, and used similar procedures as the current assessment (Figure 1). Therefore, the emphasis of the FRA 2000 ecological zone mapping project is on preparing a map for temperate areas. Due to national influences on funding for ecological/vegetation mapping projects, these temperate regions are defined by political boundaries. In addition to areas covered by the FRA1990, there are seven study regions: Central American countries (including Mexico); the European Union; countries encompassed by the Former Soviet Union; China, Australia; the Far East; and North America. The latter, which encompasses Canada and the United States, is the focus of this thesis and is used to illustrate methodologies for compilation and validation that could be adapted to other regions.

2.) Context: A Global Ecological Zone (EZ) map for the UN FAO's Forest Resource Assessment 2000

In the past two decades, there has been an increase in demand for additional information about forests, forestry, and ecology (Persson and Janz, 1997). Traditionally, forest inventories, within a nation or sub-nation area, covered parameters such as forest area, wood volume, and wood increment (the amount of wood fibre accumulated through forest growth). In addition to traditional forest inventories, the Kotka III Expert Consultation, (held in Kotka, Finland in 1997), identified a need for better understanding the character and driving forces of forests for wood supply. The need for this type of knowledge relates to the growing awareness, within some policy formulation groups, of the inter-connected nature of global processes, natural or human. From this perspective, therefore, any attempt at managing these processes requires that ecologically functional units be delineated and used as a planning unit. With improvements in data collection procedures and technological capabilities, the capacity to collect, compile and analyze relevant global environmental information to meet this need, has also improved.

2.1) Global Forest Inventories: new demands, current problems and potential improvements

New demands for information have been driven by concerns about the environment, and the issues of deforestation which have been brought to the forefront with such initiatives as the World Commission on Environment and Development's Bruntland Commission Report (1987), and the 1972 and 1992 UN Conferences on Environment and Development. Within this context of current economic, cultural, and organizational conditions; compilation of existing datasets, and the use of readily available and inexpensive data sources is the approach chosen to develop and evaluate a global ecological zone map.

Although the compilation approach to mapping ecoregions is less expensive and more politically acceptable than alternative approaches, data aggregation is often made difficult due to source inventories from various national and international agencies with highly varying data collection methodologies. In conducting a study of ecological zone map compilation, a clear set of definitions for ecological and forest resource terms are needed. The definitions must have a number of characteristics if they are to be useful for ensuring consistency in statistical reports. First, the definitions must be consistent and compatible with terminology agreed upon in international fora. This was the purpose of the 1997 Kotka III Consultation and of the World Conservation Monitoring Centre (WCMC)/FAO Ecological Zone Workshop held in July of 1999. Secondly, the definitions must be flexible enough to permit the utilization of data sets collected for other purposes. Additionally, major terms must accommodate already collected country-level data. Each country has its own unique way of classifying ecological zones, and forest statistics, and the standard definitions must be able to accommodate these differences. Lastly, the terms must also be readily understandable to a wide audience of users (Bull, 1997).

The current state of global ecological and forest inventories has not changed considerably in the past few decades. Although survey tools and computer technologies have improved, there exist three main obstacles to further improvement in the next few years. They can best be captured by three important questions: First, do the tools, concepts and classifications match the objectives of the inventory? Second, how is the issue of international harmonization to be addressed? And thirdly, what role does the international community play in building survey capacity in less developed countries? (Lanly, 1997) Although the FRA 2000 will be reporting by ecological zone, it is unlikely, that it will be able to address these obstacles fully. Nonetheless, it is a worthwhile attempt to improve and build upon existing structures.

2.2) Lessons from the past: FRA 1990.

A spatial ecological zone database, especially for temperate areas, that is reliable and relevant to the reporting of forest statistics at the global level, has not yet been developed. Such a database of tropical areas, however, was prepared for the FRA 1990, which has proven to be useful and relevant for reporting on forested areas. The task for the FRA 2000 is to extend this map to include temperate areas. This initiative is related to new directions that have called for more emphasis on reporting statistics by ecological zone. This type of approach is important not only for modeling the effects of forest changes on natural processes globally, but also for comparing how different ecological zones respond to those changes.

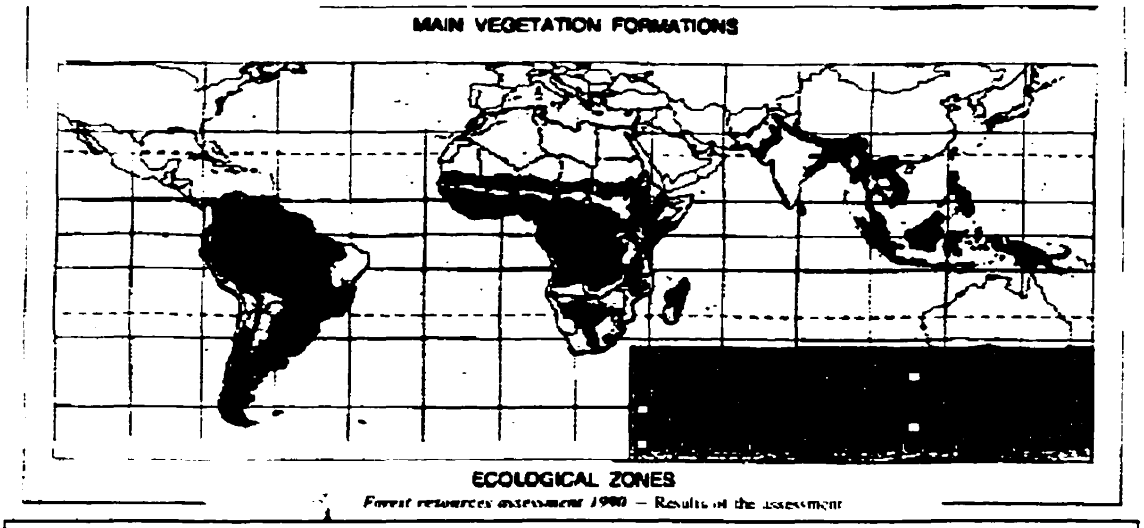


Figure 1: The Forest Resource Assessment 1990 Ecofloristic Zone Map (http://www.fao.org)

As an extension to and combination of sample survey approaches, and new computer-aided technology (including GIS), the 1990 Forest Resource Assessment used a remote sensing-based survey approach to map and report on tropical forests. This was a major global cooperative effort where institutions and individuals contributed to all phases of the implementation from statistical and analytical design, interpretation of satellite data, to the dissemination of the monitoring methodology to forestry institutions within tropical countries (FAO, 1997).

The database and map that was compiled is only a partial coverage, and excludes tropical areas within Mexico, Central America, China and the Near East. Although not a global coverage of ecoregions, this initiative proved to be a distinct advancement in reporting of forest information by functional ecological zones. Aside from the report proper, the information has proved to be very useful to other projects including the development of a deforestation model, the assessment of forest biomass, and the design of remote sensing-based sample surveys in the tropics (Singh, 1998). An example of one benefit of reporting forest statistics by ecoregion is the ability to compare two ecoregions that have an equivalent loss in forest cover. By delineating ecoregion boundaries, impacts on the loss of biological diversity and biomass flux in the two ecoregions could be monitored to compare ecosystem resilience to deforestation.

In addition to completing ecological zone maps for tropical areas not included in the FRA 1990, it was recommended, at the Kotka III meeting, that specific definitions of ecological zone classes for temperate and boreal zones would have to be specified for developed countries. Currently the FAO has no existing database on developed countries, although numerous maps and data sets have been compiled by other organizations which would serve as a logical basis for compilation (Zhu, 1998).

CHAPTER 3

3.) Types of Classification for Ecological Zone Mapping

3.1) Dividing up the Earth

The origins of ecological zone mapping stems from the science of vegetation mapping. This section will briefly describe the science of vegetation mapping, exploring how its terminology and lexicon have been perceived as ambiguous, and how it impacts and relates to ecological zone mapping.

The abundance of semantic ambiguity within vegetation mapping stems from numerous attempts and approaches to tackling the challenge of describing, classifying and explaining a phenomena which, by its very nature, is heterogeneous and multi-faceted. Even the very name used to describe the study of vegetation has varied as a function of which school of thought it derives from. They vary from *Vegetation Science* and *Botany*, to *Phytocenology* or *Phytosociology* (hereafter referred to as *Phytocenology*). The union of Phytocenology with geography places the work of phytocenologists within a spatial realm, often in the form of a vegetation or ecoregion map. The many types of maps represent the various means for integrating critically important biotic and abiotic factors.

With increasingly smaller scales, natural communities need to be mapped with broader generalizations for revealing regional correlations and global spatial distributions.

No single person invented the "vegetation map" but a great many people have contributed to the methodology of producing an image that scientifically reflects the distribution of Earth's vegetation. From the first maps made thousands of years ago, vegetation has been portrayed in some form or another, but not until the beginning of the 19th century were more scientific and systematic approaches taken to describe its distribution and character. Alexander von Humboldt, also known as the founder of modern geography, was one of the first to explicitly identify a correlation of physiognomic features to environmental zones and explore ideas of latitudinal variation in species variety (1815). This was all during a time when physiognomic aspects were rarely separated from the concept of floristics (De Laubenfels, 1975).

Küchler uses the term '*phytocenose*' ({*phyto~*} – plant, {*~cenose*} – common) to describe a unique plant community distinct from its corresponding animal community and environmental setting:

> "*A phytocenose may be defined as an aggregation of taxa which are capable of successfully competing with one another within the confines of a particular combination of environmental features they can tolerate.*" (Küchler, 1967)

There are three basic ways of classifying phytocenoses (Küchler and Zonneveld, 1988). The first approach is based on taxonomic or floristic identification of plants. The second and often less systematic approach, is based on classes related to morphological, structural or physiognomic features of plant life. The third approach is based on calibrating vegetation to one or two environmental variables that best describe their distribution. The following sections describe aspects of these classification schemes, how they are combined, and how applicable they are for global ecological zone mapping.

3.2) Floristic Classification Schemes

A *floristic* classification scheme is often based upon a 1950 international agreement for taxonomic units of vegetation. Schemes developed floristically are usually hierarchical and determined according to the binomial system that goes back to the 18th century Swedish scientist Carl Von Linne [Linnaeus] (Küchler and Zonneveld, 1988). The major levels of the taxonomic hierarchy are classes, orders, families, genera, species and subspecies.

A floristic vegetation map is a spatial representation that delineates boundaries of plant communities by floristic composition. The most widely known floristic classification scheme for mapping is the Braun-Blanquet method. In this system, floristic classes usually do not show all species within a plant community but are characterized by listing the more conspicuous or characteristic species of a community (e.g. white pine forest, oak forest). This often results in numerous categories, resulting in a general avoidance of floristic mapping at the macro scale. Other disadvantages of floristic maps include problems that arise for areas where there is species richness or unfamiliarity with the flora. Due to this, many floristic maps exist for Western Europe (familiar area, relatively few species) but in contrast, few exist for tropical areas (unfamiliar area, many species).

Large-scale floristic maps have been described as useful for tracking the historical extent of floristic compositions within an area due to the effects of historical events on species distribution (i.e. fire, avalanches, agricultural expansion, climate change). The greatest advantage of floristic approaches is that they permit a high degree of detail and accuracy at large scales, and are therefore preferred on small area maps where field surveys are logistically possible.

3.3) Physiognomic Classification Schemes

In contrast to a *floristic* approach, a *physiognomic* approach bases its scheme on appearance and growth form of vegetation rather than taxa. Two main types of physiognomic criteria are *growth form* (often called life form), and *structure*.

Growth form criteria can be interpreted as phenotypically or genetically fixed adaptations to the environment (e.g. grass with patches of broadleaf deciduous shrubs and dwarf shrubs). A growth form division of a phytocenose is technically referred to as a *synusia*, which is defined as a 'group of plants of one or several related life-forms growing under similar environmental conditions' (Küchler, 1967).

Leaf-form, height, and coverage of vegetation are the three most common criteria used to describe the *structure* of vegetation, which often require the delimitation of arbitrary boundaries for classification (e.g. canopy coverage class 1 = < 10%, canopy coverage class 2 = 10-50%, and so on) (Küchler and Zonneveld, 1988). The benefit of physiognomic criteria is that, for the most part, they are scale independent, and can be applied effectively over floristic realms and continents.

The value of a physiognomic approach is that it does not presuppose taxonomic knowledge, and can be used as an initial basis for a deeper investigation into floristic composition, environmental influences, and historical developments. In relation to this, another benefit of using physiognomic criteria is that remote-sensed data can represent, more directly, aspects of growth form and structure, than floristic attributes. Global vegetation and ecological zone maps, therefore, are generally the result of using some physiognomic criteria as the basis or as a component of the classification system used.

Many plant geographers suggest that for an optimized approach, both floristic and physiognomic characteristics should be combined (e.g. grassland-woodland conifers, shrubs-conifers). This approach may be effective at the regional or local scale, however, for a concise global vegetation map, this is somewhat debatable since the number of categories created from such a scheme could become too cumbersome and confusing. The tendency of authors is to become more selective and only choose features he/she considers essential to describe both structure and floristic composition. When this occurs, such selectivity varies, from one author to the next, one region to the other, and from one purpose to the next. If done unsystematically, this approach makes for difficult comparison if a compilation map were to be prepared.

3.4) Ecological or Biogeocenose Classification Schemes

The vegetation mapper is often called upon to map plant communities or *phytocenoses*, although, attempts to show more is not uncommon and phyto-***geo***-cenoses have been portrayed in many maps. A *phytogeocenose* is a plant community considered together with its physical environment (the terms ecoregion or ecological zone are often applied to this concept). However, on most maps that portray phytogeocenoses, the plant communities are not related to the entire complex of environmental factors but only to particularly prominent ones, such as climate or topography. As Küchler points out, it is difficult to map *phytogeocenoses* due to their complexity, and that often, *phytocenoses*, in conjunction with one or two environmental variables are used. The FRA 1990 classification scheme uses a similar strategy where, landform (e.g. lowland or montane) and precipitation (e.g. wet or dry) criteria are used to define ecological zones.

In 1898, Schimper developed a world 'ecological-vegetation' map, which assumes that occurrence and extent are based directly and perhaps exclusively on environmental conditions. Schimper suggested that the concepts of forest, grassland, and desert express directly, the concepts of wet, periodically wet, and dry climates respectively. One of Schimper's peers, Clements (1916), went even so far as to say that climate controls the vegetation to such an extent that, within the region of a given climate control, all types of vegetation evolve toward a uniform climax (Küchler, 1967). Concurrently with Schimper and Clements, Köppen used individual climatic factors (mean temperature and precipitation) to explain phytocenotic changes. For example, he used an isotherm representing areas where at least 4 months of the year had an average temperature above 10°C in conjunction with vegetation floristics and structure, to define the boundary between his temperate and boreal climate types. In 1900, Köppen developed a 'climatic regions' map of the world and since his climate regions are also based on vegetation boundaries, his map is as much a representation of vegetation as it is of climate.

One disadvantage of correlating climate to vegetation so strongly is that the parallels between a climate and a vegetation map can only exist if those climatic features responsible for vegetation changes are recorded. Some important factors, which prevent vegetation and climate maps from being co-developed, include the reaction of vegetation to changes in the soil as well as to anthropogenic influences. However, Köppen, with his adherence to temperature and precipitation variables, was able to state simply and clearly, the major aspects of climate important to the general distribution of vegetation. Interestingly, many current schemes are still based on similar approaches (e.g. Bailey's Ecoregions of North America, the FAO Global Ecological Zones, and Hou's Vegetation Classes of China).

Many research projects, often referred to as 'Ecoregion Mapping' projects, focus on the topic of calibrating vegetation to environmental variables. Some maps, referred to as 'vegetation maps', attempt to portray ecoregions or *biogeocenoses* and thus, by implication or direct statement, equate vegetation with the habitat, particularly climate (De Laubenfels, 1975). One of the reasons for this tendency is that vegetation 'faithfully portrays the character of the environment, making vegetation mapping an effective way to present the '*ecological order of our living space*' (Küchler and Zonneveld, 1988).

Although abiotic information has often accompanied vegetation maps since the early mapping days, such an accompaniment has not always been enlightening because of the author's subjectivity to often arbitrarily select a single controlling feature. In addition, critics of this approach, also point out problems that arise from the complexity of climate as an environmental factor on one hand, and the effects of species tolerance and intracommunity competition on the other. This point of view suggests that any attempts at coordinating climate and vegetation are doomed to failure. However, one could also argue that the problem lies in systematically choosing features from each of them, which are reasonably representative of the types they aim to describe.

In the early 1970s, along with concepts of holistics and developments in the study of ecology, the *land-unit* mapping approach was developed. Land-unit maps attempt to portray 'a more holistic view of the land with a more balanced emphasis on land ecology' (CEC, 1997). Not only are structure and floristics important, but equally so is the combination with other land data, such as soils, water, climate and relief. The Ecological Regions of North America, developed by the Commission for Environmental Cooperation (CEC) is an example of this *land-unit* mapping approach. It is advocated as an approach that recognizes the importance of considering a full range of physical and biotic characteristics to explain ecosystem regionality.

The inter-relationship and interaction of living organisms (plant and animal communities) and their abiotic environment leads to an exchange of materials and a flow of energy which constitutes what is referred to as an ecosystem, *ecoregion*, or less commonly, a *biome* or *biogeocenose*. What needs to be kept in mind, for map compilation purposes, is that while some authors have suggested that these terms are synonymous, others have delineated scale differences between them. For example, Walter describes a biogeocenose as a basic unit of smaller ecosystems and a biome that of a larger one (Walter, 1973). In contrast, Küchler describes a biogeocenose as synonymous with ecosystem at all scales. Nevertheless, regardless of scale and level of generalization, they represent the same concept: *a system formed by the interaction of a community of organisms with their physical environment.*

In summary, notwithstanding the varying uses of terminology, there are many approaches and objectives for classifying vegetation or ecoregions. The most effective approaches, however, are the ones that can be defended by strong theoretical concepts and definitions of phytocenology and ecology. However, as in all sciences, some degree of flexibility and subjective interpretation is required and inevitable, as Rowe and Sheard (1981) describe:

"The search for neat, mechanistic, cut-and-dried approaches to land classification is bound to be disappointing for it is based on the widespread misconception that classifications are serendipitous inductive methods by which factual data, crunched or massaged by handy algorithms, can be made to yield fruitful hypotheses."

CHAPTER 4

4.) Conceptual Frameworks for Classifying Ecoregions on a Global Scale

4.1) Source Map Compilation

Due to the logistical limitations of field sampling, and availability of remotely sensed data, small-scale maps often require compilation from source maps, and subsequent reclassification. The basic idea is simple: the author collects vegetation maps of the component parts of his area and combines them into one new map on a smaller scale (Küchler and Zonneveld, 1988). The art of compiling a vegetation map consists above all, in finding a way that leads from contrast and contradiction to unity and harmony. Contradictions in sources may require a considerable amount of research and correspondence.

Many pitfalls can be avoided by comparing map types in the light of a given purpose (Küchler, 1967). The decision to prepare a set of comparable vegetation maps produces problems of its own: What scale, area, and classification scheme should be selected? Producing a compiled small-scale ecoregion map requires the consideration of classification structure and type. Classification structure relates to whether the scheme is *hierarchical or non-hierarchical, a priori or a posteriori*. Classification type refers to whether the scheme is based on *physiognomic, floristic, climatic, holistic, or* *potential/actual* criteria. The most practical and efficient way to select maps is first to define the purpose of the map to be developed, and then identify which maps are useful based on that purpose. For the most part, the greater the resemblance of purposes and approaches between the input maps and the proposed compiled map, the better. The following sections describe the classification scheme structure and type that best suits a global ecological zone map.

4.2) Actual vs. Potential Vegetation

One issue when developing a global vegetation map is whether the source maps are representations of actual or potential vegetation, and whether they refer to anthropogenically modified vegetation. A map of actual vegetation represents the plant cover (natural or anthropogenic) at the moment of investigation. Although this may be useful in some circumstances, a major disadvantage of this approach is that the processes over time, or the dynamics of a *phytocenose* are ignored. One way of dealing with this problem is to consider the climax or potential of the natural vegetation (referred to as 'potential natural vegetation') (Kalkhoven and Van Der Werf, 1988).

Natural vegetation is the plant life that exists in the landscape unaffected by humans, and is traditionally considered to be in balance with the abiotic and biotic forces of its site. Therefore potential natural vegetation is related to *phytocenose*, but with two assumptions: 1) that humans are removed from the scene, and 2) that the 'resulting succession of plant communities is telescoped into a single moment in order to exclude the effects of climate change' (Küchler, 1967).

4.3) A *priori* and *a posteriori* schemes

Although most classification schemes for small scale vegetation maps and 'ecoregion' maps are largely based upon limited criteria for potential natural vegetation, another factor to consider is whether or not the scheme is based on an *a priori* or an *a posteriori* method (Brown *et al.*, 1980). *A priori* classification schemes are based on classes defined **before** the actual data collection takes place. In this way, classes can be standardized independent of the area and means used. The ability to standardize classes makes this approach useful at the global level for compilation purposes. In contrast, an *a posteriori* classification scheme is defined **after** clustering similarity or dissimilarity of samples. *A posteriori* approaches are often used to assess the validity of existing *a priori* classification schemes.

4.4) Hierarchical Classification Schemes

Many authors advocate that a hierarchical classification offers more flexibility because of its ability to accommodate different scales or levels of generalization of information (see Walter (1967), Adams (1997), and Bailey (1998)). To date, there have been several attempts to apply a hierarchical system of classification for *biogeocenoses* or *phytogeocenoses*. Examples include:

- Ecoregions of North America by Robert Bailey of the USDA Forest Service (1998)
- Ecological Regions of North America by the Commission for Environmental Cooperation (1997)
- National Ecological Framework for Canada by Canada's Ecological Stratification Working Group (1995)
- FGDC Vegetation Classification scheme by the U.S. Federal Geographic Data Committee(1997)
- FAO Land Cover Classification scheme by the UN Food and Agriculture Organization (1998)

Most of these examples emphasize structural/physiognomic characteristics at the broadest level, and floristic/species characteristics at the most specific level for defining classes within a hierarchy. Variations occur due to criteria and indicators used, as well as how abiotic factors are incorporated in the hierarchical system.

4.5) Future Directions for Vegetation Mapping

The most widely used methods for developing small-scale 'ecological zone' maps are currently based upon physiognomic/floristic-climatic and *land-unit* approaches. However, due to the difficulties of compiling land-unit maps in a systematic way, the more traditional approach of combining physiognomic and floristic features with one or two climatic variables is still favoured for developing a consistent global ecological zone map.

Several initiatives are currently taking place, which attempt to make approaches to *land-unit* mapping more systematic so that they can easily be incorporated into a map compilation project. Two examples include the FAO Land Cover Classification Scheme and the U.S. FGDC (draft) vegetation classification. Both schemes are hierarchical, independent of scale, *a priori*, and use a set of independent diagnostic criteria that include floristic, physiognomic and abiotic features. The diagnostic criteria systematically affect different levels of the hierarchy. Currently, there is collaboration between these two initiatives in order to harmonize their approaches (FGDC, 1997).

4.6) The FAO Ecological Zone (EZ) Classification Scheme

The approach used for developing an Ecological Zone (EZ) map supported by the UN's FAO, is based on compiling source maps to create a global coverage under a unified classification scheme. The Kotka III provided a forum for discussing the need for a global EZ map and classification scheme. The consensus at the meeting was that FRA 2000 should attempt to provide a break down of information on the state of forests, and on-going changes, by 'ecological zone'. Based on the purpose of the EZ map, some characteristics of an appropriate classification scheme can be formulated. Modifying from Adams' (1996) suggestions for an improved vegetation scheme for local and global mapping, the following section describes four basic features of an optimal universal classification scheme and relates them to the proposed FAO classification scheme in Table 1. In brief, four main qualities of an optimal universal classification scheme are:

Quality 1) *The scheme should be based on structural-physiognomic characteristics of potential natural vegetation combined with one or two climatic variables, such as temperature and precipitation.*

Quality 2) *There should be clearly defined limits for each of the categories*

Quality 3) *There should be sufficient categories to express variations, but not too many that it causes confusion.*

Quality 4) *The classification scheme should be an a priori- and a hierarchically-based one.*

Table 1: The proposed FAO Ecological Zone classification scheme (Level 1 and 2):
(Modified from the North American EZ translation table (provided by Zhu. 1998)).
Note: Level 3 (not shown) further divides classes of level 2 based on lowland vs. mountain landscape formation.

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Name	Characteristics	Name		Climatic Qualities	Phytocenose Qualities
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2. Boreal (E)	Up to 3 months over 10°C	Boreal	E		Evergreen. small leaves. closed canopy. pure stands or open woodlands
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		Continental	Dc	Coldest month under 0°C	Broadleaf deciduous forest. mixed forest
		[illegible]	[illegible]	[illegible]	[illegible]
		Arid	BW1	All months dry (or 1/2 the precipitation of BS1)	Xerophytic shrub
		[illegible]	[illegible]	[illegible]	[illegible]
		Summer Dry	Cs	Over 2 months dry (in summer)	Hardleaved evergreen trees and shrubs called sclerophyll forest
		[illegible]	[illegible]	[illegible]	[illegible]
		Arid	BW2	All months dry (or 1/2 the precipitation of BS2)	Xerophytic plants that are widely dispersed and provide negligible ground cover
		[illegible]	[illegible]	[illegible]	[illegible]
		Winter Dry	Aw	Over 2 months dry (in winter)	Deciduous forests. woodlands and savanna
		[illegible]	[illegible]	[illegible]	[illegible]
		Arid	BW3	All months dry (or 1/2 the precipitation of BS3)	Desert

In relation to the first quality. Adams suggests that structural characteristics such as height. cover. and deciduousness should be used at the global scale since they are easy to standardize and are recognize both in the field and in remotely-sensed imagery. Floristic criteria. however. is considered indirectly since it does provide an indicator of structure-type (e.g. sagebrush = dwarf shrub). Structural properties provide a common basis with which to correlate and compile separate floristic-physiognomic source maps of different areas into a global scheme. The complexity of local and regional fluctuations in land use and land cover. as a result of a variety of factors. suggests that the approach should also focus on the concept of 'potential natural vegetation.' Therefore the first quality of a global classification scheme is that it should be based on structural-physiognomic characteristics of potential natural vegetation combined with one or two climatic variables. namely temperature and precipitation. As Table 1 illustrates. the FAO EZ classes are based on temperature and precipitation ranges at the first and second level of the hierarchy. However. structural properties of vegetation contribute to the delineation of classes only at the second and third levels of the hierarchy.

The second quality of the global scheme relates to the limitations of input data sources. There needs to be a degree of flexibility to help reduce the need for either modifying boundaries or re-mapping during map compilation. However. clearly defined limits between categories need to be made explicit. The FAO Ecological Zone scheme addresses these factors with its emphasis on generalized climate-based descriptors. such as humid. semi-arid. or arid. each with temperature and precipitation limits.

The third quality of a global scheme is that it should have sufficient categories to express variations. but not too many that it causes confusion. In addition to broad climatic and vegetation characteristics. the FAO scheme differentiates between mountainous and flat geographical areas at the third level of classification for each of the classes (not shown in table). It was felt. for the purpose of the forest resource assessment project. that it would be important to differentiate between these two landscape types for reporting statistics.

Lastly, due to the frequent lack of systematic (i.e. non-explicit) methodologies in *land-unit* mapping, the more traditional approach of combining physiognomic features with one or two climatic variables is still favored for developing a comparable global EZ map. The first two levels of classification are based purely upon Köppen's climatic classification system. The approach provides a hierarchical, *a priori* and systematic approach with which to utilize temperature and precipitation data for the translation of documented climatic-vegetation source maps (quality 4).

The emphasis on developing an *a priori* classification scheme relates to the importance of deductively identifying map units according to theoretical relationships between landform, climate, biota and soils. This is in contrast to a more inductive approach of an *a posteriori* classification (e.g. cluster analysis). The *a priori* approach has often been called 'subjective', while the *a posteriori* approach 'objective'. This subjective-objective dichotomy loses its meaning when the subjective motivation and orientation, which overrides all intellectual activities, are considered. The virtue of 'objectivity' is really **explicitness** in the interests of mutual understanding (Rowe and Sheard, 1981). In relation to this view, an *a priori* approach can be considered to be more applicable to the **development** of a global classification scheme, whereas an *a posteriori* approach can be considered to be more effective in **testing and validating** the theories upon which the *a priori* approach is based. An *a posteriori* approach, using multivariate pattern recognition of NDVI, is used for validating the *a priori* FAO Ecological Classification scheme, and will be discussed in subsequent sections.

4.7) Existing North American Ecological Zone Maps

Current Ecological Zone mapping activities, conducted by FAO, are concerned only with geographic regions outside the FRA 1990 EZ map (Figure 1). For each region, the FAO will organize a study of structural, floristic, and climatic characteristics by using short-term vegetation experts from appropriate geographic regions. Using guidelines set by the

FAO, and with the support of the U.S. Geological Survey's EROS Data Center, delineation of EZ classes for each geographic region will be conducted. This will entail taking available maps, and seeing how boundaries match the FAO guidelines with respect to both bioclimatic and floristic characteristics. Within this context, developing an EZ map for the United States and Canada is just one component of the FRA 2000 project. The short term goal of this component is to provide a case study for developing a worldwide EZ map and database that is appropriate for reporting on forestry statistics. This case study was presented to the participants and experts at the EZ workshop of the FRA 2000 in Cambridge, UK in July 1999.

Existing ecological zone datasets for the United States and Canada provide an illustration of variations in ecological maps due to differences in categorical definitions and philosophies. The most widely used data sets are those developed under the direction of Robert Bailey (USDA Forest Service), and the Canadian Ecological Stratification Working Group. These two datasets were used as sources for developing a globally-oriented North American Ecological Zone map.

Bailey's ecoregion map for the United States uses a hierarchical scheme modified from Crowley (1967), and uses climate and vegetation as indicators of the extent of each unit. Similar to the FAO Ecological Zone scheme, the two most general levels or categories (domains and divisions) are based on the large ecological climate zones identified by Köppen (1931)(Table 2). Climate is emphasized at the broadest levels because of its 'overriding effect on the composition and productivity of ecosystems' (Bailey, 1995). Each is subdivided, on the basis of vegetation macrofeatures, into 'provinces', which express a more refined climatic difference than the 'domains' and 'divisions' (Figure 2).

Table 2: Regional climates, based on the Köppen system of classification (1931), as modified by Trewartha (1968) and Trewartha and others (1967) with Bailey's Equivalence

Köppen group and types	Bailey Ecoregion equivalents
A—Tropical Humid Climates	**Humid Tropical Domain (400)**
Tropical wet (Ar)	Rainforest division (420)
Tropical wet-dry (Aw)	Savanna division (410)
B—Dry Climates	**Dry Domain (300)**
Tropical/subtropical semiarid (BSh)	Tropical/subtropical steppe division (310)
Tropical/subtropical arid (BWk)	Tropical/subtropical desert division (320)
Temperate semiarid (BSk)	Temperate steppe division (330)
Temperate arid (BWk)	Temperate desert division (340)
C—Subtropical Climates	**Humid Temperate Domain (200)**
Subtropical dry summer (Cs)	Mediterranean division (260)
Humid subtropical (Cf)	Subtropical division (230)
	Prairie division (250)*
D—Temperate climates	
Temperate oceanic (Do)	Marine division (240)
Temperate continental, warm summer (Dca)	Hot continental division (220)
	Prairie division (250)*
Temperate continental, cool summer (Dcb)	Warm continental division (210)
	Prairie division (250)*
E—Boreal climates	**Polar Domain (100)**
Subarctic (E)	Subarctic division (130)
F—Polar Climates	
Tundra (Ft)	Tundra division (120)
Ice cap (Fi)	

* Köppen did not recognize the Prairie as a distinct climatic type. The ecoregion classification system represents it at the arid sides of the Cf, Dca, and Dcb types.

Definitions and Boundaries of Köppen - Trewartha System

Ar	All months above 64°F (18°C) and no dry season.
Aw	Same as Ar, but with 2 months dry in winter.
BSh	Potential evaporation exceeds precipitation, and all months above 32°F (0°C).
BWh	One-half the precipitation of BSh, and all months above 32°F (0°C).
BSk	Same as BSh, but with at least 1 month below 32°F (0°C).
BWk	Same as BWh, but with at least 1 month below 32°F (0°C).
Cs	8 months 50°F (10°C) or more, coldest month below 64°F (18°C), and summer dry.
Cf	Same as Cs, but no dry season.
Do	4 to 7 months above 50°F (10°C), coldest month above 32°F(0°C).
Dca	4to7 months above 50°F(10°C), coldest month below 32°F(0°C), warmest month above 72°F(22°C).
Dcb	Same as Dca, but warmest month below 72°F (22°C).
E	Up to 3 months above 50°F (10°C).
Ft	All months below 50°F (10°C).
Fi	All months below 32°F (0°C).

A/C boundary = Equatorial limits of frost; in marine locations, the isotherm of 64°F (18°C) for coolest month.
C/D boundary = 8 months 50°F (10°C).
D/E boundary = 4 months 50°F (10°C).
E/F boundary = 50°F (10°C) for warmest month.
B/A, B/C, B/D, B/E boundary = Potential evaporation equals precipitation.

Source: USDA Forest Service (http:/www.fs.fed.us/colorimagemap)

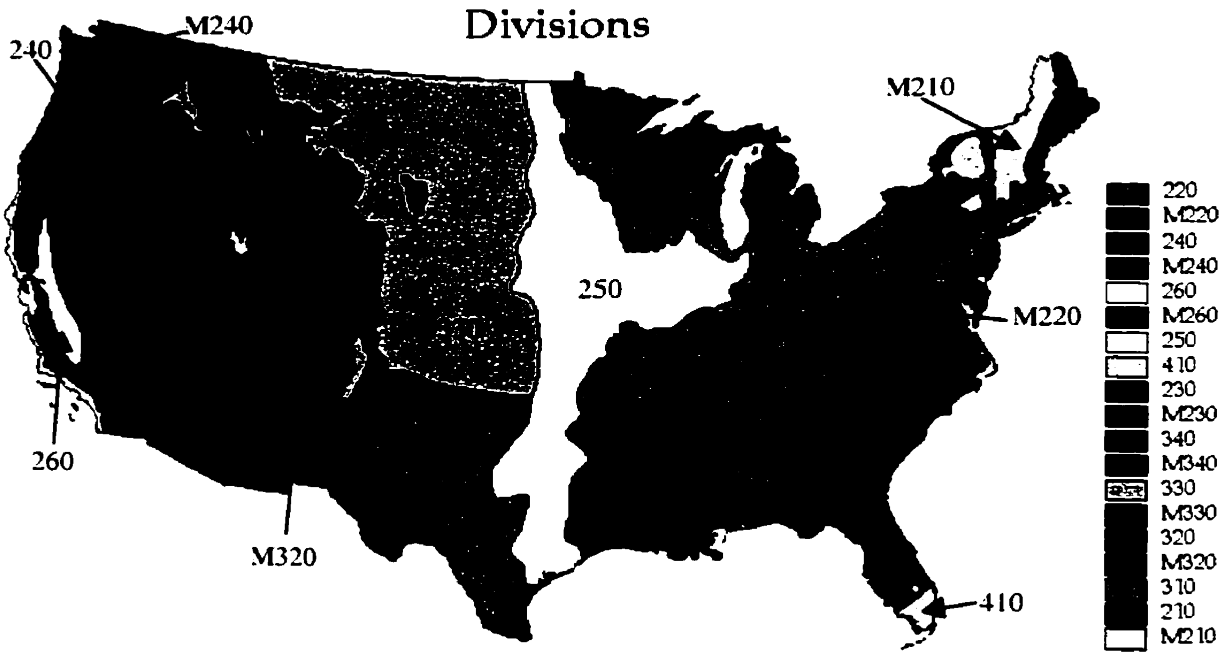


Figure 2: Bailey's Ecoregion Divisions of the United States
(see Table 2 for definition of codes, M denotes equivalent mountain division)

The approaches taken to delineate ecoregion boundaries by James Omernik in the United States, and the Ecological Stratification Working Group in Canada are somewhat different. Köppen's climatic zones were not used in any substantive way to define the boundaries. The approach is based on the belief that climate-based classifications alone do not provide enough flexibility or meaning. The authors of these initiatives have described the methodology as a more 'holistic' land-unit approach, which is considered to be an improvement upon the older systems of classification. The argument used is that the earlier ecoregion classifications evolved from schemes based on forests and climate without considering the full range of physical and biotic characteristics. The classification systems used by Omernik and by the Ecological Stratification Working Group attempt to recognize that ecosystems of any size or level are not always dominated by one particular factor, such as climate (CEC. 1997).

Ed Wiken, State of the Environment Directorate for Environment Canada (1998), provides an example of this ecosystem characteristic which appears to occur across Northern Ontario, Manitoba, and Quebec. He states that the interaction of climate with ground conditions result in very different forest ecosystem types.[1] So even at the macro scale, when climate interacts with the Hudson Plains (wet clay soils) vs. the Canadian Shield (shallow sandy soils and rock outcrop), the net results produce distinct ecosystems in the classification system.

As a follow-up to the work done by the Ecological Stratification Working Group in Canada, the Canadian Council on Ecological Areas (CCEA) led and coordinated the development of a North American ecosystem framework. This project was in response to initiatives of the Trilateral Committee on Environmental Information established by the United States, Mexican and Canadian governments. Subsequently, the Commission for Environmental Cooperation (CEC) provided further opportunities to enhance and complete this research, resulting in a compiled map of North America (Figure 3).

[1] Personal correspondence with Ed Wiken.

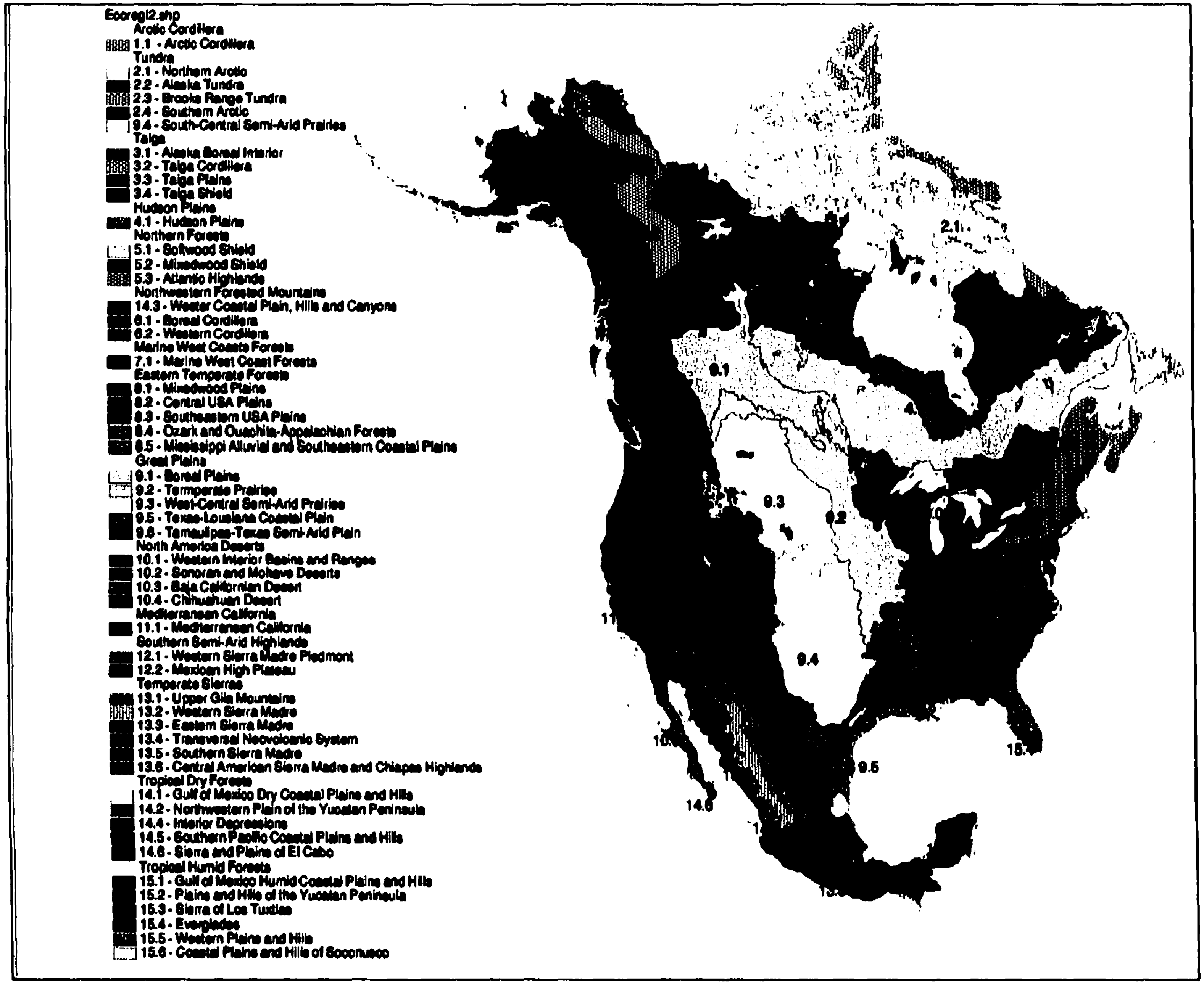


Figure 3: Commission for Environmental Cooperation's Ecological Regions of North America Level II

CHAPTER 5

5.) Methodology: Compiling and Assessing Ecological Zones of North America

5.1) Compiling Ecological Zone Maps: The USA and Canada Case Study

The first attempt at developing an EZ map for the United States and Canada involved using the most common data sources: Bailey's Ecoregions for the United States (Figure 2, including Alaska) and the CEC's Ecological Regions for Canada (Figure 3). Instead of using the entire CEC's North American map, Bailey's ecoregions for the United States were chosen because of the scheme's direct reference to Köppen's climate classes, and hence to the FAO classification scheme. Bailey's 2nd-level of classification ('Divisions'), were reclassified to the FAO Classification Scheme using the translation table that compares Köppen classes to Bailey's (Table 2). Subsequently, a contingency table for the CEC's Ecoregions of Canada was developed by comparing Bailey's reclassed regions to the CEC's class descriptions.

In Table 2, for example, Bailey describes his Humid Domain: Marine Division (240) as being equivalent to Köppen's Temperate Oceanic class (Do) (equivalent to the FAO's Level II 'Do' class). Köppen's 'Do' climate class is defined for areas where '4 to 7 months of the year are above 10°C, with the coldest month above 0°C'. For Bailey's equivalent class, climate and vegetation characteristics are given as follows:

> *" The average temperature of the warmest month is below 72°F (22°C), but at least 4 months per year have an average temperature of 50°F (10°C). The average temperature during the coldest month of the year is above 32°F (0°C). Precipitation is abundant throughout the year, but is markedly reduced during summer. Coastal mountain ranges influence precipitation markedly in these middle latitudes. The mountainous coasts of British Columbia and Alaska annually receive 60 to 80 in (1,530 to 2,040 mm) of precipitation and more. Natural vegetation in the Marine Division is needleleaf forest. In the coastal ranges of the Pacific Northwest, Douglas-fir, red cedar, and spruce grow to magnificent heights, forming some of the densest of all coniferous forests with some of the world's largest trees. "*[2]

[2] http://www.fs.fed.us/colorimagemap/ecoreg1_divisions.html, accessed August 1999.

Similarly, the CEC describes the overlapping and adjacent Marine West Coast Forest ecoregion as follows:

" *[A] maritime influence is responsible for a high level of precipitation, long growing season and moderate temperatures. Mean annual temperature range from 5°C in the north to 9°C in north California. The mean summer temperature ranges from 10°C in the north to 16°C in the south, whereas winter temperatures range from −1°C to −3°C. The annual precipitation ranges from as little as 600 mm in the gulf and San Juan Islands to over 5,000 mm along the north coast of British Columbia and Alaska. Overall, the windward slopes typically receive about 1,500 to 3,000 mm of precipitation per year. Variations in altitude create widely contrasting ecological zones within the region. They range from mild, humid coastal rain forest to cool boreal forests and alpine conditions at higher elevations. The temperate coast forests are composed of mixtures of western red cedar, yellow cedar, western hemlock, Douglas Fir, amabalis fir, Sitka spruce, California redwood and red alder. In the drier rain-shadow areas, Garry Oak and Pacific madrone occur with Douglas Fir. Sub-alpine areas are characterized by mountain hemlock and amabalis fir. Alpine tundra conditions are too severe for growth of most woody plants except in dwarf form. This zone is dominated by shrubs, herbs, mosses, and lichens.* " (CEC, 1997)[3]

As the descriptions illustrate, both schemes describe the class with similar precipitation and vegetation characteristics. However, the CEC's and Bailey's schemes use different climatic criteria to describe temperature regime. Following Köppen, Bailey uses two variables: the number of months above a certain average temperature and coldest month mean temperate. Alternatively, the CEC describes classes with temperature variables based on mean summer, winter and annual temperature ranges. Some comparison, interpretation and flexibility are needed to accommodate these differences so that they can be incorporated into the compiled map. For the West Coast example, by comparing definitions and by consulting the maps, the CEC's Marine West Coast Forest ecoregion was reclassed to the Temperate Oceanic (Do) class within the FAO Classification Scheme. Using similar comparisons, the remaining CEC ecoregions were reclassed to fit the FAO classification scheme. This approach provides flexibility for map compilation and places importance on having and comparing relevant documentation for existing classification schemes so that they can be translated to the Kotka III compliant FAO ecological classification system. Through semi-qualitative comparisons of class descriptions, existing ecological zone maps, such as the CEC's, can be incorporated into a global map. However, one disadvantage of this approach is that it restricts the number

[3] http://www.cec.org/ecomaps/eco_eng.pdf, accessed August 1999.

of ecoregion maps that can be translated to ones that are based on and documented with comparable biophysical and phenological criteria. In light of such a semi-qualitative approach, this study takes a more quantitative approach to assessing the results of the compilation. The following section describes a methodology for assessing a portion of the FAO North American Ecological Zone map by comparing it to an independent data source.

5.2) Statistical Multivariate Pattern Recognition and Accuracy Assessment of the North American Ecological Zone Map

One approach to evaluating the results of an ecological zone mapping exercise is to relate how well the ecological zone boundaries correspond to 'homogenous' geographic regions of *net primary productivity* (Bailey, 1984). Ecological zone maps are often used for assessment and resource management purposes that often require estimates of net primary productivity (NPP). Therefore, one way for assessing the quality of an EZ map is to measure how well it reflects homogenous NPP units. The working definition of NPP for this purpose is as follows:

"Basic or net primary productivity of an ecological system, community, or any part thereof, is defined as the rate at which radiant energy is stored by photosynthetic activity of producer organisms (chiefly green plants)." (Odum, 1971)

There are many approaches for assessing the accuracy of an ecological classification scheme. For example, in 1981, Rowe and Sheard used frequency and distribution of plant species for ecological productivity pattern recognition for a large-scale study area in the NorthWest Territories. In 1984, as an alternative approach, Bailey used an indicator of hydrologic productivity (i.e. runoff per unit area) measured at fifty-three USGS hydrological benchmark stations to distinguish between two broad ecosystems of the conterminous United States (Humid vs. Dry Domain). In both cases and scales, it was assumed that the assessment criteria (plant species or hydrologic runoff) were adequately correlated to NPP.

One criterion often used to assess land classification, is the normalized difference vegetation index (NDVI) derived from National Oceanic and Atmospheric Administration's Advanced Very High-Resolution Radiometer (AVHRR) data. NDVI, often referred to as the greenness index, is derived from the reflected solar radiation in the near-infrared (NIR) and red (RED) wavelength bands via the formula:

$$NDVI = (NIR - RED)/(NIR + RED)$$

It is assumed that NDVI can be used as an adequate estimator of NPP since it is sensitive to the quantity of actively photosynthesizing biomass. This characteristic makes it useful for assessing changes in seasonal vegetation phenology (Hardy and Burgan, 1999). Reed *et al.* (1994) computed twelve phenologically linked metrics based on time series NDVI to identify major land-cover types which are indicative of ecologically homogenous zones. A separate study of the Iberian Peninsula, by Lobo, Ibáñez-Martí, and Giménez (1997) suggests that temporal series of NDVI yield relevant ecological information at a scale large enough to be suitable for regional applications. As Hardy and Burgan illustrate, however, NDVI responds to complex physiological changes in vegetation, which limits its usefulness for singling out influencing factors (e.g. effects of moisture content). For a general picture of temporal and spatial NPP variation, a NDVI time series can be used as an independent data source for validating ecological zone classifications.

NDVI data are typically very noisy, and are affected by numerous phenomena, including cloud contamination, atmospheric perturbations, variable illumination and viewing geometry, each of which reduces the value of NDVI. Swets *et al.* (1999), applied a weighted least-squares linear regression approach to temporal smoothing to more efficiently reduce contamination in the NDVI signal of the conterminous United States at 1.1-km resolution. The original purpose of the smoothed data was for improving applications involving time-series NDVI, such as land cover classification, seasonal vegetation characterization, and vegetation monitoring. It is this dataset, developed by

Swets *et al.*, which is used as an independent data source for validating the North American Ecological Zone map. Regardless of spectral noise, they point out that NDVI has many benefits:

> *'One of the most commonly used vegetation indices, the normalized difference vegetation index (NDVI) takes advantage of the reflective and absorptive characteristics of plants in the red and near-infrared portions of the electromagnetic spectrum and has been used in research on vegetation productivity. The combination of the NDVI with the frequent temporal coverage and moderate spatial resolution of the advanced very high resolution radiometer (AVHRR) makes this sensor well suited for regional- to global-scale studies on ecosystem dynamics.'*

The approach used by Swets *et al.* (1999), was to first calculate the maximum value composite for a 14-day time period to reduce both cloud contamination and data volume, a common procedure performed on NDVI data. After this initial step, to further reduce residual effects due to sub-pixel clouds, prolonged cloudiness, persistent haze and other negative effects, a weighted least-squares windowed linear regression approach to temporal NDVI smoothing was used. The resulting smoothed curve statistically binds the results to the original raw data points. They suggest that land classification schemes operating on smoothed data may reduce the number of misclassified pixels due to one or more noisy values in the unsmoothed, multi-temporal satellite data.

The resulting dataset derived from the processing described above consists of three files in band sequential (BSQ) format, each approximately one gigabyte, and containing 260 bands (one for each bi-weekly 1.1-km composite from January 1989 to December 1998). The areal extents of the three files consist of strips encompassing western, mid-western, and eastern USA; each in Lambert Azimuthal Equal-Area projection with a latitudinal origin of 50°N and a longitudinal origin of 100°W.

As a validation tool. this dataset was chosen for several reasons. which include:

- the theoretical ability to relate NDVI to net primary productivity
- the applicability of the approach used to smooth the data
- the similarity of purpose between the NDVI dataset and the accuracy assessment
- the extensive. up-to-date and complete nature of the data (i.e. bi-weekly composites that are consecutive from January 1989 through December 1998).

The 260 bi-monthly smoothed NDVI composites for each of the three areal extents (western. mid-western. eastern United States). were averaged into twenty-six bands representing 10-year averages for each two-week period. This averaging was performed for two reasons. The first was to reduce data volume and processing requirements. The second was to reduce the effects of annual fluctuations on longer-term NPP trends.

The 26-band images for each of the three areal extents were then merged together. This resulted in a 26-band image of NDVI for the entire study area (conterminous United States) to be used as an independent data source to assess the FAO Ecological Zone map of the United States. A water mask derived from the Version 1.2 release of the International Geosphere-Biosphere Programme (IGBP) North America land cover characteristics database[4] was used to set all values of water bodies to *zero*. Assigning the value *zero* to all water areas assures that these regions are filtered out in the analysis (Eastman and Fulk. 1993).

The Level II FAO Ecological Zone Map developed for the United States. that was developed using the methodology described in the preceding chapters, was reprojected and rasterized to match the 26-band NDVI image. To see if NDVI values could be used to discriminate between ecological zones. an appropriate sample size needed to be chosen. For this purpose. it was assumed that the time series NDVI dataset followed a

[4] http://edcwww.cr.usgs.gov/landdaac/glcc/glcc.html. accessed August 1999. *Earth Resources Observation Systems (EROS) Data Center. Land Processes Distributed Active Archive Center (DAAC)*

multivariate normal distribution. Therefore, following the general guidelines for the number of observations required for a 'one-way analysis of variance' (ANOVA), which is computationally similar to linear discriminant analysis, an initial sample size of 1000 was chosen. In an attempt to address the issue of different ecological zone sizes, a stratified random sample approach was used. A 984-point stratified random sample, based on the nine FAO Ecological Zones for the United States, was generated using ERDAS Imagine software (6 points fell outside the study area). To test the null hypothesis, which states that the ecological zone NDVI means are not significantly different from each other, a sample size of 984 allows for an alpha (α) = 0.05, and an average power ($1-\beta$) of 0.80 (see Gatsonis and Sampson, 1989; and Taylor and Muller, 1995).[5] If the null hypothesis is true, and the mean of the NDVI values for each class are not significantly different, then the results of the discriminant functions and decision trees would be meaningless.

In addition to using a NPP indicator, Bailey (1984) suggested linear discriminant analysis (LDA), a multivariate statistical technique related to ordinary least squares regression and analysis of variance (ANOVA), as an approach for assessing *a priori* ecological boundaries. With the NDVI data ready for processing, stepwise linear discriminant analysis was performed on the 984-point sample as a first step towards testing the compiled North American Ecological Zone map.

[5] Alpha represents the probability of rejecting the null hypothesis when it is in fact true (TYPE I Error). The power of the test represents the probability of accepting the null hypothesis when it is in fact false (TYPE II Error).

Stepwise discriminant analysis (SDA) was performed on the sample to identify which of the 26 bands best characterize the spatial and temporal variance within the entire dataset. This step helps develop an efficient model that effectively predicts which group (e.g. ecological zone) a case or sample point belongs to and is similar to the process used in stepwise regression. With SDA, all the variables are reviewed and evaluated with reference to its relationship to other variables for determining which contribute most to the discrimination between groups. At each step within SDA, the variable that contributes the most to explaining variance is added to the model, and the iterative review and evaluation process starts again. The process is repeated until little or no variance is explained from further addition of variables and only the 'noise' or error variance remains.

In SDA, the stepwise procedure is "guided" by two respective F values: a F to enter and a F to remove value. For a forward SDA approach, which is the approach used here, the variable with the highest F value is added to the model. The F value is a measure of the extent to which a variable makes a unique contribution to the prediction of group membership (Jennrich, 1977). The F-statistic is essentially computed as the ratio of the **between-groups** generalized variance over the pooled (averaged) **within-group** generalized variance. If the between-group variance is significantly larger then it is assumed that there must be significant differences between means. In other words, the F-statistic is a value that represents how much each NDVI band contributes to explaining the differences in NDVI values between ecological zones.

The F-statistic, in conjunction with the multivariate correlation coefficient called Wilks' Lambda, provides an indicator of the correlation, between the predictor variables within the model and the dependent variable (*a priori* Ecological Zones). As a complement to the F-statistic, for each additional variable added to the model, the Wilk's Lambda is calculated by dividing the pooled **within-group** generalized variance by the **total** generalized variance. A value closer to *one* indicates poor separation between groups.

while values closer to *zero* indicate good separation at least between some groups. This characteristic is opposite to other commonly used correlations coefficients (Jennrich. 1977). Regardless of the reversed logic. the Wilks' Lambda can be interpreted as the multivariate counterpart of a univariate R-squared. that is. it indicates the proportion of generalized variance in the dependent variables (ecological zones) that is accounted for by the predictors (NDVI bands).

As the SDA procedure progresses. both the F-statistic and the Wilks' Lambda values should become smaller. This trend indicates that. as more and more variables are added. total and between-group variance increases. whereas pooled within-group variance decreases. However. for every additional variable added. less and less of the remaining unexplained total variance can be explained by the newly added variables. This explains why the *F*-statistic decreases. and also suggests that the remaining unexplained total variance is merely 'noise'. Both tests of correlation have corresponding p-values which represent the probability of error that is involved in accepting the observed result as valid. or as "representative of the population" (Lachenbruch. 1975). In this case. a p-value of .001 for the *F*-statistic was used to define which variables were to be used in the classification process.

Using the variables defined by the stepwise process and the 984-point sample. Linear Discriminant Analysis (LDA). was used to build discriminant functions. also known as group classification functions. for each of the ecological zones (Hand. 1997). LDA simultaneously analyses the predictor variables and identifies "patterns" of values of those variables. Technically. it determines a linear combination of the predictor variables that best predicts group membership. LDA uses estimates of significant temporal and spatial differences within NDVI to define the limits of variation between and within zones.

For each group (ecological zone). LDA produces a set of coefficients by defining a single linear combination of variables that best differentiates each class. Using generalized

(Mahalanobis) distances between group means. discriminant coefficients for each predictor variable are derived using a least-squares approach (see Appendix A for a description of the equations used by SAS). The values or discriminant scores, for each discriminant function are calculated for each pixel within the study area (i.e. each 1.1 square kilometre within the conterminous United States). In this way, a pixel is assigned a class with the highest discriminant score, or in other words, the closest class centroid measured in generalized distance.

The discriminant score for each pixel in the study area is computed as a composite of each measurement of the predictor variables, weighted by the respective discriminant function coefficients. The larger the coefficient, the greater is the contribution of the respective variable to the discrimination between groups. In addition to understanding the relationship of the predictor variables to the resulting classification, LDA contributes to understanding the strength of the relationship between classes measured in relative generalized distances between class means. Good separability between classes, as identified by the distance between the class means, suggests that the discriminant functions are successful at distinguishing between classes.

LDA was used to classify pixels into values of a categorical dependent (ecological zones) based on a set of predictor variables (a time series of NDVI). With good separability between class means and a high percentage of pixels classified into the same ecological zone as the *a priori* FAO classes, two inferences can be made. The first is that the discriminant functions derived are effective at delineating small-scale ecological zones using the NDVI inputs. The second is that both classifications are adequate representations of small-scale NPP trends. There are two main reasons why LDA was chosen as a method of assessment:

- The first was to classify pixels into classes using a traditional linear multivariate technique. The approach builds from the method used by Bailey (1984) to binomially delineate between his humid temperate and dry domains for the United States using hydrologic run-off.

- The second was to test the theory represented by the Level II FAO Ecological Zone map by observing whether pixels are classified as predicted using the NDVI time series. The discriminant function analysis provides a quantitative basis and approach for this comparison.

Discriminant function analysis is based on modeling continuous variables, which are assumed to have normal curves within each group. In relation to this requirement, the main assumptions and restrictions of LDA are as follows:

- Group sizes of the dependent variable should not be grossly different. For the FAO Ecological Zones of the United States, there were some zones which were spatially much smaller than others. The FAO's Tropical Wet (Ar) zone on the southern tip of Florida was excluded from the study due to its small size. Prior probabilities based on the proportion of samples within each class were calculated and factored into the discriminant functions (Table 4).

- Predictor variables should follow multivariate normal distributions; that is, each predictor variable has a normal distribution about fixed values of all other independents. In addition, homogeneity of variances (homoscedasticity) and covariance/correlations need to exist between and within predictor variables. Discriminant analysis is highly sensitive to outliers within the predictor variables.

Decision tree analysis was used as an alternative approach for classifying the NDVI values into homogenous units. Decision trees are described as being an alternative to the more traditional methods of Discriminant Analysis, Cluster Analysis, Nonparametric Statistics, and Nonlinear Estimation. However, although the flexibility of decision trees makes them a very attractive analysis option, this should not imply that their use is recommended to the exclusion of more traditional methods. Similar to discriminant analysis, decision tree analysis, also known as classification tree analysis, takes a set of independent variables to predict class membership. The main difference between discrimant analysis and decision trees lies in the sequence used to classify data. Discriminant analysis uses a method that simultaneously considers all variables for making decisions, whereas, the decision tree uses a hierarchical approach by considering only one variable at each stage of the classification hierarchy. Decision trees are often used for devising prediction rules that can rapidly and repeatedly be evaluated for assessing the adequacy of linear models, and for summarizing large multivariate datasets. Therefore, as an alternative approach to classification, decision tree analysis was chosen for two reasons:

- To provide an alternative non-linear exploratory approach for uncovering structure in the data. Interactions and hierarchical relationships between the independent variables (NDVI), and their relative importance for different classes are rendered explicit with trees. (Hansen *et al.*, 1996).

- To assess the adequacy of the linear discriminant model. Classification trees do not assume normality or homogeneity in the data.

A tree is constructed by recursively partitioning a data set into purer, more homogenous subsets of the variables. The method uses a deviance measure, the likelihood ratio statistic, to compare all possible splits of the data to find the one split that maximizes the dissimilarity among the resulting subsets (Hansen *et al.*, 1996). Tree-based methods are often used to classify land cover types, and in the process help reveal any hierarchical and/or non-linear interactions of the variables. The approach of decision tree analysis is to build a classifier expressed as a decision tree or as a set of rules for the purpose of predicting a sample point's class from its attribute values.

The statistical software package, Statistical Analysis System (SAS), developed by James H. Goodnight *et al.*, at SAS Institute Inc. in North Carolina, USA, was used to perform the discriminant analysis on the sample. A classification program called C5.0, developed by Ross Quinlan *et al.*, at RuleQuest Research Pty Ltd., in Australia, was used to perform the decision tree analysis. The discriminant functions and the decision tree derived from the training (sample) data were applied to the entire SDA-reduced NDVI data set to derive two *a posteriori* classified maps of the United States (see Appendix C and D). These maps were then compared, using matrix overlays and tables, to the *a priori* FAO Ecological Zone map (Figure 5).

CHAPTER 6

6) Results and Discussion

6.1) A Compiled Ecological Zone Map of North America

By adapting the CEC's ecological classification scheme in Canada, and Bailey's Ecoregions in the United States, a compiled map of FAO Ecological Zones for North America was produced. The reclassification schemes given in Tables 3 and 4 were developed by qualitatively comparing each description of the ecological zones from both data sources to the Global FAO Ecological Zone Level III Classification Scheme given in Table 1. The scheme provides three levels of detail for representing functional ecological units in a coherent and consistent way (level III, the most detailed, is shown in the tables). Figure 4 is a map of the translated classes at level III. During the development of the compiled ecological zone map, some edge-matching and conceptual problems were encountered that needed to be resolved.

Figure 6: Draft FAO Ecological Zones of North America

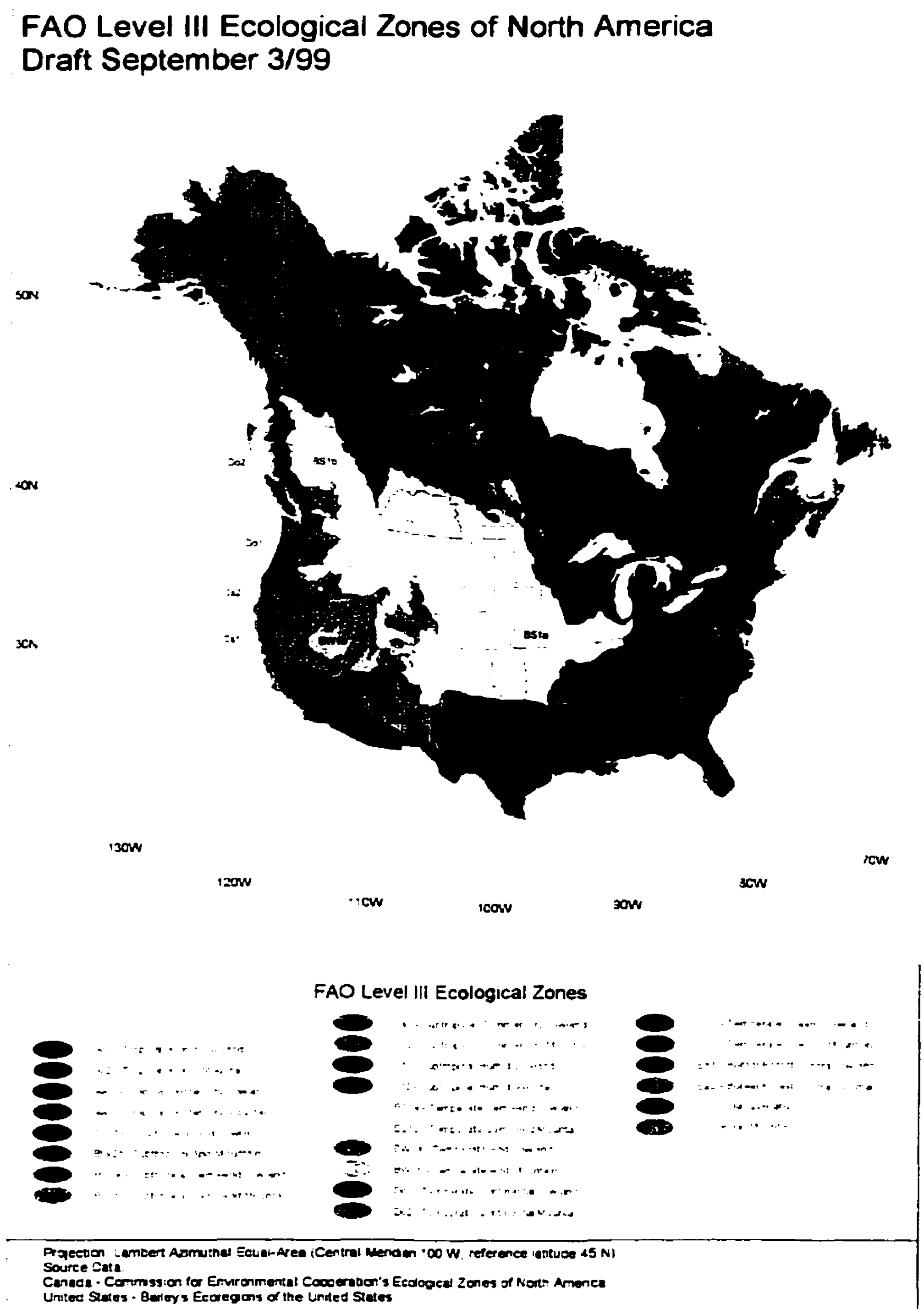


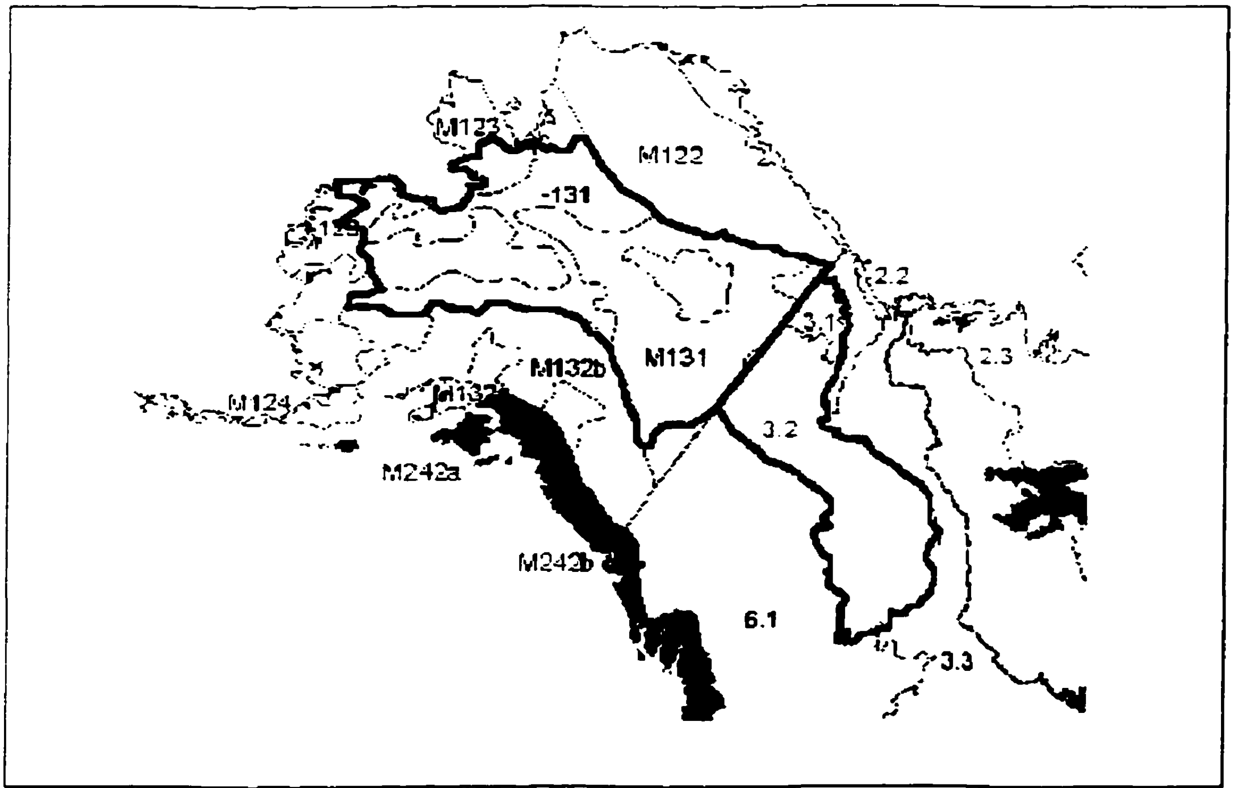
Table 3

Temperate Semi-Arid Lowland	BS1a	8 2	Central USA Plains
		9 2	Temperate Prairies
		9 3	West-Central Semi-Arid Prairies
Temperate Semi-Arid Mountai	BS1b	6 2	Western Cordillera
Temperate Arid Lowland	BW1a	10 1	Western Interior Basins and Ranges
Temperate Continental Lowlan	Dc1	5 2	Mixedwood Shield
		8 1	Mixedwood Plains
Temperate Continental Lowlan	Dc2	5 3	Atlantic Highlands
Temperate Oceanic Mountain	Do2	7 1	Marine West Coast Forest
Boreal Lowland	E1	3 1	Alaska Boreal Interior
		3 3	Taiga Plain
		3 4	Taiga Shield
		4 1	Hudson Plain
		5 1	Softwood Shield
		9 1	Boreal Plain
Boreal Mountain	E2	3 2	Taiga Cordillera
		6 1	Boreal Cordillera
Polar Lowland	F1	2 1	Northern Arctic
		2 2	Alaska Tundra
		2 4	Southern Arctic
Polar Mountain	F2	1 1	Arctic Cordillera
		2 3	Brooks Range Tundra

Table 4

Tropical Humid	Ar1	410	Savanna Division
Subtropical Arid Lowland	BW2a	320	Tropical/Subtropical Desert Division
Subtropical Semi-Arid Lowland	BS2a	250	Prairie Division (southern midwestern states)
		310	Tropical/Subtropical Steppe Division
Subtropical Semi-Arid Mountain	BS2b	M310	Tropical/Subtropical Steppe Mountains
Subtropical Summer Dry Lowland	Cs1	260	Mediterranean Division
Subtropical Summer Dry Mountain	Cs2	M260	Mediterranean Mountains
Subtropical Humid Lowland	Cf1	230	Subtropical Division
Subtropical Humid Mountain	Cf2	M230	Subtropical Mountains
Temperate Arid Lowland	BW1a	340	Temperate Desert Division
Temperate Arid Mountain	BW1b	M340	Temperate Desert Mountains
Temperate Semi-Arid Lowland	BS1a	250	Prairie Division (north midwestern states)
		330	Temperate Steppe Division
Temperate Semi-Arid Mountain	BS1b	M330	Temperate Steppe Mountains
Temperate Continental Lowland	Dc1	220	Hot Continental Division
		210	Warm Continental Division
Temperate Continental Mountain	Dc2	M220	Hot Continental Mountains
		M210	Warm Continental Mountains
Temperate Oceanic Lowland	Do1	240	Marine Division
Temperate Oceanic Mountain	Do2	M240	Marine Mountains
Boreal Lowland	E1	130	Subarctic Division
Boreal Mountain	E2	M130	Subarctic Mountains
Polar Lowland	F1	120	Tundra Division
Polar Mountain	F2	M120	Tundra Mountains

The first challenge that arose was related to the difference in definition and interpretation of Polar and Boreal Zones in the source maps. The international border between Alaska and the Yukon highlighted a difference in location set for the northern extent of the FAO Boreal Zone in the initial reclassification. In the Yukon, CEC's 3.2 - Taiga Cordillera and 3.1 - Alaska Boreal Interior (subgroups of Class 3 - Taiga) were originally reclassed as being within the Polar Zone. These CEC classes are described as having 'numerous lakes, bogs, other wetlands and forests interwoven with open shrublands and sedge meadow' (CEC, 1997). These characteristics were initially interpreted as being more typical of polar tundra under the FAO scheme. Under Bailey's scheme, adjacent areas in Alaska classified as M131 -Yukon Intermontane Plateau Tayga-Meadow Province and 131-Yukon Intermontane Plateaus Tayga Province (subgroups of M130 - Subarctic Mountain Division and 130 Subarctic Division respectively) were regrouped into the FAO Boreal Class. The descriptions for Bailey's provinces suggested that boreal forests of stunted black spruce dominate the landscape. With this initial reclassification scheme, it appeared that the southern extent of the FAO Polar Zone in the Yukon was much further south than the extent set in Alaska with an abrupt change between FAO Polar and Boreal Zones evident at the international boundary. Although Bailey and the CEC were essentially describing the same entity, their use of language and its meaning needed to be further examined to conclude that the CEC's 3.1 and 3.2 classes should be reclassed to FAO's Boreal, to more closely match adjacent boundaries in Alaska (see Figures 5 and 6). This discrepancy was due to the interpretations of the textual description of the philosophies and theoretical frameworks. For instance, delineating what is considered 'tundra' vs. 'taiga' or sparsely forested vs. forested regardless of the fact that climatic factors were similar in both systems. For reclassification purposes, 'sparsely forested areas' in the CEC's system were considered within the FAO Boreal Zone class definition.



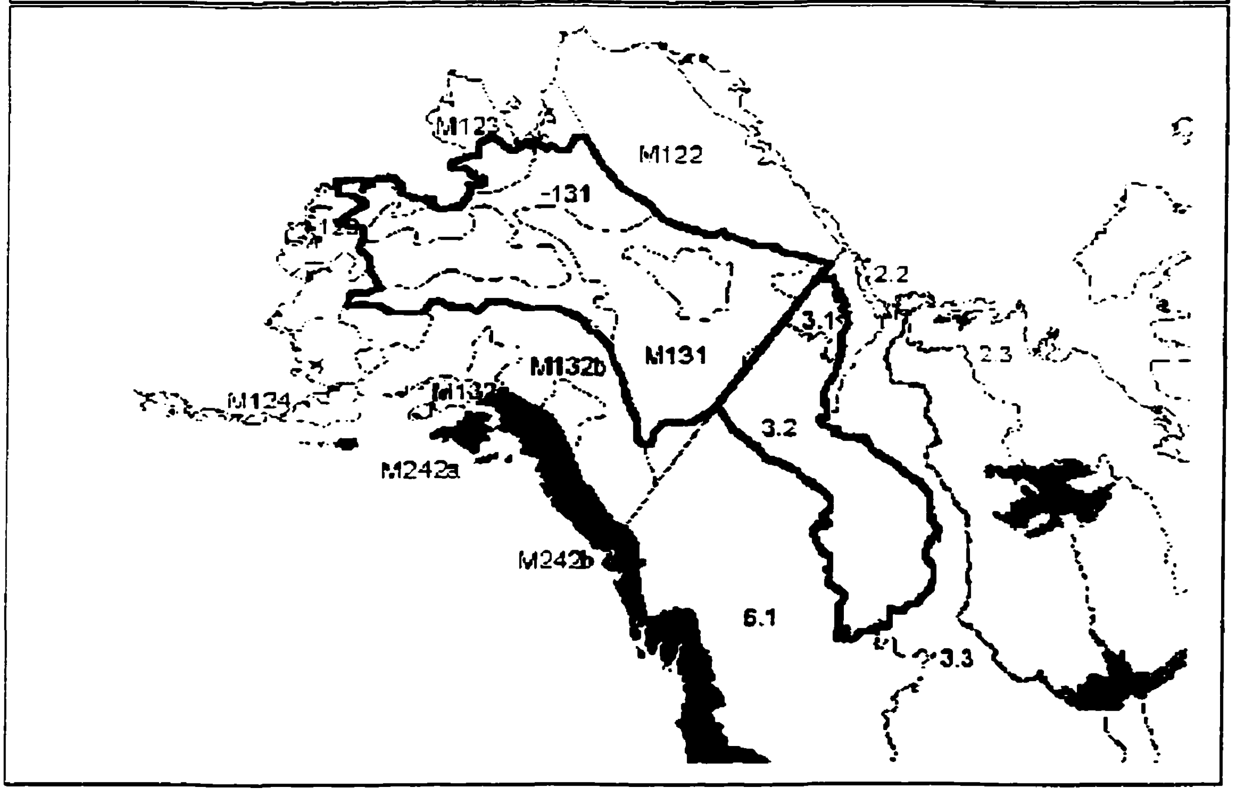


Figure 5 and 6: Alaska and the Yukon Ecoregions. In the Yukon the 3.2 and 3.1 Taiga Cordillera classes were originally regrouped into the FAO Polar Class. Upon further comparison with adjacent areas in Alaska, which fall within Bailey's M131 and 131 Yukon Intermontane Plateaus Tayga Provinces, it was decided that the FAO Boreal Class would be more appropriate for this area of the Yukon.

In addition to the Alaska/Yukon discrepancies, a transitional area that posed another challenge was along the border between Manitoba, Ontario, Quebec, New England, Northern Minnesota and the northern Appalachians. In the U.S., these areas belong to Bailey's 211 – 'Laurentian Mixed Forest Province' and M211 'Adirondack-New England Mixed Forest-Coniferous Forest-Alpine Meadow Province' (subclasses of 210 – 'Warm Continental Division and M210 – 'Warm Continental Division - Mountain Provinces' respectively). These ecoregions were reclassed into FAO's Temperate Continental Zone (Dc) in accordance with Köppen's contingency table. Directly across the border, however, the entire CEC's class 5 – 'Northern Forest' was initially reclassed to the FAO's Boreal class. Over 80 percent of this area is covered by forest, with a gradual transition of mostly coniferous species in the North to mostly deciduous in the South. It was the delineation of the boundary between Boreal and Temperate Continental Zones that needed to be re-examined. With the initial re-classification scheme, the boundary delineation between these two zones was, in most places, the international border; this initial result highlighted incongruency in the way the two data sources were translated into the FAO Scheme. To deal with this apparent incongruency, the approach was to divide the subclasses of '5 - Northern Forest' into either Boreal or Temperate Continental FAO Classes based on a more detailed analysis of species composition and climate regime descriptions. The northern class 5.1 – 'Softwood Shield', containing more coniferous species was retained as Boreal, whereas the southern classes 5.2 – 'Mixedwood Shield' and 5.3 – 'Atlantic Highlands', with more deciduous tree species and slightly warmer mean annual, summer and winter temperatures, were regrouped into Temperate Continental. This eliminated the conspicuous ecological boundary that followed the international border (see Figures 7 and 8).

Figure 7 and 8: Eastern Canada and USA Ecoregions. In Canada the 5.2 Softwood Shield and 5.3 Atlantic Highlands classes were originally regrouped into the FAO Boreal Class. Upon further comparison with adjacent areas in the United States, which fall within Bailey's M211 and 211 Mixed Forest Provinces within the Warm Continental Division, it was decided that the FAO Temperate Continental Class would be more appropriate for this area of Canada.

The case study for the United States and Canada illustrates several concerns when compiling an EZ map based on existing data sets developed within different nations. Such issues include edge-matching line work between data set boundaries, comparing differences in climatic indicators used, and resolving philosophical and terminology issues regarding the definition and translation of classes. In addition to these issues, the discrepancies encountered in the two geographical areas described above also relate to the source data classification hierarchies, and the level at which the initial reclassification was conducted. Bailey's 2nd level (Divisions) and the CEC's 1st level of classification definitions were used for the initial reclassification scheme. Further investigations of Bailey's 3rd level (Provinces) and the CEC's 2nd level of classification was needed to resolve edge-matching issues between the two data sources.

6.2) Statistical Multivariate Pattern Recognition of NDVI Composites of the Conterminous United States

6.2.1) The Stratified Random Sample

One assumption violation which was identified after processing the data, and which affected the confidence and results of the discriminant and decision tree analyses, relates to the variable sizes of ecological zones and the design of the stratified random sample. Although the option in SAS to include *a priori* probabilities of each class was used in the analysis, the design of the stratified random sample proved to be the greatest limitation when confidence was assessed. Using the 984-point sample generated, seven of the nine classes had corresponding powers greater than 0.80. The two classes with corresponding powers less than 0.80 were also the two smallest ecological zones: Do – Temperate Oceanic with only 19 samples and a power of 0.25, and Cs – Subtropical Summer Dry with a sample of 35 and a power of 0.60 (see Table 5 and Figure 9). One way to address this concern is to set the minimum number of samples within each class to 50.

Table 5: Breakdown of the 984-point stratified random sample used as training data for the linear discriminant and decision tree analyses

Breakdown of the 984 sample points derived from a stratified random sample generated:			
[redacted]			
[redacted]	19	0.019	0.019
	208	0.211	0.211
	301	0.306	0.306
	109	0.111	0.111
	129	0.131	0.131
	35	0.036	0.036
	130	0.132	0.132
	52	0.053	0.053
	1	0.001	0.001
Total	**984**	**1.000**	**1.000**

Location of stratified random sample used for training the discriminant functions and decision tree

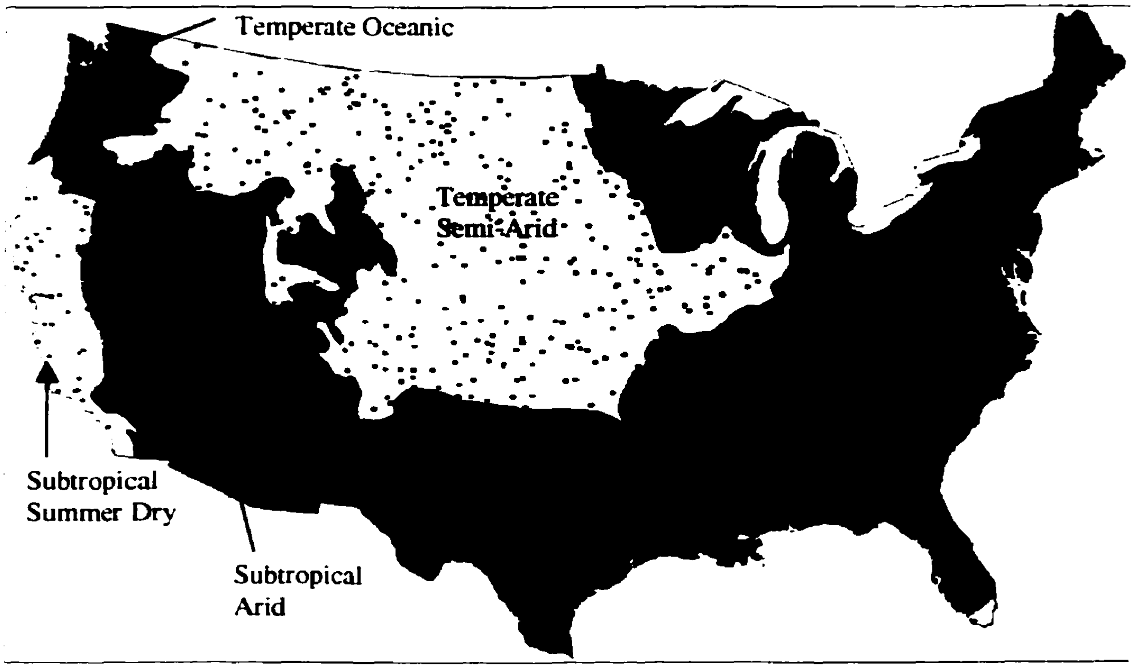


Figure 9: The 984-point stratified random sample by ecological zone

6.4.2) Stepwise Discriminant Analysis (SDA) Results

SDA was performed on the 984-point sample data to identify which of the 26 NDVI bands best characterize the spatial and temporal variance within all the images. In the SDA conducted on the NDVI composites, the images representing averaged smoothed NDVI for July 16-29, and for February 26 to March 11 explained 42.7% and 26.5% of the variance respectively (Table 6). Interestingly, the first four images correspond to periods of time during which spatially variable seasonal changes in NDVI values occur most; with summer and spring images being the two most important images followed by images representing autumn and winter. Eleven of the initial 26 composites describe 97.16% of the variance, with a calculated significance of $p = 0.0001$ for the F-statistic. These eleven bands were subsequently used in the LDA and decision tree analyses (Figure 10).

As the SDA procedure progressed, both the *F*-statistic and the Wilks' Lambda values decreased. This trend indicates that, as more variables were added to the model, total and between-group variances increased and pooled within-group variance decreased. The Wilks' Lambda for these eleven images was 0.0568, with a corresponding *F*-statistic of 5.3590, suggesting a strong relationship between the NDVI values (predictors) included in the model and at least some the FAO Ecological Zones (responses).

Stepwise Discriminant Analysis Results

In order of total amount of variance explained by each predictor (i.e. biweekly smoothed 10-year averaged NDVI composite)

1			1	0.6442	220.6430	42.7126	42.7126	0.0001	0.3568	0.0001
2			2	0.5296	137.0660	26.5336	69.2462	0.0001	0.1674	0.0001
3			3	0.2927	50.3400	9.7449	78.9911	0.0001	0.1184	0.0001
4			4	0.1369	19.2680	3.7299	82.7210	0.0001	0.1022	0.0001
5			5	0.1021	13.7990	2.6712	85.3923	0.0001	0.0917	0.0001
8			8	0.1007	13.5450	2.6221	88.0143	0.0001	0.0677	0.0001
7			7	0.0996	13.4050	2.5950	90.6093	0.0001	0.0752	0.0001
6			6	0.0894	11.3990	2.3034	92.9128	0.0001	0.0835	0.0001
9			9	0.0849	11.2080	2.1697	95.0824	0.0001	0.0619	0.0001
10			10	0.0425	5.3620	1.0380	96.1204	0.0001	0.0593	0.0001
11			11	0.0425	5.3590	1.0374	97.1578	0.0001	0.0568	0.0001
12			12	0.0234	2.8910	0.5596	97.7175	0.0035	0.0554	0.0001
13			13	0.0228	2.8080	0.5436	98.2510	0.0044	0.0542	0.0001
14			14	0.0177	2.1700	0.4201	98.6811	0.0276	0.0532	0.0001
17			15	0.0168	2.0580	0.3984	99.0795	0.0373	0.0520	0.0001
18			16	0.0153	1.8610	0.3603	99.4398	0.0628	0.0512	0.0001
15			15	0.0148	1.8010	0.3486	99.7384	0.0732	0.0524	0.0001
16			14	0.0090	1.0930	0.2116	100.0000	0.3657	0.0529	0.0001

Table 6: Stepwise Discriminant analysis results: Total amount of variance explained by each predictor variable

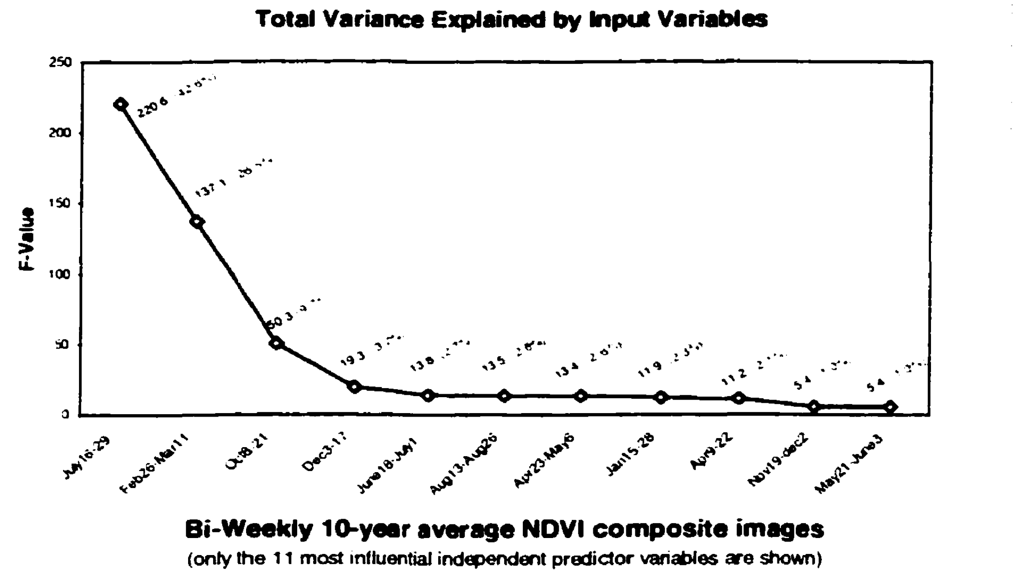


Figure 10: Graph of Total Variance explained by predictor variables

6.2.3) Linear Discriminant Analysis Results

Using the variables defined by the stepwise process and the 984-point sample, LDA was used to build discriminant functions for each of the ecological zones. Using a least-squares approach and generalized (Mahalanobis) distances between ecological zone means, a set of coefficients and constants were developed that define single linear combinations of variables that best differentiate each class. Table 7 is a matrix of generalized distances between ecological zone means. By showing generalized distance, the matrix reveals the relative relationships (distance) between the ecological zones. For example, the table reveals that average values of sample pixels classed as temperate continental are more similar, in the NDVI time-series, to those classed as temperate semi-arid ($D^2 = 9.77$), than to those classed as subtropical arid ($D^2 = 28.51$).

Table 8 gives a list of the coefficients and constants used to calculate the discriminant scores for each pixel and class. The discriminant score for each pixel in the study area is computed as a composite of each of the eleven NDVI images. weighted by the respective discriminant function coefficient. The larger the coefficient. the greater is the contribution of the respective variable to discriminate between groups. For example. the February 26 to March 11 composite image of averaged. smoothed NDVI is more important for determining if pixels are classed Subtropical Summer Dry (c2 = 2.187) than for determining whether they are Temperate Continental (c2 = 1.258). In other words. the NDVI values within the Subtropical Summer Dry Zone are more unique. during this time period. than for what is occurring within the Temperate Continental Zone. This is consistent with the fact that Spring. and hence 'green-up'. starts earlier in the subtropical summer dry zone than in the temperate continental zone. NDVI values for this time period in the temperate continental zone are not greatly different from values of other winter months.

Table 7: Matrix of Mahalanobis (D^2) values for comparison between the 9 predicted and *a priori* ecological zones*

Matrix of Mahalanobis D^2 values (generalized distances) for comparison between the 9 predicted and a priori ecological zones.
The corresponding F values are significant at the p =.0001.

	Predicted Ecological Zones							
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
A priori Ecological Zones								
[illegible]	7.89	11.18	10.67	22.65	11.12	12.86	16.81	27.56
[illegible]	15.96	3.11	9.03	22.45	16.97	27.20	18.63	31.28
[illegible]	16.20	9.77	2.37	9.80	17.04	19.76	12.62	18.02
[illegible]	26.15	21.16	7.77	4.40	22.84	24.17	8.56	9.11
[illegible]	14.96	16.01	15.34	23.17	4.06	17.38	14.02	25.66
[illegible]	14.08	23.63	15.45	21.90	14.77	6.67	16.80	25.01
[illegible]	20.65	17.69	10.94	8.91	14.04	19.42	4.05	9.85
[illegible]	29.58	28.51	14.50	7.63	23.85	25.80	8.02	5.88

* *the smaller the generalized distance the more similar the classes are to each other*

Table 7: Linear Discriminant Constants and Coefficients

Linear Discriminant Function Constants and Coefficients

Discriminant scores for each function and pixel (case) are derived from applying the following equation

DF Score = (CONSTANT) + c1(var1) + c2(var2) + c3(var3) + c4(var4) + c5(var5) + c6(var6) + c7(var7) c8(var8) + c9(var9) + c10(var10) + c11(var11)

		Df3	Df4	Df5	Df6	Df7	Df8	Df9	Df10
	CONSTANT	-303.734	-282.419	-260.511	-229.083	-310.299	-305.286	-255.308	-229.209
Coefficients of:	[redacted]	0.306	1.051	0.719	0.691	1.250	0.240	0.974	0.759
		1.924	1.258	1.496	1.205	1.554	2.187	1.077	1.531
		1.063	0.983	0.896	1.143	-0.526	1.258	1.350	0.415
		-1.920	-1.871	-1.994	-1.982	0.143	-1.823	-1.971	-1.146
		-0.022	0.213	0.361	0.268	-0.576	-0.206	0.197	-0.201
		1.564	1.551	1.262	1.124	1.425	1.521	0.985	1.238
		-0.951	-0.865	-0.623	-0.504	-0.666	-0.702	-0.157	-0.639
		1.377	1.616	1.691	1.383	1.607	1.683	0.943	1.285
		-1.035	-0.044	-0.721	-0.568	-0.499	-0.738	0.196	-0.218
		-1.620	-2.943	-2.614	-2.545	-2.375	-2.620	-3.029	-2.666
		2.924	3.473	3.830	3.832	3.329	3.938	3.902	3.689

Note: Df3 corresponds to the discriminant function for Temperate Oceanic. Df4 = Temperate Continental. Df5 = Temperate Semi-Arid. Df6 = Temperate Arid. Df7 = Subtropical Humid. Df8 = Subtropical Summer Dry. Df9 = Subtropical Semi-Arid. Df10 = Subtropical Arid.

The confusion matrix in Table 9 shows how the **training data** for the discriminant analysis were classified using the discriminant functions. The overall correlation between the predicted classes and *a priori* FAO Ecological Zone classes was 73.2%. When the same discriminant functions were applied to the entire study area, the confusion matrix revealed a correlation of 70.7% (Table 10). The largest difference between the training data prediction results and the study area extrapolation occurred in the Temperate Oceanic Zone, the class with the least confidence due to its small sample size. The discriminant functions accurately predicted 36.8% of the training data (classed Temperate Oceanic) to be Temperate Oceanic. When the discriminant functions were applied to the entire study area 42.7 % of the Temperate Oceanic pixels were classified correctly. This increase in correctly classified pixels for the entire study area in contrast to the training data may suggest the effects of the linear function to partially account for the random variation in the design set. For the remaining classes, the study area prediction results were approximately 5% to 10% less than the training data results.

Table 9: Resubstitution summary on the training data using the derived linear discriminant functions

Resubstitution Summary using the Linear Discriminant Functions									
Number of Observations and Proportion *(by row)* **classified into each Ecological Zone:**									
(training sample accuracy only)									
Derived From FAO Ecological Zones									
	7	2	4	0	3	2	0	1	19
	0.3684	0.105	0.21	0	0.158	0.105	0	0.526	1
	2	**173**	19	0	14	0	0	0	208
	0.009	**0.832**	0.091	0	0.067	0	0	0	1
	6	31	**225**	21	4	1	13	0	301
	0.002	0.01	**0.745**	0.007	0.013	0.003	0.042	0	1
	0	2	16	**72**	1	0	8	10	109
	0	0.018	0.146	**0.66**	0.009	0	0.073	0.091	1
	5	7	2	0	**104**	7	3	0	129
	0.039	0.054	0.015	0	**0.806**	0.054	0.023	0	1
	1	0	3	2	5	**22**	2	0	35
	0.029	0	0.086	0.057	0.146	**0.623**	0.051	0	1
	0	3	10	11	8	0	**82**	15	130
	0	0.023	0.077	0.085	0.062	0	**0.631**	0.115	1
	0	0	0	6	2	0	9	**35**	52
	0	0	0	0.115	0.039	0	0.173	**0.673**	1
	21	**218**	**279**	**112**	**141**	**32**	**117**	**61**	**984**
								Total Training data accuracy	**0.732**

Table 10: Resubstitution summary on the entire study area (conterminous United States) using the derived linear discriminant functions

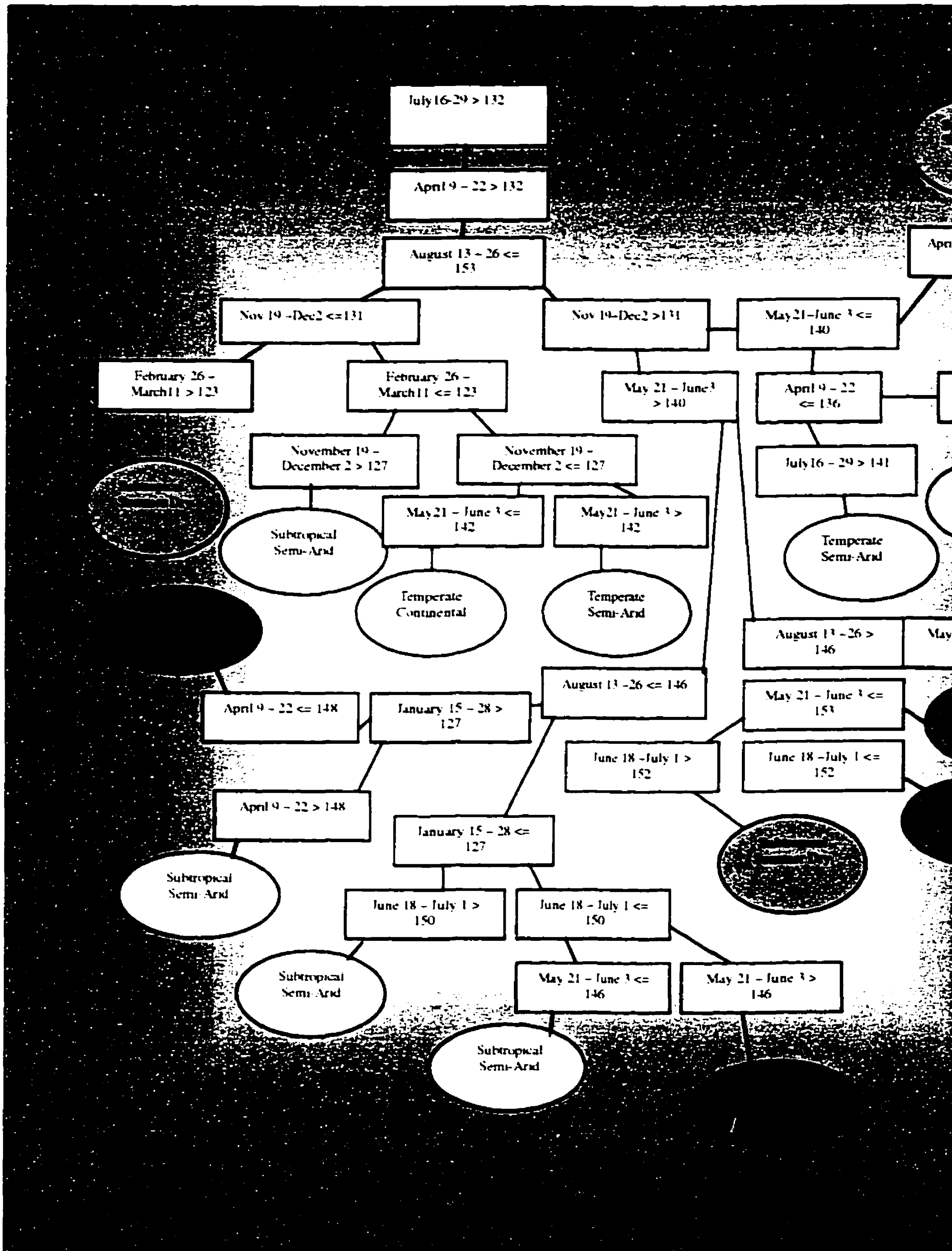
Error Matrix for Linear Discriminant Analysis									
	Linear Discriminant Predicted Group Membership								
FAO Ecological Zone	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Grand Total**
	62891	10379	21253	1025	17374	23157	9261	1729	147069
	42.76%	7.06%	14.45%	0.70%	11.81%	15.75%	6.30%	1.18%	1.94%
	11615	1240460	150257	2759	199523	295	11468	695	1617072
	0.72%	**76.71%**	9.29%	0.17%	12.34%	0.02%	0.71%	0.04%	21.34%
	50899	219511	1741577	139734	51760	7676	98853	15587	2325597
	2.19%	9.44%	**74.89%**	6.01%	2.23%	0.33%	4.25%	0.67%	30.69%
	2433	12697	155080	540992	4801	1899	52132	70867	840901
	0.29%	1.51%	18.44%	**64.33%**	0.57%	0.23%	6.20%	8.43%	11.10%
	31461	44489	14563	2140	819159	27870	36966	889	977537
	3.22%	4.55%	1.49%	0.22%	**83.80%**	2.85%	3.78%	0.09%	12.90%
	8227	1413	24706	8319	47476	155179	14247	2494	262061
	3.14%	0.54%	9.43%	3.17%	18.12%	**59.21%**	5.44%	0.95%	3.46%
	434	20860	77872	64224	119981	28690	548643	144328	1005032
	0.04%	2.08%	7.75%	6.39%	11.94%	2.85%	**54.59%**	14.36%	13.26%
	76	59	2678	29154	2948	9128	107650	250158	401851
	0.02%	0.01%	0.67%	7.25%	0.73%	2.27%	26.79%	**62.25%**	5.30%
Grand Total	168036	1549868	2187986	788347	1263022	253894	879220	486747	7577120
	2.22%	20.45%	28.88%	10.40%	16.67%	3.35%	11.60%	6.42%	**70.73%**

6.2.4) Decision Tree Analysis Results

A decision tree was generated using the same 984-point stratified random sample and 11 bi-weekly averaged, smoothed NDVI images that were used to train the discriminant functions. Using C5.0, developed by RuleQuest Research Inc., a tree with 117 terminal nodes was built with an error rate of 0.09 (or 91% of the training data were predicted to be the same class as the *a priori* class by the decision tree) (see Appendix B). In many respects, the logic of the tree is intuitively more straightforward than that of the linear discriminant functions. For illustrating how to interpret the results, Figure 11 shows approximately one sixth of the resulting tree as translated into a flowchart. Table 11 shows the confusion matrix of the training data when the decision tree was applied on them.

The decision tree method was more successful at classifying the training data because the tree 'over-fit' the data. In other words, since no thresholds were defined as to how far the tree should grow, small groupings of the independent variable were created, which merely attempt to model random variation in the design set. This resulted in a much lower error rate for the training data than for the classification of the entire study area. With this overfitting, when the same decision tree was applied to the entire study area, a much higher error rate occurred. Instead of the 91% correlation between predicted and *a priori* classes, only 67.2% of the pixels were classified the same (Table 12). Again, as with the discriminant analysis, the least correlated class was the Temperate Oceanic Zone. For the training data, 57.9% of the training data classed Temperate Oceanic were predicted correctly using the decision tree. When the decision tree was applied to the entire study area only 42.9 % of the Temperate Oceanic pixels were classified into the same *a prior* class. For the remaining classes, the study area prediction results were at least 15 to 20 % less correlated than the training data predictions. Pruning of the tree, which was not performed in this study, could have addressed the problem of 'overfit'. Pruning would have considered the predicted error rate on new cases and limited the amount of times the decision tree branched. This process would have reduced the amount of modeling performed to account for mere random variations of the design set.

Figure 8: Partial flowchart of the Decision Tree trained from the 984-point stratified sampl



Note: The rectangles in the flowchart represent classification criteria. The colored circles represent the predicted ecological zo classification criteria leading up to it are true. The flowcharts begins with a rectangle containing the text "July16-29 > 132". T classification criteria, which states that, in order to proceed down the decision tree branch, a pixel must have an NDVI value g image representing the July 16 to 29. If the pixel meets this requirement and the statement is true, then proceed to the next recta NDVI values for April 9 to 22 are greater than 132 and so on. This process is continued until a leaf (a.k.a. ecological zone) on

Table 11: Resubstitution Summary on the training data using Decision Tree analysis

Resubstitution Summary using the Decision Tree
Number of Observations and Proportion *(of row)* **classified into each Ecological Zone:**
(training sample accuracy only)

Derived From FAO Ecological Zones									
	11	0	4	0	2	2	0	0	**19**
	0.579	0.000	0.211	0.000	0.105	0.105	0.000	0.000	**1.000**
	0	**197**	4	0	7	0	0	0	**208**
	0.000	**0.947**	0.019	0.000	0.034	0.000	0.000	0.000	**1.000**
	1	9	**278**	5	2	1	4	1	**301**
	0.003	0.030	**0.924**	0.017	0.007	0.003	0.013	0.003	**1.000**
	1	0	9	**94**	1	2	1	1	**109**
	0.009	0.000	0.083	**0.862**	0.009	0.018	0.009	0.009	**1.000**
	1	1	0	0	**125**	0	2	0	**129**
	0.008	0.008	0.000	0.000	**0.969**	0.000	0.016	0.000	**1.000**
	1	0	1	1	2	**30**	0	0	**35**
	0.029	0.000	0.029	0.029	0.057	**0.857**	0.000	0.000	**1.000**
	0	2	5	2	4	0	**116**	1	**130**
	0.000	0.015	0.038	0.015	0.031	0.000	**0.892**	0.008	**1.000**
	0	0	0	3	0	1	4	**44**	**52**
	0.000	0.000	0.000	0.058	0.000	0.019	0.077	**0.846**	**1.000**
	15	**209**	**301**	**105**	**143**	**36**	**127**	**47**	**984**
								Total Training data accuracy	**0.913**

Table 12: Resubstitution Summary of the entire study area (conterminous United States) using Decision Tree analysis

Error Matrix for Decision Tree Analysis

	Decision Tree Predicted Group Membership								
FAO Ecological Zone	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Grand Total**
	63264	14571	30074	4043	9964	17527	4652	2961	147156
	42.99%	9.90%	20.44%	2.75%	6.77%	11.98%	3.16%	2.01%	1.93%
	27908	**1152746**	209768	4109	197278	12297	12423	566	1617095
	1.73%	**71.28%**	12.97%	0.25%	12.20%	0.76%	0.77%	0.04%	21.25%
	26449	240447	**1761572**	99300	48424	42587	90330	17897	2327006
	1.14%	10.33%	**75.70%**	4.27%	2.08%	1.83%	3.88%	0.77%	30.59%
	11513	19965	202295	**512010**	5515	7594	70415	11666	840973
	1.37%	2.37%	24.05%	**60.88%**	0.66%	0.90%	8.37%	1.39%	11.05%
	9729	56948	35300	3491	**760686**	45713	73037	7345	992249
	0.98%	5.74%	3.56%	0.35%	**76.66%**	4.61%	7.36%	0.74%	13.04%
	20962	9360	27889	15759	40974	**132134**	19490	3464	270032
	7.76%	3.47%	10.33%	5.84%	15.17%	**48.93%**	7.22%	1.28%	3.55%
	1745	6493	119155	88326	122873	66008	**529684**	75575	1009859
	0.17%	0.64%	11.80%	8.75%	12.17%	6.54%	**52.45%**	7.48%	13.27%
	1088	94	5227	60644	4819	11583	118765	**201707**	403927
	0.27%	0.02%	1.29%	15.01%	1.19%	2.87%	29.40%	**49.94%**	5.31%
Grand Total	162658	1500624	2391280	787682	1190533	335543	918796	321181	7508297
	2.14%	19.72%	31.43%	10.35%	15.65%	4.41%	12.08%	4.22%	**67.21%**

6.5) Assessment of the Compiled North American Ecological Zone map based on Linear Discriminant and Decision Tree Analysis Results

Even though the decision tree over-fit the data, the overall confusion matrix, which compares the resulting classification to the FAO classes, is similar to that of the linear discriminant function analysis results (70.73% correlation vs. 67.21%). Both appear to be less correlated with the FAO classes in mountainous regions. This effect can be seen along the eastern margins of both the temperate oceanic and the subtropical summer dry classes which follow the Cascade Mountain Range, the Sierra Nevada and the Rocky Mountains of the temperate semi-arid classes (see appendices C to F). This can be attributed to the effects of elevation on vegetation and hence the values of NDVI. At higher elevations and on steep slopes, vegetation may be limited due to local conditions of soil and climate. The result is a poor fit of the NDVI values in mountainous regions both for the decision tree and discriminant function models.

Both approaches reveal comparable dissimilarities to the FAO classes along the 0°C isotherm for the coldest month in which Bailey and Köppen used to delineate between Temperate and Subtropical zones. The *a priori* FAO class boundary between Temperate Semi-Arid and Subtropical Semi-Arid in North Texas/Oklahoma, as well as the boundary between Temperate Continental and Subtropical Humid on the southern side of the Appalachians to the Ozark Plateau appear not to follow the *a posteriori* classification trend. For the Texas/Oklahoma region, it appears that the isotherm is more northerly than what the variance of the NDVI values suggest. Conversely, In the Appalachian /Ozark region it appears that it is more southerly than what the NDVI values suggest. Appendices E and F show where dissimilarities between the *a priori* and *a posteriori* classifications occur. This illustrates that although the isotherm is perhaps an adequate representation of the generalized boundary between temperate and subtropical zones, regional variations do appear to exist. This may perhaps illustrate the limits of scale for the Global FAO Ecological Zone map. For these areas, a comparison to the USDA publication on World Crop Areas and Climatic Profiles (1994) suggests that climatic variables, such as average dates of last spring freeze or first autumn freeze, may be more influential in characterizing and describing the distribution of vegetation and crops than mean temperature of coldest month alone.

Another similarity between the results of the two approaches is the low correlation of the Temperate Oceanic class (42.99% for the discriminant function, and 42.76% for the decision tree). In addition to the problems outlined in preceding sections regarding sample design and the small number of samples within this class, another reason for the low correlation is related to the land form type. The class is comprised of mountainous regions and complex landforms with a wide range of vegetation from dense temperate rainforest to alpine tundra. Landscape complexity within each ecological zone, and how well the sample reflected this complexity in turn, affected how well the linear discriminant and decision tree analyses were able to predict class membership.

The largest discrepancies between the two approaches were in the subtropical arid and subtropical summer dry classes, where the decision tree was 12% and 10% less successful than the discriminant analysis, respectively. There are two possible explanations for this result. The first is that the linear discriminant function was more successful at dealing with prior probabilities related to the stratified random sample, since, in addition to the temperate oceanic class, these two classes had a smaller proportion of the training data. The second explanation relates to the way in which the approaches deal with within-class variance of the training data. In the *a posteriori* classified map of the discriminant function, it appears that the delimitation of the classes, based on highest discriminant score, are more homogenous and bounded more by changes in NDVI due to elevation than the decision tree results. Only two of the classes (Temperate Semi-Arid and Temperate Oceanic) had higher correlations using the decision tree approach, albeit only slightly (0.23% and 0.81% higher than the results of the discriminant functions respectively).

CHAPTER 7

7) Conclusion

7.1) Summary

The focus of this research was to develop a means for accurately representing ecological zones at the global scale. This research is part of a broader project for developing a spatial database of ecological zones for the UN FAO's FRA 2000 Report. The thesis begins with a brief discussion of ecological classification schemes appropriate for global applications. For map compilation, the Köppen system of climate classification within a hierarchical framework is used as a basis for the FAO ecological classification scheme. An illustration of how existing data are combined and reclassified was made using the United States and Canada as an example. Methods for assessing accuracy of ecoregions maps at the macro scale were addressed with an assessment of the draft FAO Level II Ecological Zone map. This assessment was performed by classifying 10-year average, bi-monthly, smoothed AVHRR-NDVI composites of the conterminous United States using linear discriminant and decision tree analyses. The results of the linear discriminant analysis were more significantly correlated to the FAO classes, although both approaches suggest that the classification scheme does maximize between-class variance of the NDVI temporal series.

Small-scale vegetation mapping is essential to integrate the various parts of the continents together, and establish a basis for detailed research. The task of developing a global 'Ecological Zone' (EZ) map, is part of an attempt to improve the way in which the FAO provides information on the world's forest resources on a regular basis. New information demands have led the reporting by the FAO to deal more with the 'ecological' context of forest resources. However, the complexity of the 'ecoregion' concept has created a challenging task for geographers and environmental scientists interested in projects related to 'ecoregion' mapping. The first objective of this research was to identify and evaluate the most appropriate type of classification scheme for a global ecological zone map. In support of that objective, it was deemed that the language and usage of the ecoregion concept be carefully considered. The FAO Global Ecological Zone Classification scheme was presented in Chapter 4 in the context of an optimum classification scheme. For a developing a compiled global representation of 'ecoregions,' the classification scheme has the following broad characteristics:

- the scheme is based on structural-physiognomic characteristics of potential natural vegetation combined with one or two climatic variables;
- it has clearly defined limits for each of the categories;
- it has an optimally sufficient number of categories to express variations, without causing confusion;
- and it is systematically hierarchical and *a priori*.

The second objective, to explore an approach for combining source data to achieve a coherent and consistent global FAO Ecological Zone map, was addressed with the United States and Canada as a case study. The short term goal of this component, was to provide a case study for developing a worldwide EZ map and database that is appropriate for reporting on forestry statistics. Using data sets developed under the direction of Robert Bailey (USDA Forest Service) for the United States, and the Commission for Environmental Cooperation in Canada, a compiled map of FAO Ecological Zones, directly related to Köppen's Climate classes, was developed. The compilation process used Bailey's ecoregions, with their direct references to Köppen's, as a basis for subsequently reclassifying the CEC's classes. For translation to the Kotka III compliant FAO Ecological classification system, the approach places importance on having relevant documentation of existing classification schemes. Through qualitative comparisons of class descriptions, the CEC's map was incorporated into a globally oriented North American Ecological Zone map. This macro perspective promotes the use of regionally developed ecoregion boundaries for global modeling, while also providing a global perspective to countries that utilize the dataset for regional projects.

An assessment of accuracy and error is essential for building confidence in the methods and results of a study. With this knowledge, users can evaluate and judge the most appropriate application of the methods and results. To fulfill the third objective of the research, an accuracy assessment was performed to quantify how well the FAO Ecological Zone map of the United States corresponds to temporal and spatial changes in NDVI values. The main assumption, for this approach, was that NDVI provides an adequate indicator of net primary productivity (NPP). The key hypothesis was that the FAO Ecological Zones of the United States are significantly bounded by temporal and spatial changes in NDVI. Using Linear Discriminant and Decision Tree analyses, it was concluded that the FAO Ecological Zone map adequately represents spatial and temporal homogeneity of variance within the NDVI at the macro-scale. There were several objectives that were addressed in validating the FAO Ecological Zone Map:

- The first was to identify and describe the spatial and temporal homogeneity of variance within 26 images of smoothed and averaged bi-monthly NDVI composites for a ten-year interval.
- The second objective was to assess the overall correlation of the *a priori* FAO ecological classification to the *a posteriori* classification of NDVI using linear discriminant and decision tree approaches. This involved both spatial and statistical comparisons between the datasets to show discrepancies and similarities between the representations.
- The third objective of the validation was to provide a methodology that can similarly be applied to other regions of the globe for assessing ecological zones.

The comparison of the linear discriminant and the decision tree analyses proved to be informative for describing the spatial and temporal between- and within-class variance of the NDVI within the context of ecological classification. Based on NDVI as a measure of NPP, and regardless of sample design problems, both the linear discriminant function and the decision tree results suggests that the Level II FAO Ecological Zone classes for the United States are an adequate representation of discrete ecological units at the macro scale. For the approach used here, the linear discriminant function produced a better error rate (70.73%) than the decision tree (67.21%). In addition to the error rates, the ecological zone map developed from the discriminant analysis suggests that the linear functions were able to delineate more successfully, between- and within-class variance due to effects of elevation than the decision tree. However, the similarity of results suggest that both approaches could be used to assess the accuracy of small-scale *a priori* ecological classification schemes using NDVI composites as predictor variables. In addition, these approaches could also be used to help resolve edge-matching issues between national or regional data sets, such as those between Canada and the United States

7.1) Future Research

Explorations into the structure of 'data' are a part of continuing processes of research that scientists in general and data analysts in particular are constantly engaged in. It is their task to separate the signal from the noise, and to characterize the signal based on existing theories or hypotheses of reality (Breiman *et al.*, 1984). This has been an on-going process since the beginnings of science. Two broad areas of technology are instrumental in contributing to that process, and thereby furthering the bounds of our knowledge base. One is computer-aided technology, and the second, satellite technology. It is incomprehensible how theories on the dynamics of global processes could be interpreted and modeled efficiently without the integration of these two technologies.

From the AVHRR sensor orbiting around the earth, the software written to process and analyze the data collected, to the visualization systems to view them, both these technologies have become powerful vehicles through which the structural complexity and dynamics of macro scale ecological processes can be visualized, characterized, and understood. Satellite technology is also a powerful contributor to increasing streams of data that can be processed and analyzed. NASA's Earth Observing System (EOS) is an example of how a vast repository of data and data products are being created with the broad objective of studying global change in all domains of the Earth's ecological system.

In concert with the benefits of these two technologies, data analysis tools such as discriminant analysis and decision trees provide complementary support to make sense of the analyzed data, and aid in deciding on varying hypotheses of reality. This research is a clear illustration of how these two analytical tools can be applied to investigate the validity of ecological classification schemes. The development of both these approaches, in the past 20 years, is related to the ability of technology to enable analysts to sieve through and handle massive amounts of data. The following areas of further research are related to different approaches for applying ecological theory to large amounts of geographic data. The first area of investigation relates to the nature of the input data. Since smoothed NDVI data, as was used in this study, is not available for the entire world, an investigation into applying the same approach on non-smoothed NDVI data would be useful for gaining insight into the effects of the smoothing function. Secondly, for an alternative approach to incorporating annual fluctuations of NDVI, it has been suggested that the integral of temporal NDVI (another possible estimator of NPP), or the area under the curve, could also be used (Goward *et al.*, 1987). In addition to NDVI values, incorporation of elevation, slope information and other data sources as predictor variables would be useful for further investigating the stratification of ecological regions.

Next, problems due to sampling structure could be investigated. These include varying the sample size of the training data, and setting minimum limits on the number of sample points within each class. In addition, the number and scale of *a priori* ecoregions used as the dependent variable could be varied. For example, instead of using Bailey's boundaries for Ecoregion Divisions (2nd level of hierarchy), his boundaries for Ecoregion Provinces could be used (3rd level). As an alternative and for a comparison between theories, Omernik's ecoregions boundaries could be used instead of Bailey's.

Finally, for visualizing the confidence of the *a posteriori* classifications, an explicit approach for devising a confidence map would help in visualizing the confidence of the two approaches. An in-house add-on to C5.0 called "C50mapi" at the EROS Data Center was used to generate a confidence map from the decision tree although it is not presented here due to its 'black box' approach. Further refinement of the decision tree could also be investigated through pruning and 'bootstrapping' to analyze the effects of these tools on the final classified image. For developing a confidence map of the discriminant analysis, such questions as to how to quantify the similarity of discriminant scores need to be addressed (i.e. for each pixel, how similar were the first 2, 3 or 4 scores?).

The primary aim of this research focussed on the development of a methodology for producing a reliable global database of ecological zones. It identified and evaluated appropriate classification schemes, explored combinations and reclassification mechanisms for existing data, and investigated how error and uncertainty contribute to the final quality. This research also illustrated the integrative merits of well-proven statistical and alternative data analysis methods and tools brought to bear on remotely-sensed satellite-derived data using computer-aided data processing and visualization systems. This research has drawn together various areas of scientific knowledge, processes, philosophies, and analyses. From this synthesis, it can be concluded that the FAO Ecological Zone Map adequately exemplifies the temporal and spatial homogeneity of variance within the NDVI.

Bibliography

Adams. J. 1998. "*A Suggestion for an improved vegetation scheme for local and global mapping and monitoring*". http://www.esd.ornl.gov/projects/qen/veg.html

Adams. J. 1996. "Towards a better vegetation scheme for global mapping and monitoring" *Global Ecology and Biogeography Letters* 5: 3-6.

Eastman. J.R. and M. Fulk. 1993. 'Long Sequence Time Series Evaluation Using Standardized Principal Components.' *Photogrammetric Engineering and Remote Sensing*. 59. 6:991-996.

FAO. 1997 "Resource Assessment. Statistics and Outlook Studies" FAO Forestry Information Notes Feb. 1997. (http://www.fao.org/fo%2a/infonote/infnote/allstude.pdf)

FAO. 1997. "*Minutes of the Meeting between representatives of U.S. Federal Data Committee. Earth Cover Working Group and Vegetation Subcommittee and Food and Agriculture Organization of the United Nations.*" http://www.fao.org/waicent/FaoInfo/Agricult/AGL/agls/fgdcfao.htm

Bailey. R.G. 1984. "Testing an Ecosystem Regionalization." *Journal of Environmental Management* 19: 239-248.

Breiman. L. Freidman. J.H.. Olshen. R.A.and C.J. Stone. 1984. "Classification and Regression Trees." Wadsworth Inc.

Brown. D.E.. Lowe. C.H. and C.P. Pase. 1980. "*A Digitized Systematic Classification for Ecosystems with an Illustrated Summary of the Natural Vegetation of North America.*" USDA Forest Service. General Technical Report: RM-73.

Bull. G. 1997. "FAO's Global Fibre Supply Study: context. method and modeling the future." Unasylva. vol.49. no.193. (http://www.fao.org/fo%2a/unasylva/193/E/193E.htm)

Commission on Environmental Cooperation. 1997. "Ecological Regions of North America: Toward a Comparative Perspective". (http://www.cec.org/english/resources/publications/index.cfm)

De Laubenfels. D.J. 1975. *Mapping the World's Vegetation: Regionalization of Formations and Flora*. Kluwer Academic Publishers. Boston.

Di Gregorio. A. 1996. *"FAO Land Cover Classification: A Dichotomous, Modular-Hierarchical Approach."* http://www.fao.org/WAICENT/FAOINFO/SUSTDEV/eIdirect/eIre0019.htm

Gatsonis. C. and A.R. Sampson. 1989. "Multiple Correlation: Exact Power and Sample Size Calculations" *Psychological Bulletin* 106:3:516-524.

Goward. S.N.. Dye. D.. Kerber. A.. and V. Kalb. 1987 "Comparison of North and South American Biomes from AVHRR Observations" *Geocarto International* 1:27-39.

Hand. D.J. 1997. *Construction and Assessment of Classification Rules.* John Wiley & Sons. Toronto.

Hansen. M.. Dubayah. R. and R. Defries. "Classification trees: an alternative to traditional land cover classifiers." *International Journal of Remote Sensing.* Vol.17. 5:1075-1081.

Hardy. C.C.. and Burgan. R.E. 1999. "Evaluation of NDVI for Monitoring Live Moisture in Three Vegetation Types of the Western United States." *Photogrammetric Engineering and Remote Sensing.* Vol.65. 5:603-610.

Hou. Xue-yu. 1983. "Vegetation of China with reference to its geographic distribution." *Annals Missouri Bot. Garden.* 70:509-548.

Hypertext Webster's Dictionary - http://www.fin.gov.nt.ca/webster.htm

Jennrich. R. I. (1977). Stepwise discriminant analysis. In K. Enslein. A. Ralston. & H.S. Wilf (Eds.). *Statistical methods for digital computers.* New York: Wiley.

Küchler. A.W. 1967. *Vegetation Mapping.* Kluwer Academic Publishers. Boston.

Küchler. A.W. and I.S. Zonneveld (editors). 1988. *Vegetation Mapping: Handbook of vegetation science.* Kluwer Academic Publishers. Boston.

Lachenbruch. Peter A. 1975. *Discriminant Analysis.* Hafner Press. New York.

Lanly. J.P. 1997. "World Forest Resources: Situation and Prospects." *Unasylva.* vol. 48. no.190/191. (http://www.fao.org/fo%2a/unasylva/190/E/190E.htm)

Lobo A.. Ibáñez Martí. J.J.. and C. Carrera Giménez-Cassina. 1997. "Regional scale hierarchical classification of temporal series of AVHRR vegetation index." *International Journal of Remote Sensing*. Vol.18. 15: 3167-3193.

Odum. E.P. 1971. Fundamentals of Ecology. 3rd Edition. W.P. Saunders Company.

Persson R.. and K. Janz. 1997. "Assessment and Monitoring of Forest and Tree Resources" *XI World Forestry Congress*. Antalya. Turkey. 13 to 22 October. (http://www.fao.org/fo%2a/wforcong/publi/v1/tle/1%202.htm).

Rowe. J.S. and J.W. Sheard. 1981. "Ecological Land Classification: A Survey Approach" *Environmental Management*. Vol.5. 5:451–464.

Singh. K.D. 1998. "Outline of FAO/EDC/WCMC Paper: Ecofloristic Zone Maps and Statistics for FRA2000" (unpublished paper).

Swets. D.L.. Reed B.C.. Rowland. J.D.. and S.E. Marko. "A Weighted Least-Squares Approach to Temporal NDVI Smoothing" (unpublished paper)

Taylor. D.J. and K.E. Muller. 1995. "Computing Confidence Bounds for Power and Sample Size of the General Linear Univariate Model" *The American Statistician* 49:1:43–47.

United States Department of Agriculture. 1994. "Major World Crop Areas and Climatic Profiles: Agricultural Handbook No. 664." World Agricultural Outlook Board and the Joint Agricultural Weather Facility. Washington D.C.

Walter. H. 1973. *Vegetation of the Earth and Ecological Systems of the Geo-Biosphere* (translation from 5th edition in German. by J. Weiser). Springer-Verlag. Heidelberg.

Wardle. P. 1997. "The Evolution of Forestry Statistics from 1945 to 2000". *Unasylva*. vol.46. no.182 (http://www.fao.org/fo%2a/unasylva/182/185e.htm)

Zhu. Z.. 1998. "Ecoregion Map for FRA2000. June 1998" (unpublished paper).

Zhu. Z.. 1998. "Minutes of Meeting. Washington D.C.. June 7th & 8th" (unpublished paper).

Zhu. Z. 1997. *"Producing a Global Ecoregions Classification"*. (unpublished paper).

Appendix A: Equations used by SAS for the calculating linear discriminant functions

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV^{-1} \, (\bar{X}_i - \bar{X}_j) - 2 \ln PRIOR_j$$

Calculations for constants and coefficients for each linear discriminant function

$$Constant = -.5 \, \bar{X}_j' \, COV^{-1} \, \bar{X}_j + \ln PRIOR_j \qquad Coefficient\ Vector = COV^{-1} \, \bar{X}_j$$

Generalized Squared Distance Function of observations to class means:

$$D_j^2(X) = (X - \bar{X}_j)' \, COV^{-1} \, (X - \bar{X}_j) - 2 \ln PRIOR_j$$

Posterior Probability of Membership in each ecological zone:

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \,/\, SUM_k \exp(-.5 \, D_k^2(X))$$

where: D^2 = generalized square distance (Mahalanobis distance)
PRIOR = prior probability of class (based on proportion of samples within each class)
COV = within-group covariance
i,j = subscripts to distinguish between ecological zones
X = p-dimensional vector containing the quantitative variables (NDVI) of an observation (pixel)

$\bar{X}$ = a p-dimensional vector containing the variable means (of NDVI) within each ecological zone

Appendix B: Decision Tree trained from 984-point stratified random sample using C5.0

The first part identifies the version of C5.0. the run date. and the options with which the system was invoked. C5.0 constructs a decision tree from the 984 training cases in the file sampled.data. The tree itself can be paraphrased as:

"If NDVI value for the bi-weekly average. smoothed composite for July 16 - 29 is less than or equal to 132. and if January 15 – 28 is less than 108 and if June 18 – July 1 is less than 108 and if February 26 to March 11 is greater than 106. then the pixel is Subtropical Arid……"

Every leaf of the decision tree is followed by (n) or (n/m). Where the value of n is the number of cases in the file that are mapped to this leaf. and m is the number of them classified incorrectly by the leaf.

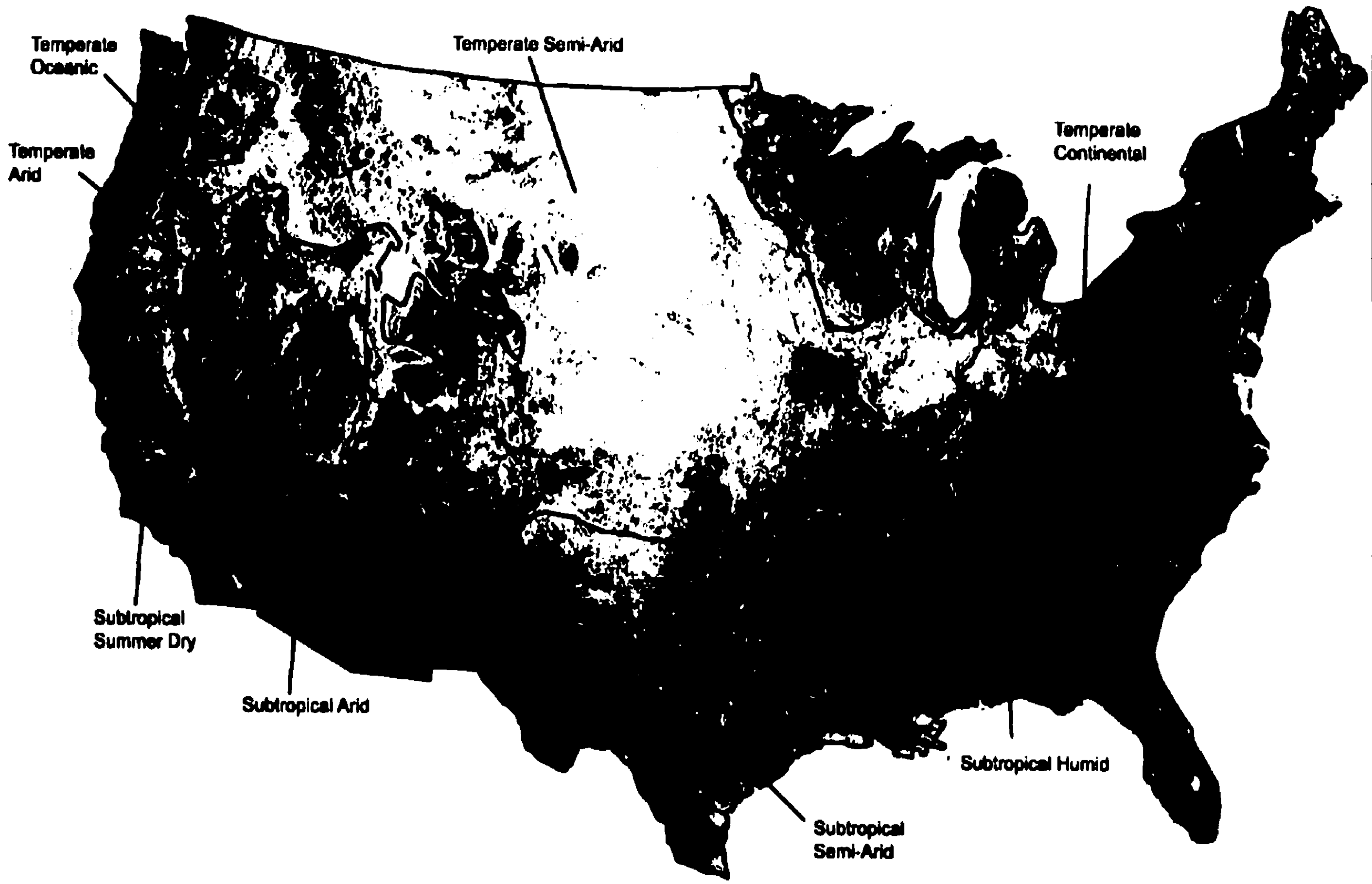
```
C5.0 INDUCTION SYSTEM [Release 1.09]  Sat Jul 10 23:38:01 1999
-----------------------------------

    Options:
      File stem <sampled>

Class specified by attribute Eco

Read 984 cases (11 attributes) from sampled.data

Decision tree:

july16-29 <= 132:
:...jan15-28 <= 108:
:   :...june18-july1 <= 108:
:   :   :...feb26-mar11 > 106: Subtropical Arid (8.0)
:   :   :   feb26-mar11 <= 106:
:   :   :   :...apr9-22 <= 105: Temperate Arid (2.0)
:   :   :       apr9-22 > 105:
:   :   :       :...oct8-21 <= 104: Subtropical Arid (3.0)
:   :   :           oct8-21 > 104:
:   :   :           :...june18-july1 <= 105: Subtropical Arid (3.0/1.0)
:   :   :               june18-july1 > 105: Temperate Arid (2.0)
:   :   june18-july1 > 108:
:   :   :...july16-29 <= 121: Temperate Arid (50.0/4.0)
:   :       july16-29 > 121:
:   :       :...jan15-28 <= 105:
:   :           :...june18-july1 <= 121: Temperate Semi-Arid (3.0/1.0)
:   :           :   june18-july1 > 121: Temperate Arid (9.0/1.0)
:   :           jan15-28 > 105:
:   :           :...feb26-mar11 > 111: Temperate Arid (2.0/1.0)
:   :               feb26-mar11 <= 111:
:   :               :...nov19-dec2 <= 112: Temperate Semi-Arid (10.0)
:   :                   nov19-dec2 > 112:
:   :                   :...feb26-mar11 <= 107: Temperate Arid (2.0)
:   :                       feb26-mar11 > 107:
:   :                       :...apr9-22 <= 117: Temperate Semi-Arid (9.0/1.0)
:   :                           apr9-22 > 117: Temperate Arid (2.0)
:   jan15-28 > 108:
:   :...june18-july1 > 121:
:       :...nov19-dec2 > 122:
:       :   :...oct8-21 <= 128:
:       :   :   :...jan15-28 <= 120:
:       :   :   :   :...feb26-mar11 <= 114: Temperate Arid (2.0/1.0)
:       :   :   :   :   feb26-mar11 > 114: Subtropical Semi-Arid (18.0/5.0)
:       :   :   :   jan15-28 > 120:
:       :   :   :   :...dec3-17 <= 125: Subtropical Summer Dry (4.0/1.0)
```

```
:       :   :   :       dec3-17 > 125: Subtropical Semi-Arid (2.0)
:       :   :   oct8-21 > 128:
:       :   :   :...dec3-17 > 129:
:       :   :       :...jan15-28 <= 124: Subtropical Semi-Arid (2.0)
:       :   :       :   jan15-28 > 124: Subtropical Humid (2.0)
:       :   :       dec3-17 <= 129:
:       :   :       :...dec3-17 <= 124: Subtropical Semi-Arid (21.0)
:       :   :           dec3-17 > 124:
:       :   :           :...nov19-dec2 <= 129: Temperate Arid (4.0/1.0)
:       :   :               nov19-dec2 > 129: Subtropical Semi-Arid (13.0)
:       :   nov19-dec2 <= 122:
:       :   :...jan15-28 > 120: Subtropical Summer Dry (2.0)
:       :       jan15-28 <= 120:
:       :       :...june18-july1 <= 123:
:       :           :...feb26-mar11 <= 116: Subtropical Semi-Arid (4.0/1.0)
:       :           :   feb26-mar11 > 116: Temperate Arid (2.0/1.0)
:       :           june18-july1 > 123:
:       :           :...aug13-26 > 122: Temperate Semi-Arid (22.0/1.0)
:       :               aug13-26 <= 122:
:       :               :...nov19-dec2 > 118: Subtropical Semi-Arid (2.0)
:       :                   nov19-dec2 <= 118:
:       :                   :...june18-july1 <= 128: Temperate Arid (2.0)
:       :                       june18-july1 > 128: Temperate Semi-Arid (8.0/2.0)
:       june18-july1 <= 121:
:       :...feb26-mar11 > 119:
:           :...feb26-mar11 <= 121: Subtropical Arid (3.0)
:           :   feb26-mar11 > 121: Subtropical Summer Dry (3.0/1.0)
:           feb26-mar11 <= 119:
:           :...jan15-28 <= 109:
:               :...aug13-26 <= 109: Subtropical Semi-Arid (3.0)
:               :   aug13-26 > 109:
:               :   :...nov19-dec2 <= 112: Temperate Arid (3.0)
:               :       nov19-dec2 > 112:
:               :       :...june18-july1 <= 117: Subtropical Semi-Arid (3.0)
:               :           june18-july1 > 117: Temperate Arid (3.0/1.0)
:               jan15-28 > 109:
:               :...july16-29 <= 112: Subtropical Arid (14.0)
:                   july16-29 > 112:
:                   :...oct8-21 > 121:
:                       :...apr9-22 <= 113: Subtropical Semi-Arid (4.0)
:                       :   apr9-22 > 113:
:                       :   :...july16-29 > 120: Subtropical Semi-Arid (3.0)
:                       :       july16-29 <= 120:
:                       :       :...aug13-26 <= 121: Subtropical Arid (9.0)
:                       :           aug13-26 > 121: Subtropical Semi-Arid (3.0/1.0)
:                       oct8-21 <= 121:
:                       :...oct8-21 <= 114: Temperate Arid (4.0/1.0)
:                           oct8-21 > 114:
:                           :...july16-29 > 117:
:                               :...feb26-mar11 <= 114: Temperate Arid (5.0)
:                               :   feb26-mar11 > 114: Subtropical Semi-Arid (2.0/1.0)
:                               july16-29 <= 117:
:                               :...aug13-26 > 114: Subtropical Semi-Arid (11.0)
:                                   aug13-26 <= 114:
:                                   :...nov19-dec2 <= 115: Subtropical Semi-Arid (4.0)
:                                       nov19-dec2 > 115: Subtropical Arid (4.0/1.0)
july16-29 > 132:
:...apr9-22 > 132:
    :...aug13-26 <= 153:
    :   :...nov19-dec2 <= 131:
    :   :   :...feb26-mar11 > 123: Subtropical Summer Dry (13.0/3.0)
    :   :   :   feb26-mar11 <= 123:
    :   :   :   :...nov19-dec2 > 127: Subtropical Semi-Arid (5.0)
    :   :   :       nov19-dec2 <= 127:
    :   :   :       :...may21-june3 <= 142: Temperate Continental (2.0/1.0)
    :   :   :           may21-june3 > 142: Temperate Semi-Arid (6.0)
    :   :   nov19-dec2 > 131:
    :   :   :...may21-june3 <= 140:
    :   :       :...apr9-22 > 136: Subtropical Summer Dry (2.0/1.0)
```

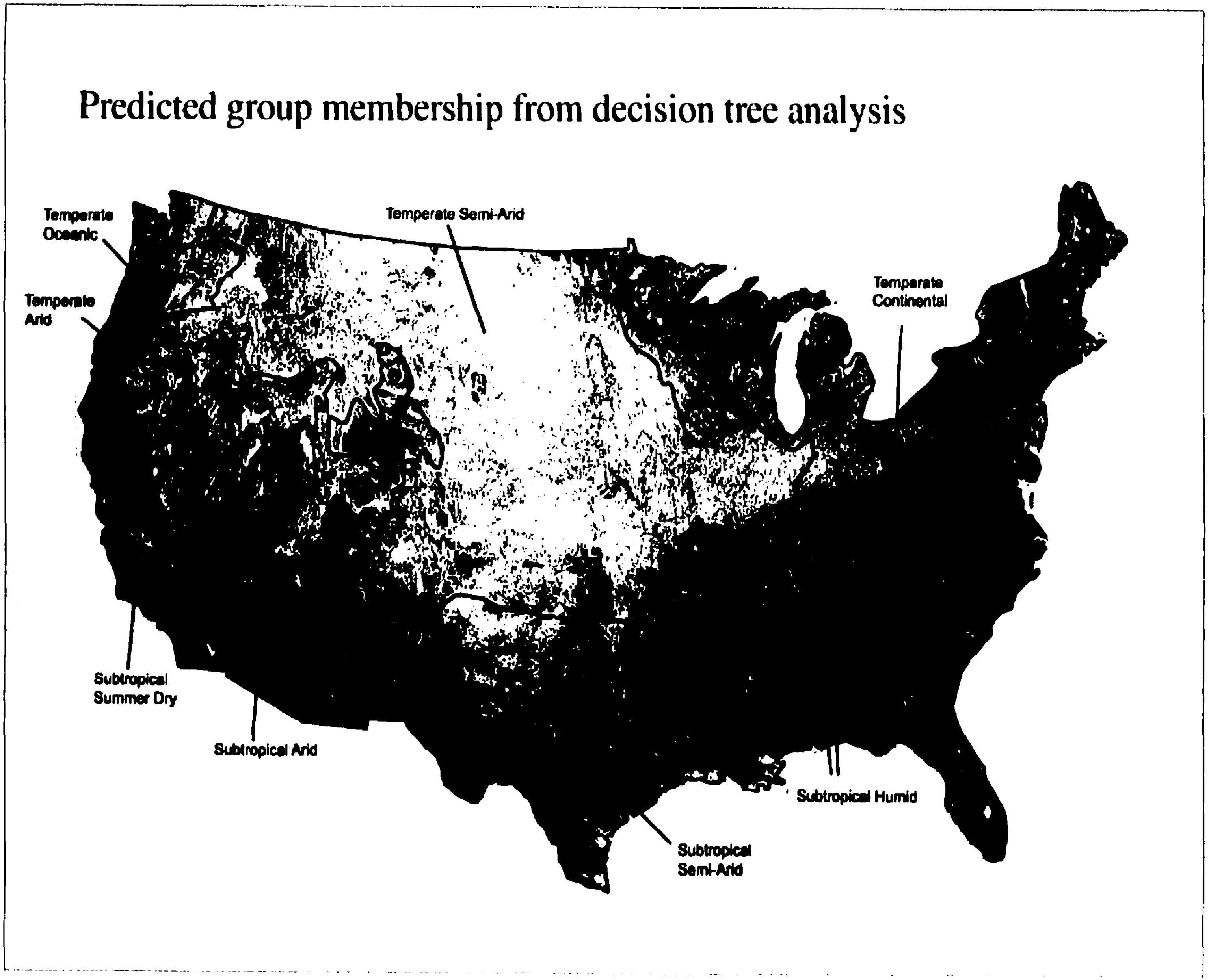
```
:   :       :   apr9-22 <= 136:
:   :       :   :...july16-29 <= 141: Subtropical Semi-Arid (3.0)
:   :       :       july16-29 > 141: Temperate Semi-Arid (2.0)
:   :       may21-june3 > 140:
:   :       :...aug13-26 > 146:
:   :           :...may21-june3 > 153: Subtropical Humid (52.0/8.0)
:   :           :   may21-june3 <= 153:
:   :           :   :...june18-july1 <= 152: Subtropical Humid (23.0/1.0)
:   :           :       june18-july1 > 152: Subtropical Summer Dry (4.0)
:   :           aug13-26 <= 146:
:   :           :...jan15-28 > 127:
:   :               :...apr9-22 <= 148: Subtropical Humid (36.0/4.0)
:   :               :   apr9-22 > 148: Subtropical Semi-Arid (3.0/1.0)
:   :               jan15-28 <= 127:
:   :               :...june18-july1 > 150: Subtropical Semi-Arid (4.0)
:   :                   june18-july1 <= 150:
:   :                   :...may21-june3 <= 146: Subtropical Semi-Arid (9.0/1.0)
:   :                       may21-june3 > 146: Subtropical Humid (5.0/1.0)
:   aug13-26 > 153:
:   :...jan15-28 > 135:
:       :...june18-july1 <= 160: Subtropical Summer Dry (4.0)
:       :   june18-july1 > 160: Temperate Oceanic (3.0)
:       jan15-28 <= 135:
:       :...june18-july1 <= 154:
:           :...july16-29 <= 155: Subtropical Humid (10.0/2.0)
:           :   july16-29 > 155: Subtropical Summer Dry (3.0/1.0)
:           june18-july1 > 154:
:           :...jan15-28 <= 117: Temperate Semi-Arid (3.0/1.0)
:               jan15-28 > 117:
:               :...apr9-22 <= 139: Temperate Continental (28.0/1.0)
:                   apr9-22 > 139:
:                   :...aug13-26 > 157: Temperate Oceanic (6.0/1.0)
:                       aug13-26 <= 157:
:                       :...jan15-28 <= 122: Temperate Oceanic (2.0/1.0)
:                           jan15-28 > 122:
:                           :...jan15-28 > 130: Subtropical Humid (2.0)
:                               jan15-28 <= 130:
:                               :...aug13-26 > 154: Temperate Continental (9.0)
:                                   aug13-26 <= 154:
:                                   :...jan15-28 <= 126: Temperate Continental (2.0)
:                                       jan15-28 > 126: Subtropical Humid (2.0)
apr9-22 <= 132:
:...july16-29 > 159:
    :...june18-july1 > 156:
    :   :...aug13-26 <= 156: Temperate Semi-Arid (2.0/1.0)
    :   :   aug13-26 > 156: Temperate Continental (103.0/4.0)
    :   june18-july1 <= 156:
    :   :...july16-29 > 163: Temperate Oceanic (2.0/1.0)
    :       july16-29 <= 163:
    :       :...aug13-26 <= 161: Temperate Continental (4.0)
    :           aug13-26 > 161: Temperate Semi-Arid (7.0/2.0)
    july16-29 <= 159:
    :...aug13-26 <= 154:
        :...nov19-dec2 > 128:
        :   :...june18-july1 <= 140:
        :   :   :...feb26-mar11 > 119: Subtropical Arid (3.0/1.0)
        :   :   :   feb26-mar11 <= 119:
        :   :   :   :...apr9-22 <= 125: Temperate Oceanic (2.0/1.0)
        :   :   :       apr9-22 > 125: Subtropical Semi-Arid (3.0)
        :   :   june18-july1 > 140:
        :   :   :...may21-june3 <= 140:
        :   :       :...nov19-dec2 <= 135: Temperate Semi-Arid (14.0/1.0)
        :   :       :   nov19-dec2 > 135:
        :   :       :   :...jan15-28 <= 122: Temperate Arid (2.0)
        :   :       :       jan15-28 > 122: Temperate Semi-Arid (3.0/1.0)
        :   :       may21-june3 > 140:
        :   :       :...dec3-17 <= 119: Subtropical Humid (3.0/1.0)
        :   :           dec3-17 > 119:
        :   :           :...nov19-dec2 > 136: Temperate Continental (2.0/1.0)
```

```
:   :                nov19-dec2 <= 136:
:   :                :...july16-29 <= 147: Subtropical Semi-Arid (2.0)
:   :                    july16-29 > 147: Temperate Semi-Arid (5.0/1.0)
:   nov19-dec2 <= 128:
:   :...apr23-may6 <= 119:
:       :...apr9-22 <= 116: Temperate Semi-Arid (74.0/3.0)
:       :   apr9-22 > 116: Subtropical Semi-Arid (3.0/1.0)
:       apr23-may6 > 119:
:       :...jan15-28 > 110:
:           :...aug13-26 > 133: Temperate Semi-Arid (47.0/3.0)
:           :   aug13-26 <= 133:
:           :   :...june18-july1 <= 136: Subtropical Summer Dry (2.0)
:           :       june18-july1 > 136:
:           :       :...feb26-mar11 <= 113: Temperate Arid (2.0)
:           :           feb26-mar11 > 113: Temperate Semi-Arid (5.0/2.0)
:           jan15-28 <= 110:
:           :...oct8-21 <= 128:
:               :...apr9-22 <= 120: Temperate Semi-Arid (14.0/1.0)
:               :   apr9-22 > 120:
:               :   :...may21-june3 <= 142: Temperate Arid (4.0)
:               :       may21-june3 > 142: Temperate Semi-Arid (2.0/1.0)
:               oct8-21 > 128:
:               :...july16-29 > 157: Temperate Arid (3.0)
:                   july16-29 <= 157:
:                   :...may21-june3 <= 135: Temperate Continental (9.0/1.0)
:                       may21-june3 > 135:
:                       :...aug13-26 <= 142: Temperate Continental (3.0/1.0)
:                           aug13-26 > 142: Temperate Semi-Arid (6.0)
aug13-26 > 154:
:...apr23-may6 <= 119: Temperate Semi-Arid (27.0/1.0)
    apr23-may6 > 119:
    :...june18-july1 <= 131: Subtropical Humid (4.0)
        june18-july1 > 131:
        :...feb26-mar11 > 122:
            :...feb26-mar11 <= 126: Subtropical Humid (4.0/1.0)
            :   feb26-mar11 > 126: Temperate Semi-Arid (2.0)
            feb26-mar11 <= 122:
            :...jan15-28 > 115: Temperate Continental (10.0)
                jan15-28 <= 115:
                :...jan15-28 <= 108:
                    :...oct8-21 > 143: Temperate Continental (10.0)
                    :   oct8-21 <= 143:
                    :   :...dec3-17 > 114: Temperate Semi-Arid (3.0)
                    :       dec3-17 <= 114:
                    :       :...feb26-mar11<= 111: Temperate Continental (14.0/2.0)
                    :           feb26-mar11 > 111: Temperate Semi-Arid (2.0)
                    jan15-28 > 108:
                    :...may21-june3 > 146: Temperate Semi-Arid (8.0)
                        may21-june3 <= 146:
                        :...oct8-21 > 145: Temperate Continental (4.0)
                            oct8-21 <= 145:
                            :...may21-june3 > 138: Temperate Semi-Arid (9.0)
                                may21-june3 <= 138:
                                :...apr23-may6>123: Temperate Continental (7.0/1.0)
                                    apr23-may6 <= 123:
                                    :...july16-29 > 151: Temperate Semi-Arid (6.0)
                                        july16-29 <= 151:
                                        :...oct8-21<=135: Temperate Semi-Arid (2.0)
                                            oct8-21>135:Temperate Continental (2.0)
```

Predicted group membership from linear discriminant analysis

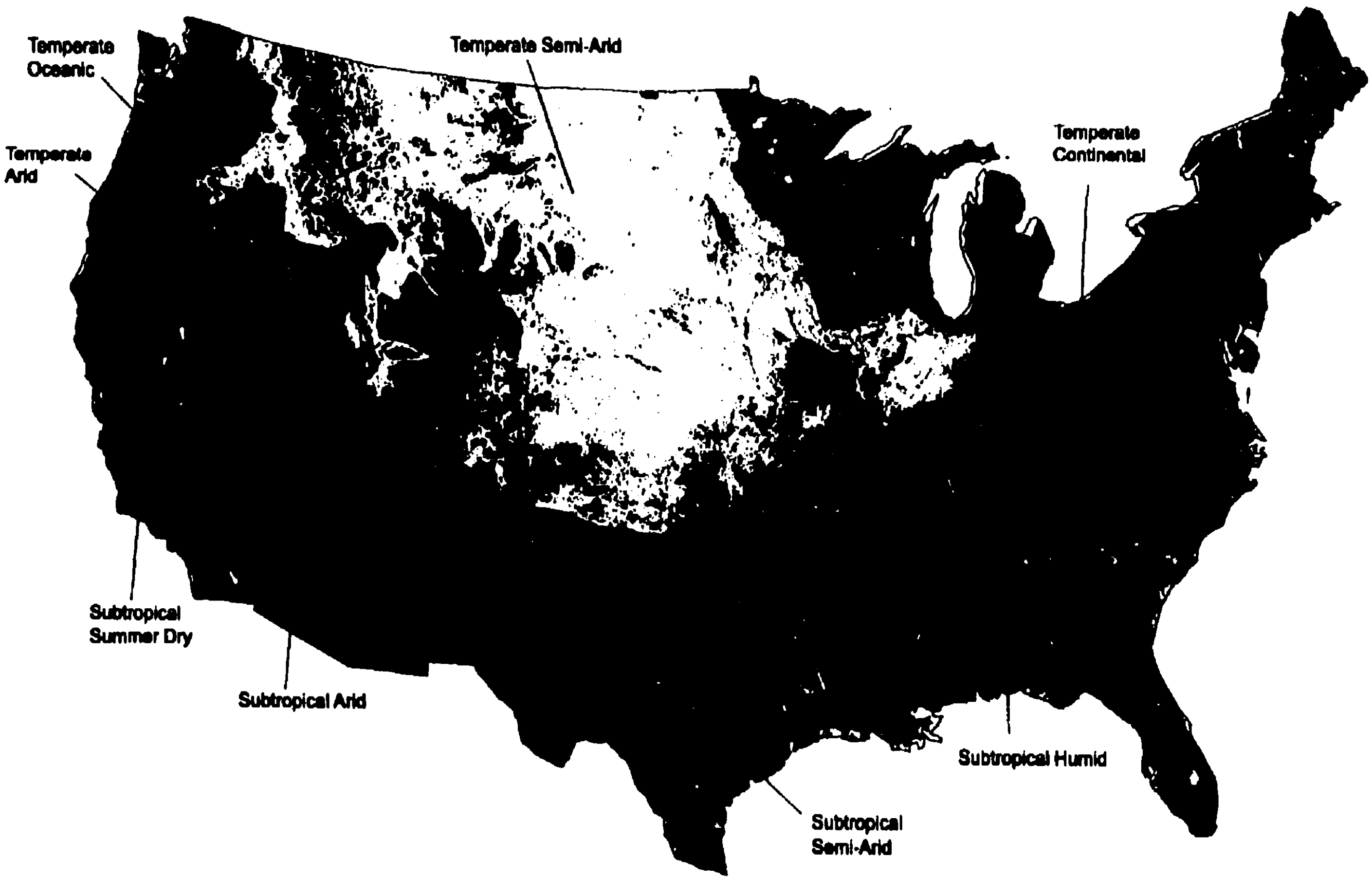


Appendix C: Map of Predicted Group Membership from Linear Discriminant Analysis



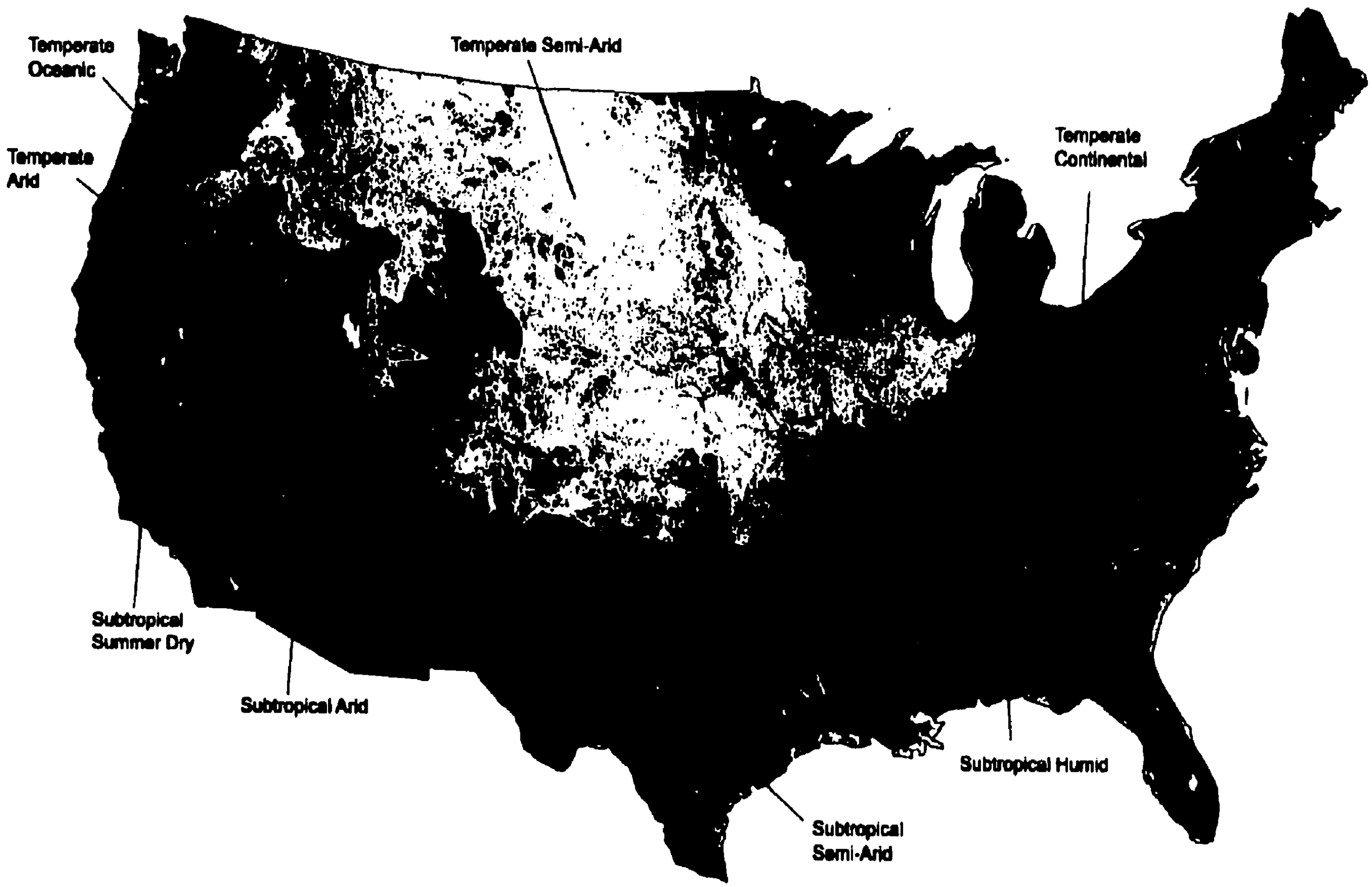
Appendix D: Map of Predicted Group Membership from Decision Tree Analysis

(black areas indicate pixels with predicted group membership not equal to "known")



Appendix E: Difference Map of **Predicted** Group Membership from Linear Discriminant Analysis

(black areas indicate pixels with predicted group membership not equal to "known")



Appendix F: Difference Map of **Predicted** Group Membership from Decision Tree Analysis