AN ABSTRACT OF THE THESIS OF

Sarah Jane Hash for the degree of Master of Science in Soil Science presented on
May 29, 2008.
Title: Use of Decision Tree Analysis for Predictive Soils Mapping and
Implementation on the Malheur County, Oregon Initial Soil Survey

Abstract approved:

_____

Jay S. Noller

Soil surveys provide essential information for making land use and management

decisions on publicly-owned lands in the semi-arid Great Basin. Soil maps produced with

conventional mapping techniques are time-consuming, costly, and do not explicitly

document the soil scientist's mental soil-landscape model. Predictive soils mapping using

decision tree analysis (DTA) can increase mapping efficiency and accuracy by extracting

relationships between soil types and environmental variables, applying these relationships

to predict soil types for unmapped areas, and explicitly documenting the process. While

DTA has been used for soils mapping in the past, no research exists concerning the use of

DTA for predictive soils mapping on an active Natural Resources Conservation Service

(NRCS) soil survey. This research documents the procedure for producing and validating

preliminary soils maps using DTA on the Malheur County, Southern Part Soil Survey

(MCSPSS), documents interactions with survey staff, and proposes a system for predictive

mapping implementation on the MCSPSS and other soil surveys.

In the early stages of the project, four sets of predictive maps were produced. The

June 16, 2007 predictive map used Owyhee County, Idaho soil survey information as

training data to predict soil types for three adjacent quads in Malheur County. Predictive

accuracy was low (67% at the order level, 61% at the suborder level, and 35% at the great group level) and this approach was abandoned. Subsequent predictive runs used soils mapping within the MCSPSS area, completed during the 2006 and 2007 field seasons, to predict soil map units (SMUs) for adjacent quads. The July 23, 2007 predictive map used training data from four east-west trending quads to SMUs for a surrounding annulus of quads. Accuracy improved significantly (87% at the order level, 73% at the suborder level, 67% at the great group level, and 53% at the subgroup level). The August 3, 2007 predictive map used all completed mapping from the 2006 and 2007 field seasons to predict SMUs for adjacent quads. Introduction of additional training data, representing greater area and more environmental variable combinations, did not improve accuracy (80% at the order level, 67% at the suborder level, 55% at the great group level, and 32% at the subgroup level). The September 14, 2007 predictive run used the same input variables and soils training data, but the soils information was recoded from SMUs to the subgroup level of soil classification. Accuracy was expected to increase because contiguous mapped areas would be larger and encompass more potential variation at a higher level of classification. However, accuracy was not significantly different from the August 3, 2007 predictive run. These results led to the prediction of SMUs for one quad, Threemile Hill, based on training data from four adjacent quads (October 12, 2007 predictive run). Accuracy improved substantially (91% at the order level, 83% at the suborder level, 83% at the great group level, 34% at the subgroup level). Confusion matrix analysis (for SMUs) was performed for the training area and the predicted area, and yielded overall accuracies and Kappa coefficients of 98.5% and 0.98 and 74.0% and 0.67 for the training and predicted areas, respectively. The April 10, 2008 predictive run used the same input

variables and training data as the October 12, 2007 run, but introduced surficial geology variables (surficial age and surficial lithology). The overall accuracy and Kappa coefficient were not significantly different from the October 12, 2007 run, but line placement was more precise at delineation boundaries and the locations of prediction errors were different.

DTA is effective for producing preliminary soils maps, but personnel, computer hardware, and software constraints currently prevent implementation on the MCSPSS. Skepticism and resistance among field scientists is also a major barrier. The proposed system for implementation requires certain personnel: an experienced field scientist with area-specific knowledge who would carry out the DTA and produce pre-maps, a GIS analyst who would provide data acquisition, data preparation and digitization support, and a soil survey project leader who would oversee the predictive mapping process. All would attend a week-long predictive mapping training session. An agency employee would conduct training sessions and serve as the support and issue resolution contact. Full documentation of procedures would be provided as an NRCS technical note. Adequate support in early stages of adoption is essential for long-term success.

Use of Decision Tree Analysis for Predictive Soils Mapping and
Implementation on the Malheur County, Oregon Initial Soil Survey

by

Sarah Jane Hash

A THESIS

Submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented May 29, 2008
Commencement June 2009

Master of Science thesis of Sarah Jane Hash presented on May 29, 2008.

APPROVED:

_____

Major Professor, representing Soil Science

_____

Head of the Department of Crop and Soil Science

_____

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

_____

Sarah Jane Hash, Author

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF APPENDIX FIGURES

LIST OF APPENDIX TABLES

Use of Decision Tree Analysis for Predictive Soils Mapping and Implementation on the Malheur County, Oregon Initial Soil Survey

1. Introduction

Soil mapping activities in the United States are carried out by the National Cooperative Soil Survey (NCSS), a nationwide partnership of federal, state and local agencies that work together to inventory, document, and encourage use of information about soils. The Natural Resources Conservation Service (NRCS), a division of the United States Department of Agriculture, leads and coordinates NCSS activities and employs numerous field scientists who develop comprehensive soil surveys for counties or regions of interest (U.S. Department of Agriculture, 2007).

Developing an inventory of an area's soil resources is critical for enabling land owners to make good management decisions. The complexity and variation in soils across the landscape greatly impact the potential land use—whether land is suitable for agricultural or forestry production, can be used as a foundation for roads, homes, and commercial buildings, or will support sanitary facilities and septic systems. These are some of the classic reasons for interest in a comprehensive soil survey. As a result, surveys have tended to exist only for counties with significant agricultural production and high population densities. However, many government agencies and private organizations now recognize the value of soil surveys for making decisions related to ecological health, habitat assessments, hydrologic projects, recreational use, wild land management, and grazing issues (just to name a few). This interest has sparked an increase in mapping activities for the traditionally unmapped, dominantly publicly-owned areas of the west.

The process of soil survey is often a tedious, time-consuming, and costly endeavor. Well-trained field scientists must cover tremendous expanses of land, dig and describe soil profiles, take detailed notes on physical site characteristics, make decisions on taxonomic classification, and carry out mental interpolation with the collected data to accurately draw soil unit boundaries. When drawing the unit boundaries, the soil scientist uses his or her knowledge of the soil forming factors—spatial and temporal environmental variables whose unique combinations produce unique soils—and how variations in these factors across the landscape dictate where a certain soil type will end and another begin. Traditional soil survey has drawn marked criticism because the process relies on the mapper's qualitative (and often undocumented) mental assessments of soil properties in relation to formative variables. The soil scientist's working model for soil development and distribution is the *state factor theory*. It was first proposed (though not in name) in the late nineteenth century by Dokuchaev (1883) and was later popularized by Hans Jenny (1941) in his famous book. Over the decades, it has stood the test of time and remains the accepted paradigm. The state factor theory is given as:

$$s = f\,(cl, o, r, p, t...)$$

where *s*oil is a *f*unction of independent variables such as *cl*imate, *o*rganisms and *r*elief acting on *p*arent material through *t*ime. The ellipses represent other factors not accounted for that may prove to be important. The state factor theory implies that soil properties can be inferred if values or conditions for the soil-forming factors are known. In addition to establishing the framework for traditional soils mapping, Jenny's model also allows a knowledgeable GIS analyst to use digital data sets that represent soil-forming factors (or other environmental variables co-varying with soils) and digital soils maps to extract

landscape relationships that can be used to predict soil types for unmapped areas. This predictive mapping process provides a thorough analysis of soil-landscape relationships, documents the "knowledge" used in creating the map, gives an estimation of the map errors, and can significantly reduce time and resources spent on mapping.

This research builds upon the work of Elnaggar and Noller (2008a; 2008b), who found that decision tree analysis (DTA) could be used to retrieve expert knowledge imbedded within soil survey data and then apply that knowledge to create predictive soils maps for adjacent areas. This research examines the feasibility of adapting those digital soils mapping (DSM) and pedometric techniques to develop predictive soils maps within an active soil survey area. Despite the abundance of literature exploring the use of DTA for natural resources mapping, no research exists that examines actual implementation on an active NRCS soils mapping project. The goals of this thesis are: 1) show that soils of the semi-arid Great Basin can be predicted using DTA to extract soil-landscape relationships from digital representations of environmental variables and available soil survey data, 2) identify critically important independent variables for use in the DTA, 3) document a protocol to quickly utilize soils information as it is generated to produce preliminary soils maps for subsequent mapping, often within the same field season, 4) analyze the accuracy of those maps, and 5) document interactions with NRCS and Bureau of Land Management (BLM) survey personnel over the course of this research. The second chapter will review literature pertaining to study area geology, climatic and vegetation history of eastern Oregon, study area soils, use of DTA in predictive soils mapping, and relevant technology acceptance issues. The third chapter will cover the methods, results and discussion of the predictive soils mapping research in Malheur County. This chapter is meant to be

subsequently published in Soil Science Society of America Journal and so it will read somewhat like a stand-alone document. The fourth chapter will document interactions with NRCS and BLM survey crews, and summarize recommendations for implementing predictive soils mapping on the Malheur County survey. The fifth chapter will provide a summary conclusion.

2. Literature Review

2.1. Geologic Setting

2.1.1. Regional Geology

Eastern Oregon's volcanic stratigraphy can be generally summarized with the following five stratigraphic divisions, listed oldest to youngest: (1) a foundation of pre-Tertiary oceanic sedimentary terranes that adjoin the North American pre-Cambrian craton to the east, (2) Oligocene to early Miocene calc-alkalic lavas and pyroclastic rocks, (3) mid-Miocene flood basalts--Steens basalts, Malheur Gorge-region basalts, and the Columbia River Basalt Group (CRBG), (4) localized but widespread calc-alkalic basalts and pyroclastic rocks associated with mid-Miocene to Holocene lithospheric extension in the Basin and Range province, and (5) late-Miocene to Holocene eruptions of high-alumina olivine tholeiitic basalts (Camp, 1995).

Much of the geologic research in eastern Oregon has focused on the chronology and chemical properties of the major flood basalt sequences. Steens basalt volcanism began about 16.6 Ma, coeval or just prior to the Imnaha and Grande Ronde basalt eruptions (CRBG). Scientists argue whether Steens basalt volcanism represents the earliest manifestation of the CRBG (Brandon & Goles, 1988; Camp, 1995; Geist & Richards, 1993; Hooper, 1997; Takahashi et al., 1998) or constitutes a distinct flood basalt sequence encompassing a greater time span and more eruptive centers than originally believed (Brueseke et al., 2007). Though often questioned in the past, it is now generally accepted that Steens basalt and CRBG volcanism occurred as a result of the impingement of the Yellowstone mantle hotspot, and that later-occurring small-scale calc-alkalic eruptions

resulted from lithospheric extension (Camp et al., 2003; Hooper et al., 2002). Much of

the study area is characterized by large lava plateaus formed in Steens and Oregon Plateau

basalts. The Owyhee Plateau and river canyon may serve as the best record of

mid-Miocene to Holocene basaltic volcanism in the Northwestern U.S. (Shoemaker & Hart,

2002).

2.1.2. Local Area Geology

The Geologic Map of Oregon (Walker & MacLeod, 1991) serves as a useful tool to

guide discussion of bedrock geology in the study area vicinity. This map shows Holocene

basalt and basaltic andesite in the vicinity of Cow Lakes and in the Jordan Craters area.

These flows originated from several small spatter cones and one major tephra cone, known

as Coffeepot Crater, and cover an area of approximately 72 km^2 (Otto & Hutchison, 1977).

An olivine basalt flow issuing from Coffeepot Crater dammed Cow Creek to create Upper

and Lower Cow Lakes (Kittleman, 1973; Otto & Hutchison, 1977; Hart & Mertzman,

1983). Thorough descriptions of the flow genesis and volcanic features at Jordan Craters

Volcanic Field (JCVF) are provided by Otto and Hutchison (1977) and Chitwood (1994).

The Walker and MacLeod (1991) geologic map legend states that relations to Mazama

pumice deposits indicate that the Jordan Craters basalt flows are less than 6,800 years old.

Otto and Hutchison (1977) give an estimated age of 4,000 to 9,000 years based on lichen

growth rates and degree of basalt weathering. Ferns (1997) cites an age of 2,000 to 4,000

years B.P. (source or dating method not provided). A report by Mehringer (1987) provides

a radiocarbon age of 3,200 years. In her popular layperson's book, *In Search of Ancient*

Oregon: A Geological and Natural History, E. M. Bishop (2003) states that the youngest

of the Jordan Craters basalts is 150,000 years old—an improbable overestimation. The

reference Bishop uses (Hart & Mertzman, 1983) gives 150,000 years as the estimated

maximum age of one piece of basalt from the main crater wall, and does not describe how

this estimate was derived. Lack of soil cover, preservation of surface flow features, lichen

growth rates, and degree of weathering all indicate that the basalt is Holocene in age.

Bondre (2006) identifies 14 monogenetic volcanoes that contributed late Miocene to

Holocene basalt flows in the Jordan Valley Volcanic Field (JVVF), which includes the

JCVF. This study describes the vent alignments, geochronology, compositional

heterogeneity, and petrology of the JVVF in great detail.

The young Jordan Craters basalt flows are flanked by hundreds-of-meters-thick

accumulations of thin individual flows of dyktytaxitic and porphyritic olivine basalt

(Miocene through Pleistocene in age) that stretch north- and westward to the Owyhee

River, eastward to the easternmost edge of Upper Cow Lake, and southward to Antelope

Reservoir. Pliocene and Miocene rhyolite and dacite begin on the southern flanks of

Mahogany Mountain, and trend southward along the perimeter of the Miocene and

Pleistocene basalt flows, before terminating at the edge of young basalt flows (found to the

south and west) and Miocene tuffaceous sedimentary rocks (found to the east all the way to

the Idaho border). These tuffaceous sedimentary rocks (from lacustrine and fluvial

deposits) trend southward to the vicinity of the Jordan Valley community and contain

inclusions of mafic vent rocks and olivine basalt flows. The town of Jordan Valley is

situated on the northeast edge of the east-west trending Antelope Valley graben (Pliocene

to Holocene in age), which clips the southern end of the Oregon-Idaho graben (Ferns,

1997). The Oregon-Idaho graben was formed during the late stages of Columbia River basalt volcanism (Cummings et al., 2000).

Hart and Mertzman (1983) provide a detailed explanation of the stratigraphy, chemical composition, and geochronology of the volcanic rocks and lacustrine sediments in the area between the Owyhee River and Jordan Valley. To the south of Jordan Valley, older tuffaceous sedimentary rocks, tuffs and ash flows flank the Miocene basalt and andesite of Juniper Ridge. Thick deposits of Quaternary sediments are found along Jordan Creek in the farming communities of Arock, Danner, and Jordan Valley and along Cow Creek, Indian Fort Creek, Dry Creek, and Soldier Creek. Outcrops of these deposits are also present along the Owyhee River near the Rome community.

2.2 Vegetation and Climatic History

Evidence for very different paleoclimatic conditions in eastern Oregon can be readily seen in soil profiles, ancient geomorphic surfaces, and relict landscape features. Palynology, limnology, and macro- and microfossil studies provide more specific documentation on plant communities and vegetation succession through time. Although few, if any, historic and pre-historic climatic studies have been carried out in the immediate study area, studies in other parts of eastern Oregon shed light on major climatic shifts and associated vegetation changes. Because changes in climate and vegetation are so intimately intertwined, they will be discussed simultaneously here. To constrain the scope of this discussion, only studies focused on eastern Oregon and the Northwestern Great Basin will be reviewed. Pre-Pleistocene vegetation and climatic history are summarized in Appendix I.

2.2.1. Pleistocene

Encroaching northern ice sheets did not invade Oregon territory during the Pleistocene. Ice caps and small glaciers developed at higher elevations in the Cascade Range, Klamath Mountains, Strawberry Range, Wallowa Mountains, Elkhorn Mountains, and Steens Mountain, and cyclically descended to lower elevations and then ascended according to fluctuating global glacial-interglacial temperature and moisture regimes. During the more humid times of the late Pliocene and Pleistocene, fault-block basins with no external drainage partially filled with water to form extensive lakes in eastern Oregon. The Alvord Desert and Coyote Lake playas are remnants of pluvial lakes found in Harney and Malheur Counties. Relict features from these closed-drainage pluvial lake systems are critical tools for teasing out the details of Pleistocene climate (Fritz, 1996; Negrini et al., 2000). Reheis (1999) and Reheis et al. (2002) suggest that a long-term drying trend from early to late Pleistocene, not seen in global marine oxygen isotope records, is evidenced by temporally decreasing shoreline elevations for western Great Basin pluvial lakes. Reheis (1999) estimates that effective moisture likely fell by a factor of 1.2 to 3 to produce the lake level decreases observed. During the Pleistocene as a whole, temperatures were generally lower and effective moisture was generally higher than present-day. In the Great Basin, full glacial periods had an effective moisture of about 7 to 9 cm yr^{-1} greater (Chadwick et al., 1995) and summer temperatures approximately 10° C cooler (Thompson, 1990) than present day; interglacial periods had moisture conditions and temperatures similar to today. Pluvial lake basins filled with water during glacial periods but later dried out and deflated during interglacials, resulting in increased aeolian activity. Studies by Negrini et al. (2000) and Cohen et al. (2000) look at sediment cores from Summer Lake, a remnant of

Pleistocene pluvial Lake Chewaucan in south-central Oregon, to derive a detailed

history of temperature and moisture fluctuations for the last 250,000 years in the northern

Great Basin. Fossil diatoms and pollen from Upper Klamath Lake provide higher

resolution for the last 45 ka and show that the last glacial maximum in this area occurred ca.

21 ka, with an abrupt decrease in glacial extent ca. 16 ka (Bradbury et al., 2004; Colman et

al., 2004). As temperatures warmed at the ends of glacial periods during the Pleistocene,

outburst floods from pluvial lakes often occurred as ice and debris dams melted or

collapsed and meltwater caused lake levels to rise. The floods often scoured channels,

altered connectivity of drainage networks, stripped soil and vegetative cover, and

eventually left thick deposits of rock, sediment and debris. The closest documented glacial

outburst flood to the study area occurred at Lake Alvord in the late Pleistocene, and its

present day landscape signature and estimated discharge are described by Carter et al.

(2006). Freshwater diatom fossils suggest deep open-basin conditions existed at ca. 13 to

14 ka, and the authors believe the most recent flood dates to that time, at the waning of the

last glacial period. Lake Alvord breached its eastern rim and spilled into Coyote Basin,

releasing approximately 11.3 km^3 of water. The water then escaped Coyote Basin through

two outlets, flowed into Crooked Creek, and eventually reached the Owyhee River just

north of Rome. The floodwaters eroded canyons and bedrock surfaces, and deposited

imbricated clasts as large as 4.1 m in diameter up to 30 m above the present-day Crooked

Creek channel! These outburst floods, powerful agents of geomorphic change, likely

occurred in many other locations across the Great Basin during the Pleistocene.

Detailed vegetation history covering the past 30 ka (since prior to the last full

glacial) has been elucidated by examining pollen and macrofossils in lake sediments and

woodrat (*Neotoma* spp.) middens in eastern Oregon and the northern Great Basin. Work by Nowak et al. (1994a; 1994b) suggests that although specific modern assemblages of xeric plant species in the northwestern Great Basin is a late Holocene occurrence, the individual taxa in the woodlands and shrub steppe have been present in their current ranges for at least 30 ka; Thompson (1990) believes that sagebrush steppe limits expanded downslope in the late Pleistocene. Mehringer (1987) states that sagebrush steppe has been the dominant vegetation on the Snake River Plain for the last 70 ka and that sagebrush and grass pollen dominate the record of the past 33 ka at Carp Lake in southeastern Washington (Mehringer, 1985). Cronquist (1978) envisioned cyclical migrations of plant communities during the glacial-interglacial reversals, where mesic species (conifer forests) would replace xeric species (shrub steppe) during cool, wet times and vice versa. Wells (1983) describes continuous conifer stands in many parts of the Great Basin during the late Pleistocene that retreated to their modern distributions at higher elevations as temperatures warmed heading into the Holocene. At Fish Lake on Steens Mountain, pollen records show that sagebrush followed retreating glaciers to an elevation of 2,300 m, its current upper elevation limit, by 12 ka (Mehringer, 1985). It becomes evident from extensive reading that although specific community composition and exact distributions may have fluctuated since the middle Pleistocene, many of the familiar sagebrush steppe, woodland and grassland taxa have been present for tens or hundreds of thousands of years. The complex topography and microclimates of the Great Basin allowed many different species to coexist during the Pleistocene, while high phenotypic plasticity and genetic diversity allowed those species to persist in changing environments.

2.2.2. Holocene

Climate began to warm and dry appreciably at the Pleistocene-Holocene transition. In the interior west, early Holocene warming was not manifested in higher mean annual temperatures, but rather came in the form of significantly hotter summers (Nowak et al., 1994a). Thompson (1990) suggests that while early Holocene temperatures were much warmer than that of Pleistocene glacial periods, they were still several degrees cooler than present-day; MAP was lower than that of glacial times as well. Mehringer (1987; 1985) reports that, based on sagebrush-to-grass pollen ratios, 8700 to 5400 b.p. was a time of deficient effective moisture at Fish Lake, Steens Mountain. The pollen record at Diamond Pond, in the Diamond Craters area south-southeast of Burns, shows alternating dominance of greasewood-shadscale desert, sagebrush steppe, and juniper grassland for the past 6,000 years, primarily as a function of MAP (Wigand, 1987; Mehringer, 1985; Mehringer & Wigand, 1990; Mehringer, 1987). Woolard (1937) resolved climatic fluctuations back to 1268 B.P. in eastern Oregon based on growth ring analysis of ponderosa pine (*Pinus ponderosa*) near Klamath Falls. Historically documented floods and droughts were accurately identified in the tree ring record. While there have been drastic fluctuations between drought and ample moisture during the past 700+ years, Woolard (1937) assessed that no overall trend toward wetter or drier conditions existed during the time period analyzed. This agrees with work performed by Meyer (1934) in the Blue Mountains of northeastern Oregon, so the general observations are most likely applicable to the whole of eastern Oregon.

Although humans have influenced landscape processes and vegetative communities for the last 12,000 years, the effect has been the most profound since the

arrival of pioneering settlers in the mid nineteenth century. Since Great Basin plant communities evolved with little pressure from grazers, the inception of widespread, uncontrolled grazing was devastating to many of the native plants and delicate cryptogamic soil crusts. Clearing of land, human-induced fire regimes, and water diversion also impacted plant communities. However, introduction of certain invasive annual forbs and grasses, namely cheatgrass (*Bromus tectorum* L.), has likely been man's most calamitous mark on the Great Basin landscape. Knapp (1996) provides a thorough literature review of the ecological history of cheatgrass in the Great Basin, examines its physiological characteristics that facilitate the filling of ecological niches, and explores how human interactions with the landscape exacerbate the problem of invasive species. Norton et al. (2002; 2004) examine how cheatgrass invasion impacts soil organic matter dynamics and alters other soil physical and biological properties in the sagebrush-steppe ecosystem. Other invasive species of concern in eastern Oregon are medusahead rye (*Taeniatherum caputmedusae*), various knapweeds (*Centaurea* spp.), musk thistle (*Carduus nutans*), Scotch thistle (*Onopordum acanthium*), Russian thistle (*Salsola iberica*), and toadflaxes (*Linaria* spp.), among others (Wilson & Young, 2008).

2.3. Study Area Soils

2.3.1. Surface Features

Some surface features of semi-arid soils in eastern Oregon, such as cryptogamic crusts and desert pavements, have an indisputable effect on soil function and development. Cryptogamic crusts (also known as biological, cryptobiotic, microphytic, microfloral, organogenic, or microbiotic crusts) (West, 1990) directly influence primary production,

soil fertility and nutrient cycling, structure development, water infiltration and retention, erosion severity, and soil thermal properties. The specific mechanisms of desert pavement genesis have been debated in the literature for several decades, but their impact on pedogenesis and hydrologic processes in desert soils is widely acknowledged. They are known to armor the soil surface against erosion, minimize infiltration and increase runoff, trap aeolian sediment, and adversely affect or exclude plant growth (Cooke et al., 1993). The literature examining the roles of cryptogamic crusts and desert pavements in arid and semi-arid ecosystems is vast, and will not be discussed in great detail beyond sufficiently documenting their importance in the study area.

Cryptogamic crusts are actually complex assemblages of tiny non-vascular plants, such as mosses, lichens, liverworts and algae, and other organisms such as fungi and cyanobacteria. Review papers by Belnap (2003) and West (1990) provide thorough overviews of the structure, composition, function, and disturbance responses of cryptogamic soil crusts, and also address their use as environmental indicators. Work by Ponzetti and McCune (2001) focuses specifically on crusts in eastern Oregon's shrub-steppe communities, and how composition varies with livestock impact, soil pH, dissolved salts, and other environmental variables and terrain attributes. Johansen (1993) summarizes research specific to North America and highlights the dispute over the importance of cryptogamic crusts in range management scenarios. Although ecologists and soil scientists generally believe that cryptogamic crusts are beneficial, many ranchers and land managers disagree. The work of Savory (1988) suggests that soil crusts inhibit the germination and establishment of vascular plant seedlings, and that short, intense periods of livestock grazing should be implemented to break up the crusts and allow

desired grasses and forbs to establish. Does the immediate reward of a potential increase

in plant establishment outweigh the long-term ecosystem services rendered by the soil

crusts? This is an important question in making land management decisions, because once

crusts are destroyed by vehicle, foot or animal traffic, recovery times are measured on

scales of decades to centuries (Belnap, 2003). Currently, cryptogamic soil crusts are not

catalogued in the Bureau of Land Management (BLM) Ecological Unit Inventory

procedure in Malheur County. They are plentiful throughout the study area in the

otherwise-bare spaces between shrubs and bunchgrasses. Although the crusts have little, if

any, direct effect on above-ground biomass estimates, their indirect effects on potential

production (through soil carbon contribution, nitrogen fixation, nutrient-rich dust capture,

and erosion mitigation, among others) may be quite significant.

Desert pavements are characterized by a surface layer of closely packed clasts that

overlies a soil matrix with relatively fewer (or no) clasts, and are particularly common on

poorly sorted alluvial deposits (Cooke et al., 1993). The traditional model for pavement

formation involves the depletion of fines by wind and sheetflow, upward migration of

stones via freeze-thaw activity, shrink-swell clay expansion, or salt heaving, and

subsequent rearranging of those stones by sheetflow at the surface. While many authors

contend that this specific combination of mechanisms may be responsible for pavement

formation on certain alluvial deposits where climatic conditions are favorable (Cooke,

1970; Cooke et al., 1993; Williams & Zimbelman, 1994), much research in the past 30

years suggests that this model doesn't account for the presence of many desert pavements.

Early works by Jessup (1960) and Cooke (1970) suggested that deflation might be

relatively unimportant, and put greater emphasis on the role of sheetflow. Williams and

Zimbelman (1994) point out that although the role of deflation is no longer implicated in

initial pavement formation, it is still considered important for the restoration of disturbed

pavements. Their research in the Mojave Desert emphasizes the importance of sheetflow

in the formation of pavements atop relatively young basalt surfaces where soil has never

been present, and therefore played no role in pavement formation. This work supports the

findings of Wells et al. (1985; 1987) and McFadden et al. (1987) who suggest that

pavements initiate at the surface with the weathering of fresh bedrock or basalt flows.

Bedrock is broken down via chemical and mechanical processes, and aeolian silt in-fills

the fractures to form the finer-textured soil beneath. Colluvial processes may also carry

detached clasts into aeolian silt-filled depressions. In this model, clasts are never deeply

buried in the soil matrix. Wetting and drying of aeolian material and salt crystal formation

displaces surface clasts enough to open fractures and allow additional in-filling of aeolian

fines. McFadden et al. (1987) go on to say that stone pavements on alluvial fans on desert

piedmonts throughout the western U.S. also likely form via this mechanism, and cite

similarities in the chemical composition of varnishes on clasts from surfaces of varying

ages. Once pavement formation begins, clasts act as a natural sieve for windblown

material. Physical breakdown of clasts, through insolation-induced stresses, diurnal

temperature changes and salt shattering, produces a tighter pavement that favors silt

accumulation and subsequent formation of Av (vesicular) horizons (McFadden et al., 2005;

Valentine & Harrington, 2006).

Relict periglacial features (frost-sorted patterned ground and stone stripes) have

been documented in the Snake River Plain of Idaho (Malde, 1961; Malde, 1964) and are

present in the study area. These features are readily visible in aerial photographs. Cox and

Hunt (1990) suggest that similar stone stripes on Oregon's Columbia Plateau are the product of collapsed pocket gopher tunnels, whereas Malde's (1961; 1964) position that stone stripes are simply solifluction-induced extensions of frost-sorted polygonal ground conforms more with aerial observations of features in eastern Malheur County. Many of the stripes occur on very low slopes, and Goldthwait (1976) confirms that stone stripes can form via solifluction on slopes as shallow as 3°.

2.3.2. A Horizons and Epipedons

Vesicular A horizons (Av horizons) are quite common in the study area. Vesicular horizons are fine-textured and characterized by disconnected ovoid pores that presumably form due to the heat expansion of trapped air and reduced soil shear strength after rain events (Yaalon, 1974; McFadden et al., 1998). Vesicular horizons almost always occur beneath a stone pavement. A rough surface is required to capture fine-textured dust, and the pavement then serves to protect the horizon from deflation and erosion (McFadden et al., 1998). Vesicular horizons most often have strong columnar parting to platy or platy soil structure. Increased horizon thickness from aeolian deposition and the subsequent flattening of vesicles likely causes the platy structure (1998). These horizons tend to decrease infiltration and increase runoff, having an obvious effect on plant communities, subsequent pedogenesis, and landscape development (Anderson et al., 2002). These horizons often occur over deep, well-developed soil profiles with argillic horizons that seem incongruent with what we know about the behavior of vesicular horizons. Some researchers provide evidence that Av horizons were formed exclusively during the Holocene (McFadden et al., 1986; Wells et al., 1985; Chadwick & Davis, 1990), and

resulted from increased dust flux at the Pleistocene-Holocene climate transition. If this is the case, desert soil profiles with argillic horizons would have formed during wetter Pleistocene periods, prior to vesicular horizon formation. Others, however, point to correlations between Av horizon and stone pavement properties on older surfaces to suggest that many Av horizons may be much older than Holocene (McDonald, 1994; Reheis et al., 1995). McFadden et al. (1998) suggest that because presently subhumid areas in or adjacent to North American deserts do not exhibit vesicular horizons or pavements, wetter Pleistocene climates would not have facilitated their development either. This evidence, along with thermoluminescence age estimates on flows of different ages in the Cima Volcanic Field, suggests that most vesicular horizons are Holocene in age. Micromorphology of vesicular horizons is covered by Sullivan and Koppi (1991) and Anderson et al. (2002).   Ap (plowed) horizons may be present where animal traffic has been sufficiently intense.

Epipedons in the study area may be mollic, umbric or ochric. This characteristic is an important one for predictive mapping, because slight differences in depth or color between two identical soils may result in one being classified as a Mollisols and the other as an Aridisols although no real differences in management or function exist. Malheur County soil survey staff recently developed an area-specific rubric for making the Mollisols-Aridisols decision and assigning expected subgroups (Vitrandic, Vitrixerandic, or Vitritorrandic) based on elevation, vegetation, precipitation zone and temperature regime (Alina Rice, personal communication, May 8, 2008). However, this approach was not documented or consistently used until the later stages of this study. This information is presented in Appendix II.

2.3.3. Subsurface Horizons and Features

Subsurface accumulations of illuvial clay, carbonates, or silica opal are the most common pedogenic features in soils of semi-arid and arid environments (Peterson, 1980). Argillic horizons in the semi-arid Great Basin are generally assumed to be relicts of Pleistocene pluvial periods that had more effective leaching regimes than present-day (Gile et al., 1966; Nettleton et al., 1975; Peterson, 1980; Southard & Southard, 1985), although work by Alexander and Nettleton (1977) and Peterson (1980) shows that argillic horizons can form much more quickly in the presence of high salinity. Holocene-age Natrargids may be found near lake margins and in playa basins where the dispersive action of exchangeable sodium accelerates rates of clay illuviation.

As salts, carbonates and clays are released through parent material weathering, salts move to the greatest depths within the soil profile and concentrate near the maximum depth of the wetting front. Carbonates are less soluble, and do not move as deep. Clays move in suspension or in solution with soil water and are deposited as the energy state of water decreases, and will be concentrated at even shallower depths (Nettleton et al., 1975). Many soils within the study area do not readily conform to this construct. It is common to find salts and carbonates (as free carbonates or in association with silica in a duripan) within or between argillic horizons. These polygenetic soils were likely free of salts and carbonates during the wetter Pleistocene. In the Great Basin, Holocene-age saline playas have provided a tremendous amount of aeolian dust, rich in readily weatherable minerals as well as carbonates and salts. These materials are deposited on the soil surface, enter the profile, and remain near the maximum wetting depth since modern-day effective moisture is insufficient to flush them through.

Pedogenic silica in eastern Oregon is derived mainly from aeolian dust deposits that are rich in volcanic ash and tephra, comprised largely of highly-soluble amorphous silica. The depth at which silica accumulates is commonly close to the top of the Bt horizon (Chadwick et al., 1989; Harden et al., 1991). The tendency of silica to precipitate at contrasting textural boundaries is evidenced by the formation of cicada burrow-shaped durinodes in the study area. Abandoned cicada burrows are infilled with coarser-textured surface soil, and as silica-rich water infiltrates and is temporarily perched at the textural discontinuity, SiO_2 precipitates to form a "cast" of the insect burrow. A duripan is present if silica accumulation results in relatively continuous induration and root restriction, and if the layer does not break down in the presence of hydrochloric acid (Soil Survey Staff, 2006a).

Opaline silica and calcium carbonate tend to occur together in duripans, although their precipitation chemistries and depths of maximum concentration differ somewhat. Chadwick et al. (1987a; 1987b) describe a concise depositional model for calcium carbonate and opaline silica in semi-arid soils. Calcium carbonate is more soluble, moves deeper in the profile during wetter years, and preferentially precipitates on previously formed calcite crystals, eventually plugging soil voids near the maximum wetting depth. Pedogenic silica does not dissolve as easily, but once in solution diffuses into smaller pores and requires high activation energy for the Si-O bond breakage necessary for precipitation. Although silica stays in solution longer than calcium carbonate, it tends to precipitate higher in the profile of fine-textured soils because $Si(OH)_4$ is readily adsorbed by hydroxyl groups on clay mineral surfaces. As the soil dries, these molecules precipitate to form SiO_2-OH.

Duripan studies on Idaho's Snake River Plain and Owyhee Plateau (Blank & Fosberg, 1991; Blank et al., 1998) suggest that many duripans in the study area are polygenetic. Repeated subaerial exposure and reburial by aeolian dust during Pleistocene climate shifts provided the silica-rich material needed for the observed thickness and degree of development (Blank et al., 1998). In addition, encapsulated dust particles (incorporated through fracture fill or encasement in precipitating calcite) likely provided an in-situ source of soluble silica (Blank & Fosberg, 1991), in contrast with the accepted model of duripan formation where soluble silica is derived exclusively from overlying horizons and deposited by percolating soil water.

2.3.4. Regional Soil Surveys

The Malheur County, Southern Part (OR644) survey area is bounded by three counties that have published soil surveys available online: Harney County Area, Oregon (OR628); Humboldt County, Nevada, East Part (NV777); and Owyhee County Area, Idaho (ID675). Of these four, only the Owyhee County Area, Idaho survey is close enough to the study area to provide a potentially relevant source of training data. The Malheur County, Northern Part (OR645) soil survey is also currently in progress.

According to the memorandum of understanding (MOU) for the Malheur County, Oregon, Southern Part soil survey (U.S. Department of Agriculture, 2005), a soil survey of the Fort McDermitt Indian Reservation (located on the border between Malheur County and Humboldt County, Nevada) was completed in 1974, but was never correlated and data was never entered into the National Soil Information System (NASIS). A private contractor mapped approximately 17,000 acres of privately owned farmland in the Jordan

Valley, Arock, and Rome areas in 1989 (2005). This data will require careful evaluation and validation prior to being incorporated into the current survey.

2.4. Decision-Tree Analysis in Predictive Soils Mapping

Predictive soils mapping techniques (PSM) seek to develop numerical or statistical models that quantify the relationships among soil type and values of various environmental variables that affect soil formation. The information in the model can then be applied to a new geographic area to produce a predictive soil map. McBratney et al. (2000; 2003) provide two excellent reviews of digital soils mapping and pedometric techniques, including linear models, generalized additive models, classification and regression trees, neural networks, fuzzy logic and geostatistical techniques (e.g. Kriging and co-Kriging). Scull et al. (2003) provide a separate review of literature pertaining to predictive soils mapping techniques and also examine the different data components required for effective characterization of environmental variables. This section will focus on decision and classification tree analysis (DTA and CTA) for predictive soils mapping.

DTA has been used extensively for natural resources mapping, especially for vegetation studies (Lees & Ritman, 1991; Franklin, 1995; Franklin, 1998; Vayssieres et al., 2000; Miller & Franklin, 2002; Simard et al., 2002; Brown de Colstoun et al., 2003; Thuiller et al., 2004) and habitat assessments (Guisan & Zimmerman, 2000). Australian soil scientists have performed a tremendous amount of research on predictive soils mapping using DTA. Bui et al. (1999) successfully derived predictive rules from a 1:100,000 scale soil survey in the Toowoomba region of Queensland, Australia, using only geology, a digital elevation model (DEM), and DEM-derived terrain attributes as

predicting variables. Moran and Bui (2002) improved the quality of the predictive rules for the same area by incorporating LandSat MSS data and additional terrain attributes and using boosting to reduce misclassification error. In 2003, Bui and Moran (2003) used DTA to produce a new 1:250,000 scale soil map for the Murray-Darling Basin, Australia's most important agricultural region. Henderson et al. (2005) adopted a DTA approach to use information from an Australian national soils point database along with environmental variables to build piece-wise linear models for the distribution of soil properties such as pH, organic carbon, total phosphorus, total nitrogen, thickness, texture, and clay percentage. This work suggests that widely-distributed point data can be analyzed with DTA to make predictions for soil characteristics over large land areas. McKenzie and Ryan (1999) also used DTA and linear models to map soil properties for 50,000 ha of forested land in New South Wales, Australia, but found that optimal selection of environmental variables was more critical than the modeling technique employed.

Luoto and Hjort (2005) explored several predictive mapping techniques (generalized linear models, generalized additive models, classification tree analysis, neural networks and multiple adaptive regression splines) for geomorphological mapping of patterned ground in northern Finland. Their work found that CTA performed the best when evaluated on training data, but had limited predictive power when extrapolated to new areas. The tree was likely grown too large in this instance and overfitted the rule sets to the training data. Scull et al. (2005) used CTA to develop a preliminary soil map for part of California's Mojave Desert, and modeled basins and mountains separately to increase map accuracy. The research also found that individual variables could be removed from the model without significantly reducing accuracy. In contrast with the findings of

McKenzie and Ryan (1999), variable selection had little impact on the overall model fitness, but greatly affected the types of rules that were generated.

Mapping, digitizing, and GIS analyses all introduce additional error into both soil maps and predictive variables. Lagacherie and Holmes (1997) explore error propagation in classification tree analysis, and propose an approach to take errors into account by manipulating the size of the final tree. Controlling the tree size via a "stopping rule" based on a risk threshold parameter prevents overfitting to the training data. Other authors discuss the use of "pruning" (Moran & Bui, 2002; Bui & Moran, 2003; Luoto & Hjort, 2005; Scull et al., 2005), a cross-validation method that reduces tree size by removing splits that are predicted to have a high error rate. Qi and Zhu (2003) and Qi (2004) focus on accounting for line misplacement error in source maps and reducing noise in predicted maps by only sampling modal pixels (pixels representing the central concept for a soil class) from environmental variable data sets.

2.5. Technology Acceptance

One of the main goals of this research is to examine how predictive soils mapping using DTA can be incorporated into an active soil survey. How likely are field scientists within the NRCS and BLM to accept and use a new computer-based tool? What factors contribute to technology acceptance in the workplace? What barriers exist for successful implementation? Understanding what factors provide for or prevent successful adoption will assist in tailoring the product to the needs of the agency and its individuals.

Most technology acceptance research has used the technology acceptance model (TAM) (Davis, 1989; Davis et al., 1989) or some variation as its basis. The TAM says that

individuals choose to use a new technology based on *perceived usefulness* (degree to

which use will increase performance) and *perceived ease of use* (how effortless using the

technology will be) (Venkatesh & Morris, 2000). Mathieson (1991) suggests that simple

barriers to use (affecting perceived ease of use), such as hardware or time constraints, often

prevent adoption regardless of recognized benefits of the technology (perceived

usefulness). Rogers (1995) describes how individuals can be classified according to when

they first begin using a new idea or technology, or their degree of "innovativeness", and

what influences affect adoption in each category. The plot of frequency of adoption over

time typically follows a bell-shaped curve, and the cumulative number of adopters follows

an S-shaped curve. Of course, this generalization only applies to technologies that are

successfully disseminated to most of the potential users. In Rogers' system, the very first

users of a technology (more than two standard deviations below the mean, or ~2.5%) are

called *innovators.* Innovators are, ideally, wealthy (or control money), can understand

complex technical information, and are comfortable with risk. *Early adopters* (between

one and two standard deviations below the mean, or ~13.5%) are normally leaders in the

community or workplace. Other potential adopters may look to them for advice. They

have likely garnered respect due to prudent and thoughtful adoption of previous

innovations. The *early majority* (between one standard deviation from the mean and the

mean itself, or ~34%) adopt new ideas just prior to the average individual in the

community. They do not typically hold leadership positions, but are important in the

dissemination process because they interact with other individuals in the later adopter

categories. Individuals in the *late majority* (between the mean and one standard deviation

above the mean, another 34%) usually adopt due to peer pressure or economic necessity,

and are typically skeptical of the proposed change. *Laggards* are the last individuals to adopt (the last 16%, or more than one standard deviation above the mean). They are typically isolated, suspicious, and traditional, look to the past for reference, and may have very limited resources. Rogers suggests that earlier adopters are not necessarily younger, as many assume, and doesn't mention gender. Earlier adopters are, however, typically more educated and literate, less dogmatic, and belong to larger, more developed social and employment networks.

Brown and Venkatesh (2003) applied Rogers' adopter categories to personal computer ownership and usage, focusing on each group's expected outcomes from technology acceptance. They found that earlier adopters were influenced primarily by hedonic and social status outcomes, while later adopters were driven by utilitarian outcomes and the social influences of their peers. Cost, knowledge requirements and fear of obsolescence may be major barriers for laggards. Venkatesh and Morris (2000) examine how social influence and gender affect individual technology acceptance. They summarize literature that suggests women are more concerned with ease of use (women often feel more computer use anxiety), while men focus on the perceived usefulness (men tend to be motivated by achievement, recognition and advancement). Women generally respond to social influence more so than men. Women are typically more people-oriented, responsive to social cues, aware of the feelings of others, and focused on group dynamics while men are typically more rebellious, independent and competitive. Bergeron et al. (1990) and Venkatesh and Morris (2000) suggest that providing technology support staff is critical in early stages of learning and practice to help overcome barriers (real and perceived) and provide a social support system. When seeking to introduce a new

technology, such as predictive soils mapping, into agency practices, it becomes

necessary to evaluate needs, resources and implementation barriers at the agency level as

well as at the working group and individual level.

3. Use of Decision Tree Classifier Software for Predictive Soils Mapping on the Initial Soil Survey of Malheur County, Oregon

3.1 Study area setting

Malheur County is located in the extreme southeastern corner of the state of Oregon. The southern Malheur County Survey Area covers approximately the bottom half of the county, as shown in Figure 3.1-1. Figure 3.1-2 shows the specific 7.5' quadrangles included in the study area and also shows the soils mapping that has been completed to date. In the study area vicinity, Cow Creek delineates the northern border of the survey area. According to the survey MOU (U.S. Department of Agriculture, 2005), the survey area totals 3,080,108 acres and land ownership is as summarized in Table 3.1-1. Most of the area is BLM rangeland and is dominantly used for livestock grazing, wildlife habitat, and recreation. Irrigated cropland is concentrated in stream valleys and terraces near the communities of Jordan Valley, Danner, Arock and Rome.

Mean elevation of the study area ranges from 970 m in the Owyhee River Canyon to 1850 m on Parsnip Peak (see Figure 3.1-3). Mean annual precipitation (MAP) is 32 cm (~12.5 in.), and ranges from about 22 cm (~8.5 in) along the Owyhee River near the community of Rome to about 54 cm (~21 in) on Parsnip Peak, near the Idaho border (see Figure 3.1-4). This range of MAP translates to soil moisture regimes that are aridic to xeric, respectively. Mean annual soil temperature (MAST) regimes are mesic at lower elevations and frigid at higher elevations, covariant with MAP. The survey crew has installed soil data loggers in several locations throughout the survey area to determine at what elevation the mesic-frigid break occurs and how the elevation break shifts with latitude.

Vegetation in the study area is composed primarily of sagebrush, other native shrubs, bunchgrasses and introduced grasses that occur in complicated mosaics. According to the Oregon Gap Analysis Project final report (Kagan et al., 1999), sagebrush steppe and big sagebrush shrubland are the two dominant vegetation cover classes in the area. These two cover classes are differentiated by the relative amounts of grasses and shrubs present. Sagebrush steppe vegetation has historically been dominated by bunchgrasses, but grazing pressure has resulted in increased dominance of sagebrush and a blurring of the distinction between the two types. Because of overgrazing and altered fire regimes, cheatgrass (*Bromus tectorum*), an invasive annual, is now the dominant understory species in both classes. Common shrub species include basin big sagebrush (*Artemisia tridentata* var. *tridentata*), Wyoming big sagebrush (*A. tridentata* var. *wyomingensis*), mountain big sagebrush (*A. tridentata* var. *vaseyana*), rigid sagebrush (*A. rigida*), low sage (*A. arbuscula*), silver sage (*A. cana*), threetip sagebrush (A. tripartata), rabbitbrush (*Chrysothamnus viscidiflorus* and *C. nauseosus*) and bitterbrush (*Purshia tridentata*). Common native bunchgrasses include bluebunch wheatgrass (*Agropyron spicatum*), Idaho fescue (*Festuca idahoensis*), basin wildrye (Leymus cinereus), Sandbergs bluegrass (*Poa secunda*), junegrass (*Koeleria macrantha*), Thurber needlegrass (*Stipa thurberiana*) and Indian ricegrass (*Oryzopsis hymenoides*). Crested wheatgrass (*Agropyron cristatum*), an introduced perennial grass that may compete with native grasses, is also common. It is worthy of notice, that if properly managed, *A. cristatum* can serve as a "stepping stone" for directed succession to reestablish native perennials and rebuild depleted or lost A and B horizons in terms of organic matter content and beneficial mycorrhizal components (Raven, 2004).

The study area extends across two Major Land Resource Areas (MLRA),

Owyhee High Plateau (MLRA 25) and Malheur High Plateau (MLRA 23). MLRA 25 is

characterized by lava plateaus and valleys bordered by long, gently sloping alluvial fans,

broad alluvium-filled basins, and north-south trending fault block mountain ranges

separating those basins. MLRA 25 is drained by the Owyhee River, which is part of the

Columbia River drainage system. MLRA 23 has the same landform types, but is

characterized by closed basins with no external drainage (Soil Survey Staff, 2006b). Small

areas of MLRA 24 (Humboldt Area) occur in the southern and western parts of the

Southern Malheur survey area, but it does not occur in the study area described here.

3.2. Soil Survey Framework

The following information is summarized from the Malheur County, Southern Part

Soil Survey MOU (U.S. Department of Agriculture, 2005). The Malheur County,

Southern Part Soil Survey (MCSPSS) commenced in summer 2005, and initial field

mapping began in summer 2006. Expected completion date is October 2014. The survey

is a cooperative effort between the Natural Resources Conservation Service (NRCS) and

the Bureau of Land Management (BLM) and includes both a traditional soil survey and an

ecological site inventory (ESI). ESI is a survey method developed by the BLM to

document vegetation cover and structure, species composition, potential natural vegetation,

annual biomass production, major soil features, hydrologic properties, physiographic

features, climatic data and rangeland health (U.S. Department of the Interior, 2001).

Individual ecological sites have distinct soils, hydrology and vegetation that result in

unique management considerations, and serve as minimum delineations for range

management decisions. An ESI must be completed either after or in coordination with a soil survey. Although this study focuses on the use of predictive mapping for soil survey, it is important to document the dual survey goals since final mapping units encompass both soils and ESI information.

Mapping intensity varies according to land use, environmental sensitivity, and potential production capability. Irrigated and non-irrigated cropland, wetlands and riparian areas are mapped at Order 2 intensity. Minimum map unit size for Order 2 is 20 acres, but map units as small as 5 acres may be delineated for areas of high resource value. All rangeland is mapped at Order 3 intensity. The minimum map unit size for Order 3 is 40 acres, but map units as small as 15 acres may be delineated for areas of high resource value. Order 2 intensity mapping requires that 75 percent of the delineations of each map unit be verified by direct observation, while Order 3 intensity mapping requires that 50 percent of delineations be verified. Other agency standards regarding sampling density and pedon description detail must be taken into account when using predictive soils maps. All field mapping is recorded on 1:24,000 scale Mylar aerial photographs acquired by the National Agriculture Imagery Program.

3.3. Methods - The Predictive Mapping Process

This project is designed to produce predictive soils maps and introduce this methodology to soil survey crews to facilitate, expedite and improve initial soil survey mapping in this region. Predictive soils maps in this study were produced according to the methodology of Elnaggar (2007) and explicitly documented by Hash (2006). The main mode of delineating map units for the MCSPSS is to be the distribution of the state factors

of vegetation according to MAP, elevation (proxy for MAST), geology and slope, based

on discussions with the survey project leader and NRCS MO-1 region staff. A general

summary of the procedures used in Malheur County for data preparation, decision-tree

analysis, verification and map production follows.

3.3.1 Data Collection and Pre-Processing

Once an area of interest is delineated, raster datasets representing independent

variables (variables that influence or reflect the dependent variable, i.e. soil type) are

collected or derived from existing data sets. Selecting data sets which quantify or describe

the different processes and conditions that affect soil formation for a given location is a

critical step. Because this survey is based on soil-forming factors, each must be

represented by digital data sets. Subclasses of the factors are described by one or more

databases (layers). For example, when trying to account for the effect of topography and

elevation, the analyst will download a digital elevation model (DEM) to represent the

specific elevation at each location. They may also derive slope and aspect grids from the

DEM. Solar Analyst (Fu & Rich, 1999), an ArcView 3.x extension, is one of several

programs that may be used to determine the type and amount of solar insolation at each

point across the area. Topographic position, plan and profile curvature, and drainage

networks can also be derived from the DEM. Vegetation characterization may include

representations of present and historical vegetative cover, along with measured reflectance

and band ratios from satellite imagery. Some independent variables are selected because,

as opposed to being causative in soil formation, they are believed to co-vary with soil

properties and thus serve as useful predictors. Satellite imagery and resulting band

calculations are examples of this data type. A layer representing the dependent variable (the variable we are seeking to predict), soils, is also needed. All datasets in the vector format must be converted to raster format. Then, they are projected to a common coordinate system (NAD 83 datum, UTM zone 11N), clipped to the study area extent (covering both the reference and unmapped areas), and resampled to a common cell size (30 meters).

Descriptive statistics are generated for each data set and box plots are created to detect outliers in continuous data sets. Rasters must first be converted to point shapefiles in ArcMap. ArcMap's Geostatistical Analyst provides the most comprehensive exploratory data analysis tools, but only works with feature layers. Using the Geostatistical Analyst extension to create histograms provides a comprehensive list of descriptive statistics: count, min, max, mean, median, standard deviation, 1st quartile, 3rd quartile, skewness (a measure of symmetry) and kurtosis (a measure of outlier likelihood). Histograms serve as useful tools for visualizing data distribution and are helpful if one is interested in the actual number of occurrences in a particular class or range. These summary statistics are only applicable to continuous data sets. For categorical data sets, the only analysis needed is a review of the bar chart or attribute table to ensure that values are valid and the distribution reasonable. Once rasters are converted to point shapefiles, ArcMap's Create Graph Wizard can be used to create box plots. Using the Create Graph Wizard with the original raster results in analysis based on the *range* of values of the data—that is to say, the tool looks at the attribute table and each potential value is considered as one data point, regardless of how many cells actually have that value. Values with a low number of occurrences have more influence on the overall statistics than they should, while values

with high occurrences are under-represented. Box plots created by ArcMap follow the

construct of Tukey (1977). The horizontal line in the middle of the box marks the median

of the data, while the lower and upper edges of the box (called the "hinges") represent the

first and third quartiles, respectively. Therefore, the box contains the middle 50% of the

data points and represents the interquartile range, or IQR (sometimes called the "box

length"). The ends of the whiskers (lines that extend out from the upper and lower hinges)

mark the minimum and maximum values, unless there are outliers. If there are outliers, the

whiskers will extend to no more than 1.5 times the IQR away from the lower and upper

hinges. Outliers are values that occur outside of this range. Mild outliers occur no more

than 3 times the box length away from a hinge, while extreme outliers occur more than 3

times the box length away from a hinge. After identifying the approximate values of

outliers, they are selected and highlighted. Outliers are then overlaid on the original raster

and a hillshade of the terrain. Visualization of outlier location helps the analyst decide

whether individual data points are errors or meaningful pieces of data. All erroneous

outliers must either be converted to new values or to no data. Histograms, descriptive

statistics and box plots were generated for all data sets in all runs. Histograms, descriptive

statistics, box plots and outlier maps for all continuous data sets used for the October 12,

2007 predictive map (described in Section 3.6.1), along with a brief discussion, are

presented in Appendix III.

3.3.2 Decision-Tree Analysis

Decision-tree analysis is conducted using three software packages--CART (Earth

Satellite Corporation, 2003), which runs as an extension in ERDAS IMAGINE (Leica

Geosystems, 2008), and See5 (Quinlan, 2007). The CART sampling tool in IMAGINE

is used to carry out file conversions, establish sampling grids, and set parameters needed

for the See5 decision-tree analysis software. The CART sampling tool allows the user to

set the filepaths to locate dependent and independent variables, specify the amount of data

to be used for training and validation, define the sampling method for selecting that data,

and establish and populate files needed for the See5 analysis. A file (the .names file) is

created and must be edited (using any text editing software) to specify whether each layer

is composed of continuous or discreet data. Discreet data layers must be accompanied by a

list of the possible values. The .names file directs See5 to sample a pre-defined percentage

of the area for generating rule sets to classify soil types, and another percentage to test the

validity of the rule sets.

     See5 uses the training sample pixels specified by the CART sampling tool to

generate the rule sets. It examines each pixel, looks at the values for each of the

independent variables associated with that pixel, and generates a decision tree which can

be converted to an "If-Then" set of rules for quantifying relationships for the independent

variables (environmental predictors) and the dependent variable (soil type). Once the rules

have been generated, See5 uses the validation sample pixels specified by the sampling tool

to check the accuracy of the rule sets. It applies the rules to each of the validation pixels,

generates a prediction, and then compares its prediction to the actual value for the pixel

specified by the dependent variable band. The program will then create a detailed set of

statistics (for each rule and for the overall rule set) describing its accuracy in predicting the

dependent variable. Two parameters in the See5 program, boosting and pruning, increase

the accuracy of predictive maps by revisiting misclassified pixels to build new trees and

controlling the classification tree size to prevent overfitting of the model. Boosting

allows See5 to create multiple classification trees, instead of just one, to reduce

misclassification error. After the program builds one decision tree and classifies pixels, it

checks the accuracy of the rules by making predictions for the validation data. Inevitably,

there will be errors in the prediction. The program then builds a second tree, concentrating

on the misclassified cases from the previous analysis. While the overall error for this tree

may be higher, more of the misclassified cases are correctly classified. Each classifier, up

to 10 total, votes for its predicted class for each pixel. The result is a lower overall error

after all classifiers have weighed in with their prediction. Ten boosting trials were used for

all predictions described here. Pruning controls the size of the final decision tree to prevent

overfitting. Decision trees are built so that they fully classify all of the training data based

on the independent variables provided. However, all of the quantified relationships may

not prove to be meaningful and may result in high misclassification error when applied to a

new area. Removal of rule sets, called pruning, removes parts of the tree that are observed

to have a high error rate. All of the classification trees described here were pruned by 25%

(i.e. the upper quartile of the worst prediction rules), and every branch in the tree was

required to describe at least two cases. The See5 output includes a list of all the rules that

were generated for every boosting trial, detailed statistics about the number of cases

classified by each rule, and the overall accuracy of the predictive rule sets. The built rules

are tested once on the training data (used to build them) and once on the separate subset of

validation data. Resulting accuracy assessments are provided individually. See5 provides

a ranking of independent variable importance (given as a percentage, which represents the

theoretical decrease in prediction accuracy if that variable were removed), and specifies if

any variables were winnowed (shown not be statistically significant) in building the decision tree. Table 3.3-1 shows all independent variables with their sources and original scale/resolution.

After performing the decision-tree analysis, the rule sets can be applied to the area of interest using the See5 Run tool, found in the CART extension of ERDAS IMAGINE 8.7. This tool examines the independent variable values in area of interest, pixel by pixel, and assigns a dependent variable value based on the applicable rule with the highest confidence. The result is a raster representation of predicted soil types.

3.4. Approaches to Prediction

Training data selection can vary depending on the amount, type and location of available soil survey information as well as variations in environmental conditions across the study area. Five different approaches to predictive map generation are graphically summarized in Figure 3.4-1. In approach A, a one-to-one prediction is made for the study area based on soils information already available for that area. This approach is useful for model development and accuracy assessment (Elnaggar & Noller, 2008a). See5 uses this approach to test the rule sets it creates. The rule sets are then used to generate a prediction according to one of the other four approaches shown here. In approach B1, soils training data for a mapped area is used to predict an adjacent unknown area (Elnaggar and Noller, 2008b; Pace, 2007). This approach would be used to take soil survey data from one county and predict an area in a bordering county. Approach B2 uses a small mapped area to predict outward for the surrounding area on all sides, while approach B3 uses available training data to fill in holes in the map coverage. Approaches B1, B2 and B3 will be

discussed in the following section. Approach C uses available training data to predict

for a distant, non-adjacent area. This approach would be useful within a given MLRA

where soil-forming conditions are assumed to be analogous (Malone, in preparation).

3.5. Results and Discussion – Early Predictive Mapping Runs

This study was initiated during the second field season of the MCSPSS and thus

dates of predictive soils maps are an important documentation of the progression in

development and application of results by field crews (see Tables 3.5-1 and 3.6-1). All

predictive maps were validated using soil pedon descriptions collected at referenced GPS

points during the 2006 and 2007 field seasons. The first predictive map was generated for

the study area on June 16, 2007. This map was based on training data from the Owyhee

County, Idaho soil survey. The Stonehouse Creek, Flint, Swisher Mountain and De Lamar

quads in Idaho were used to predict the Jordan Valley, Antelope Reservoir, Hooker Creek

and Downey Canyon quads in Oregon (prediction approach B1). A small selection of

independent variables was used (elevation, slope, aspect, mean annual precipitation,

minimum January temperature and maximum July temperature). Predictive accuracy was

generally low (67% at the order level, 61% at the suborder level, and 35% at the great

group level). This approach was abandoned for several reasons related to NRCS concerns

over inapplicability of the Owyhee County soil taxonomy and soil-landscape model.

Subsequent predictive mapping runs used data collected during the 2006 and 2007

MCSPSS field seasons as training data, and introduced additional independent variables.

The predictive mapping run completed on July 23, 2007 used mapping from an east-west

trending set of quads (Juniper Ridge, Little Grassy Mountain, Dry Creek Rim and Scott

Reservoir) to predict a surrounding annulus of quads (prediction approach B2). Overall

accuracy improved substantially (87% at the order level, 73% at the suborder level, 67% at

the great group level, and 53% at the subgroup level) but there was tremendous variability

in prediction accuracy from quad to quad. Map units were predicted with very low overall

accuracy (14%). Accuracy also decreased with increasing distance from training data

boundaries. The ten most important predictor variables (all with an estimated importance

of >10% as ranked by See5) were elevation, mean annual precipitation, minimum January

temperature, average minimum temperature, maximum July temperature, bedrock geology,

average maximum temperature, slope, GAP analysis vegetation cover, and LandSat TM

band 6 reflectance. The next predictive run, completed on August 3, 2007, incorporated

additional training data digitized from recently-produced preliminary field maps for three

quads (Danner, Threemile Hill and Jordan Valley) and also utilized all previously available

training data from the 2006 field season (10 quads total) (prediction approach B2). It was

expected that incorporating more training data (more area with more environmental

variable combinations) would increase accuracy. However, accuracy decreased relative to

the July 23 run (80% at the order level, 67% at the suborder level, 55% at the great group

level, and 32% at the subgroup level). The nine most important predictor variables (all

with an estimated importance of >10% as ranked by See5) were elevation, mean annual

precipitation, bedrock geology, minimum January temperature, average maximum

temperature, July maximum temperature, average maximum temperature, GAP analysis

vegetation cover, and historic vegetation. For the September 14, 2007 predictive run, the

soils training data from the August 3 run was used again, but all of the data were recoded to

the subgroup level of classification (prediction approach B2). It was expected that this

would increase accuracy because contiguous mapped areas would be larger, encompass more potential variation, and require coarser separations than at the map unit level. Again, the results did not support this hypothesis. The accuracy was not significantly different (79% at the order level, 65% at the suborder level, 50% at the great group level, and 30% at the subgroup level). The eleven most important predictor variables (all with an estimated importance of >10% as ranked by See5) were elevation, mean annual precipitation, bedrock geology, minimum January temperature, average minimum temperature, July maximum temperature, diffuse radiation received (W-h/m^2), average maximum temperature, GAP analysis vegetation cover, historic vegetation, and LandSat TM band 6 reflectance. Predicted soils maps, sampling points, accuracy summaries (overall and quad-by-quad) and soil map unit descriptions for all early runs are shown in Appendix IV. Table 3.5-1 summarizes the prediction approach used and the resulting accuracy for each of the early runs, and Table 3.6-1 summarizes information for the two Threemile Hill predictive runs that are described in the following section.

3.6. Results and Discussion - Threemile Hill Predictive Maps

3.6.1. October 12, 2007 Predictive Map

In light of results from early predictive runs (accuracy is highest near borders with training data, incorporating more training data from a larger geographical extent doesn't necessarily improve accuracy), a predictive map of one quad, Threemile Hill was produced on October 12, 2007. A raster representation of the soils map for four surrounding quads (Jordan Craters South, Arock, Danner and Dry Creek Rim) was used as training data (prediction approach B3). Figure 3.6-1 shows the soil map units for the training area.

Refer to Appendix V for map unit descriptions. Accuracy assessment for the predicted

map was initially assessed based on GPS-referenced soil pedon descriptions collected

during the 2007 field season. Figure 3.6-2 shows the predicted soil map units for the whole

area, and Figure 3.6-3 shows a comparison of the predicted and actual soil map units for the

Threemile Hill quad. Accuracy increased to 91% at the order level, 83% at the suborder

level, and 83% at the great group level. However, accuracy was still only 34% at the

subgroup level. The eleven most important predictor variables (all with an estimated

importance of >10% as ranked by See5) were bedrock geology, mean annual precipitation,

average minimum temperature, January minimum temperature, elevation, average

minimum temperature, LandSat TM band 6 reflectance, diffuse radiation received

(W-h/m^2), slope, July maximum temperature, and LandSat TM 4/3 band ratio. In light of

accuracy improvements relative to other runs, confusion-matrix analysis was performed to

generate descriptive statistics regarding accuracy among individual map units and to get a

better view of overall map accuracy.

3.6.2. Confusion Matrix Analysis for October 12, 2007 Predictive Map

A confusion matrix is a table that catalogs the agreement between all pixels in the

predicted and actual soil maps. It summarizes the number of pixels assigned to each soil

class (potential pixel value) that were correctly classified, and then summarizes the number

of incorrectly classified pixels and the classes to which they were assigned. This provides

a more comprehensive summary of the predictive map's accuracy than what could be

assessed based on data from field-check points. An error map can also be generated that

shows the actual location of all incorrectly classified pixels. Identifying the conditions

under which the misclassification occurs facilitates model improvement by allowing the

analyst to select additional (or different) input variables or perhaps modify the structure or

unit nomenclature of the training map. Certain summary statistics (overall accuracy, user's

and producer's accuracy for each class, and Kappa index) are also calculated when the

error matrix is produced.

The *overall accuracy* is simply the total number of correctly classified cases

divided by the total number of cases. User's and producer's accuracy are calculated for

each individual class. The *user's accuracy* for a given class (say class x) measures the

accuracy of the prediction by looking across the rows of the error matrix. It is a measure of

commission error—when a pixel that is actually from another class is committed to class x.

The *producer's accuracy* measures the prediction accuracy by looking down the columns

of the error matrix and is a measure of *omission* error—when a pixel that is actually in class

x is omitted and assigned to a different class. These measures tell the analyst if pixels from

a certain class are being classified incorrectly (low producer's accuracy) or if that class is

"catching" pixels that should be in other classes (low user's accuracy). Consider the

situation where nearly all of the pixels that should be in class x are correctly assigned to

class x (high producer's accuracy) but many pixels that actually belong to class y and z are

also assigned as class x (low user's accuracy). This would tell the analyst that while class x

is very accurately predicted and easily distinguishable from other classes, classes y and z

must have some characteristics similar to x. Examining their distributions and thinking

about their differences in terms of formative factors or predicting variables might lead to

selection of another variable that helps to discriminate between the classes. The *Kappa*

index (Cohen, 1960) is a discreet multivariate statistic used to measure the agreement

between actual and predicted soils maps. Kappa is a more robust measure than the

overall accuracy, because it takes into account the amount of agreement that would occur

by chance. Kappa, as adopted for use in image analysis (Mather, 2004), is calculated as:

$$K = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} * x_{+i})}$$

where N is the total number of pixels, r is the number of rows in the matrix, x_{ii} are the

diagonal entries of the error matrix, x_{i+} and x_{+i} are the sums of row i and column i,

respectively, and K is the Kappa value. Kappa values may range from 0 (no agreement

above that expected by chance) to 1 (perfect agreement).

A new approach was developed for creating confusion matrices. ENVI 4.3 (ITT

Industries, 2006), a geospatial imagery analysis software package, has the capability to

produce a confusion matrix to compare two classified images. However, the tool will only

work with an image that has been classified in ENVI. The images used with this tool are

generally multi-spectral satellite images that have undergone supervised classification

whereby unique regions of interest (ROI's) are delineated that represent activities or

conditions on the ground (i.e. land use, drought severity). These ROI's are then used as a

sort of "training data" to classify all the pixels in the image. The selection of ROI's can be

a time-intensive process when classifying an image because of the continuous nature of

reflectance data in multiple bands—the analyst needs to delineate representative ROI's that

fully capture the variation in reflectance within each class. Although ENVI's supervised

classification tools are not intended for classifying rasters that already represent categorical

data, simply defining one ROI for each soil class (only a few pixels are needed) allows

the classifier to identically reproduce the soil map units as a classified image. Since there

is only one possible pixel value for each soil class, there is no error in the classification.

Completing this process for both the predicted and the ground truth soil map results in two

classified images that can be compared and the results summarized in the confusion matrix.

The specific procedure is detailed in Appendix VI. Two confusion matrices were

generated for the October 12, 2007 predictive map—one for the training area (Jordan

Craters South, Danner, Dry Creek Rim, and Arock quads) and another for the predicted

area (Threemile Hill quad). The confusion matrices and summary statistics are also shown

in Appendix VI.

Examining the confusion matrix for the training area gives an idea about the overall

appropriateness of the model and independent variable selection and can also provide clues

about where misclassification is likely to occur in the unmapped area. Because the

confusion matrix compares a predicted map to the actual map from which the training data

was derived, accuracy should be very high. In this case, the overall accuracy is 98.5%,

whereas the Kappa coefficient is 0.98. Where is error occurring in the training area?

Evaluating the error in every map unit would be time consuming and impractical since less

than half of the map units (16 of 37) comprise 95% of the total area (see Tables 3.6-2 and

3.6-3). This section will address these major map units. Of those 16 units, 13 units (140C,

108C, 184C, 199B, 998, 102C, 189B, 196C, 127C, 122C, 175C, 120D, 266B) had a

prediction accuracy of 96% or greater, and the error map (Figure 3.6-4) shows that nearly

all misclassified pixels occur at unit boundaries. Unit 186C, Vitrixerandic Haplocambids,

has 86% of its pixels classified correctly. Five percent of the pixels were classed as 108C

(Vitrixerandic Argidurids). Another eight percent were incorrectly placed into units that do include Haplocambids (196C and 187C). Unit 200D, another Haplocambids unit, had 84% of its pixels classified correctly. Of the incorrectly classified pixels, 13% are identified as 998 (lava flows), in part because the largest 200D delineation borders the 998 delineation in the northern part of the Jordan Craters South quad. Haplocambids units have been poorly predicted in past experiments, but it is likely that a surficial geology map would reconcile this problem. Haplocambids likely occur on younger aeolian-influenced surfaces, while the more ubiquitous Argidurids likely occur on older surfaces subjected to Pleistocene weathering and moisture regimes.   Unit 191C (Vitrixerandic Argidurids-Vitrixerandic Haplargids complex) had a 92% prediction accuracy. Of the misclassified pixels, 5% were classified as 102C, which is the same series as the main component of 191C (Muni series). Another 2% were classified as 120D (rock outcrop-rubbleland complex), with which this delineation shares an extensive border. In reality, soil types do not have discreet boundaries—they vary continuously across the landscape and blend into one another. It appears that most of the error in predicting the training area is a result of the uncertainty inherent in placing discreet map unit boundaries.

   A confusion matrix should be created for the area of interest once the analyst has the ground-truthed field map. Because the map is created by extrapolating rule sets outside the area of creation, accuracy is expected to be lower. Constructing the confusion matrix requires that every class from the predicted map have an equivalent class on the ground-truth map. Because of this constraint, it is common to omit units from this analysis. Units 129A, 222 and 224 were actually mapped in the Threemile Hill quad but were not part of the predictive map, so they were not included in the confusion-matrix analysis.

Unit 129A was not present anywhere in the training data, so its prediction in the area of

interest is impossible. Together, these three units comprise a statistically insignificant

areal percentage (<2%). For the Threemile Hill area, the overall accuracy is 74% and the

Kappa coefficient is 0.67 (at the map unit level) (see Table 3.6-1). Unit 140C covers the

most area in the Threemile Hill quad (41%). This unit was also predicted with the highest

accuracy (94%). Unit 140C is comprised of shallow Abruptic Xeric Argidurids and seems

to have a strong correlation to broad, low-sloping Tertiary tuffaceous sedimentary rock

plateaus. Unit 184C, which covers 22% of the area, is comprised of 65% rock outcrop 25%

Vitrixerandic Argidurids (Clarksbutte series). It was predicted with 89% accuracy. About

six percent of the pixels were incorrectly classified as 189B, which is comprised of 40%

Vitrixerandic Argidurids (also Clarksbutte series), 30% rock outcrop, and 15% Typic

Haplargids. 184C occurs on younger Quaternary basalt flows, while 189B occurs on

slightly older (still Quaternary), more weathered flows. Another four percent of the pixels

were misclassified as 140C and occur near the unit boundary. Unit 189B, described above,

covers 13% of the total area and was predicted with 85% accuracy. Most of the

misclassified pixels (~10%) were classified as 196C (60% Durinodic Haplocambids and

25% Typic Argidurids). Likewise, unit 196C (9% of the area) was predicted with 89%

accuracy and had most of its misclassified pixels classified as 189B. These two units

border one another, and parent material and landform seem to be major factors

distinguishing the two. When looking at the geology map, hillshade and digital orthophoto

quad, it appears that the Quaternary basalt that 189B occupies may have been mapped a

little too extensively. In this instance, it appears that the predictive map may have

delineated these two units with higher accuracy (using reflectance data and other factors in

the rule sets) than the actual field map, but since there are no sampling points in this overlapping region it is not possible to say so with much certainty. Unit 108C (shallow Vitrixerandic Argidurids) covers approximately four percent of the total area, but was mapped with only 25% accuracy. Most of the misclassified pixels were classified as 140C (shallow Abruptic Xeric Argidurids). These two soil types are very similar in their morphologies—the clay content and increase with depth in the Bt horizon is the only thing that necessarily separates them. Another 14% were mapped as 196C (60% Durinodic Haplocambids, 25% Typic Argidurids), which is a bordering unit. Unit 108C was predicted with very high accuracy in the training area (97%), but it appears that it has been inconsistently mapped with respect to landform and geology in the Threemile Hill area resulting in very low prediction accuracy. Unit 186C (Vitrixerandic Haplocambids) is an aspect unit occurring on slopes greater than 15% and covers 3% of the total area. It was predicted with 43% accuracy. 45% of the pixels were misclassified as 140C. Where unit 186C does appear in the predicted map, it appears in the correct places. However, it is not carried all the way around scarp faces or down drainages to its actual full extent. Although unit 140C is officially mapped on 2 to 15% slopes, some areas mapped as 140C have much greater slopes and this likely carried over into the rule sets. Since 186C is mapped on the steeper sections at the boundaries of 140C, this misclassification is easily understood. The remaining units cover a small percentage of the total area, and will not be discussed. Many of the errors in the Threemile Hill predicted map can be easily explained and accounted for in subsequent predictions by adjusting map unit boundaries, running slope analyses to delineate certain aspect units, or aggregating similar units after predicting for a higher level of taxonomy (i.e. subgroup) as opposed to map units.

3.6.3. April 10, 2008 Predictive Map

A second predictive map was produced for the Threemile Hill quad on April 10, 2008. Some of the deficiencies in the October 12, 2007 predictive run were believed to be related to the lack of surficial geology information. During winter 2008, surficial geology maps of several quads in the MCSPSS were created based on bedrock geology map analysis and aerial photo interpretation (Noller, 2008), and this information was incorporated into April 10, 2008 the model. Individual layers were created to represent surficial age (e.g. Holocene, Pleistocene), surficial lithology (e.g. basalt, landslide deposits), and a combination of the two (e.g. Holocene basalt). All other independent variables were kept the same. The ranking of variable importance changed significantly relative to the October 12, 2007 predictive run. The fifteen most important predictor variables (all with an estimated importance of >10% as ranked by See5) were: iron oxide minerals (LandSat TM Band3/Band1 ratio), surficial age, average minimum temperature, elevation, mean annual precipitation, vegetation index (LandSat TM Band4/Band3 ratio), surficial lithology, January minimum temperature, July maximum temperature, diffuse radiation received (W-h/m^2), average maximum temperature, LandSat TM Band 4, bedrock geology, LandSat TM Band 6, and slope. Four soil map units from the predicted map that do not occur on the actual map (136B, 182C, 199B and 213) and one soil map unit from the actual map that does not occur on the predicted map (273) were excluded from the error-matrix analysis. All five of these units combined account for approximately one percent of the total mapped area within the Threemile Hill quad. Overall accuracy was not significantly different compared to the October 12, 2007 run. Figure 3.6-5 shows a comparison of the predicted and actual soil map units for the Threemile Hill quad. The

overall accuracy and Kappa coefficient for map units in the training area were 98.6% and 0.98, respectively. The overall accuracy and Kappa coefficient for map units in the area of interest, the Threemile Hill quad, were 74.5% and 0.68, respectively. Overall accuracies and Kappa coefficients were also calculated for the order, suborder, great group, and subgroup levels of map unit taxonomy (summarized in Table 3.6-1). These results initially seem to support the findings of Scull et al. (2005), who found that removing individual variables did not greatly impact map accuracy. Here, as opposed to removing variables from the analysis, additional variables were introduced that were thought to be of great importance for predicting soils in the area. Although these variables were ranked with high importance by the decision-tree classifier, accuracy did not improve relative to a previous predictive run that excluded these variables from consideration. However, comparing the October 12, 2007 and April 10, 2008 error images for the Threemile Hill area (see Figure 3.6-6) shows that the distribution of prediction errors varies. The surficial geology information results in more precise map unit boundary placement and less error near those boundaries.
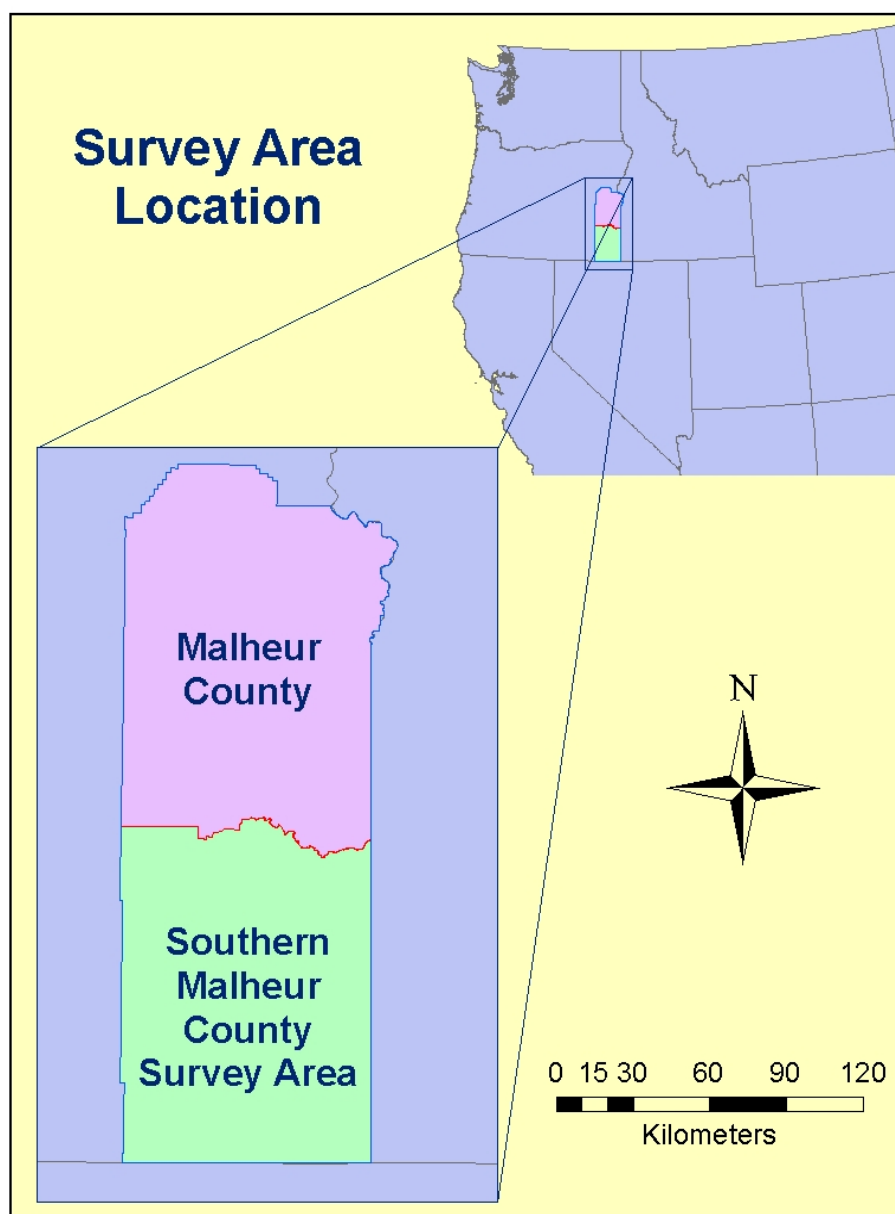
Figure 3.1-1: Malheur County, Southern Part Soil Survey Area Location
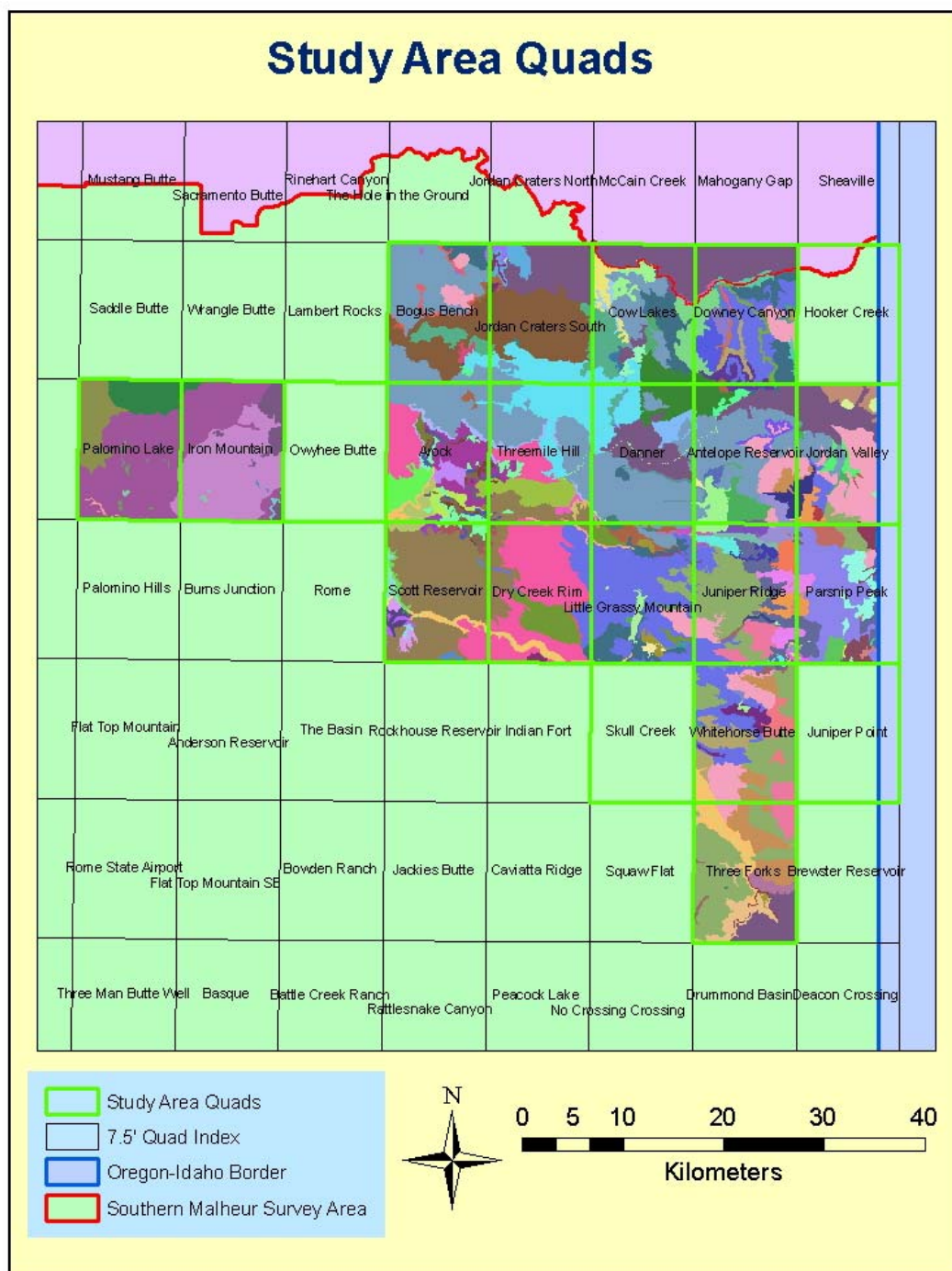
Figure 3.1-2: Study Area Quads
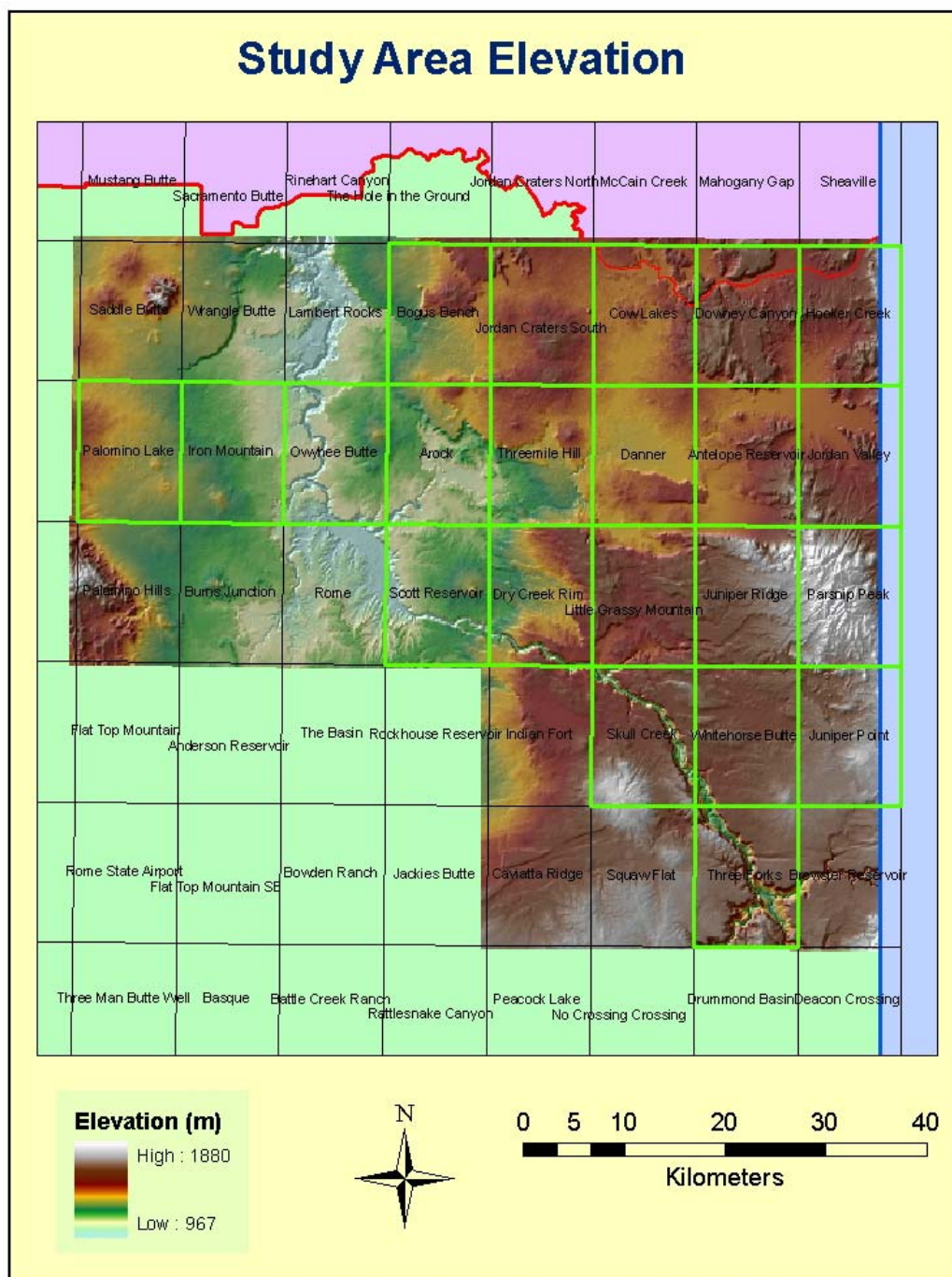
Figure 3.1-3: Study Area Mean Elevation
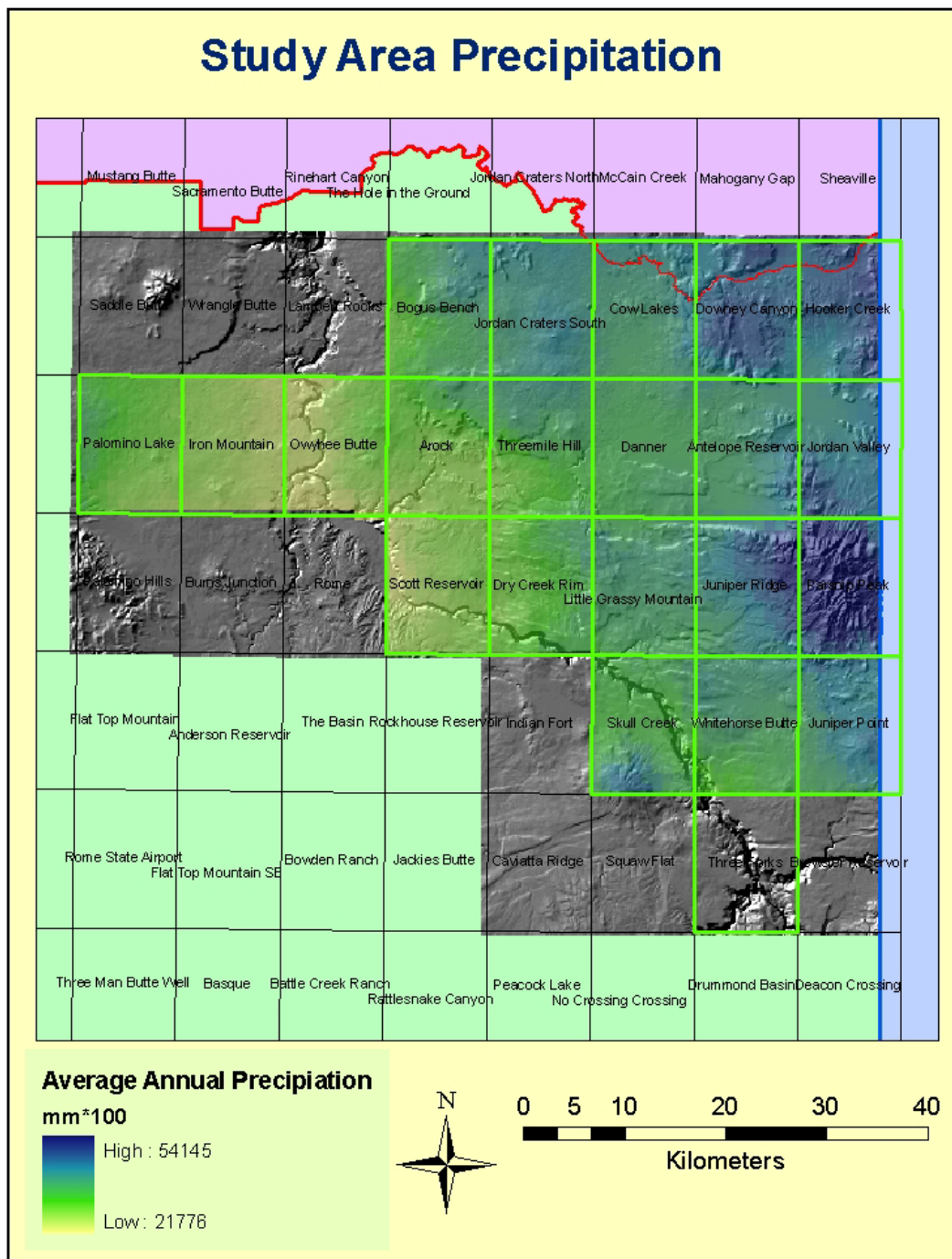
Figure 3.1-4: Study Area Mean Annual Precipitation (MAP)
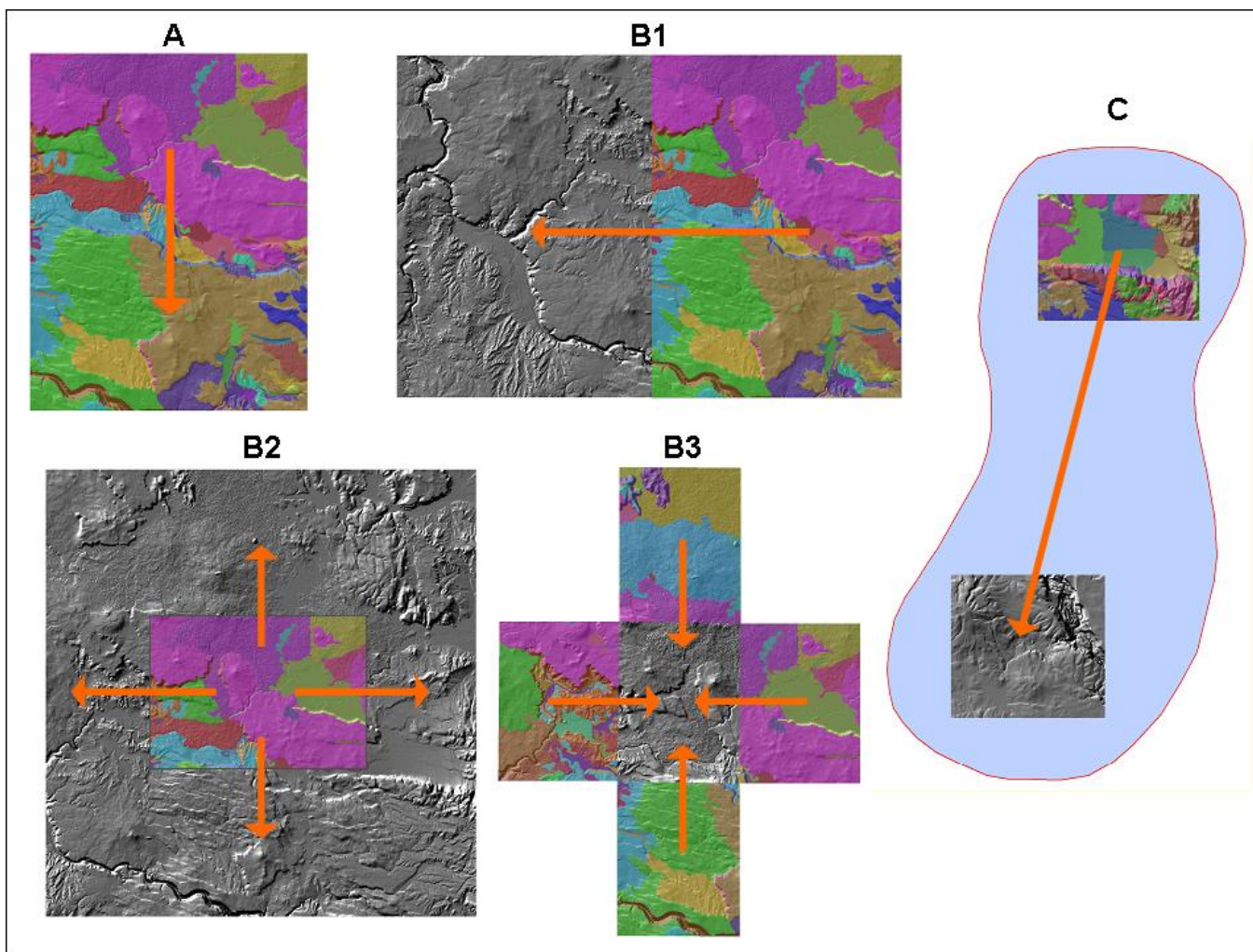
Figure 3.4-1: Approaches to Prediction
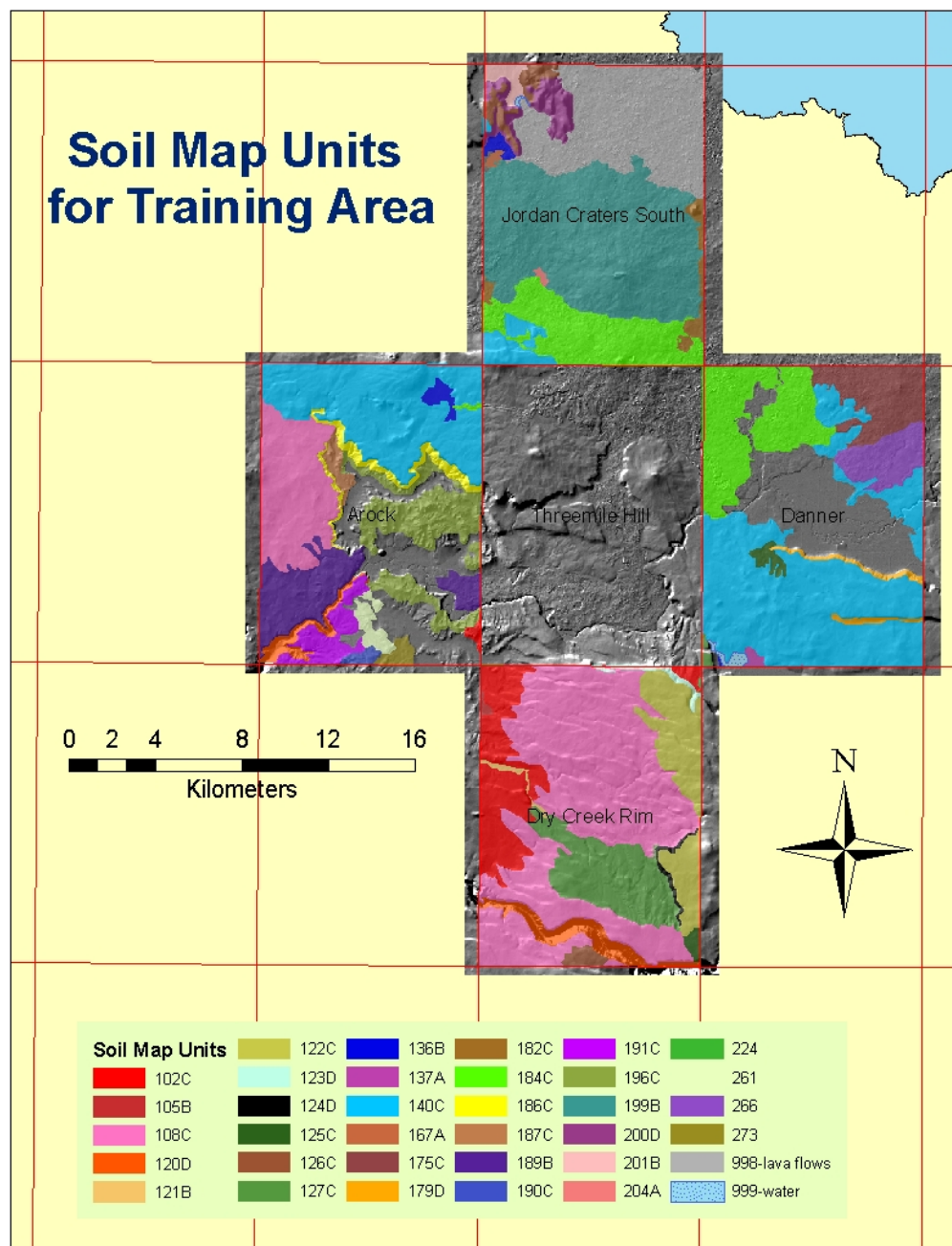
Figure 3.6-1: Soil Map Units for Training Area - October 12, 2007 Predictive Map
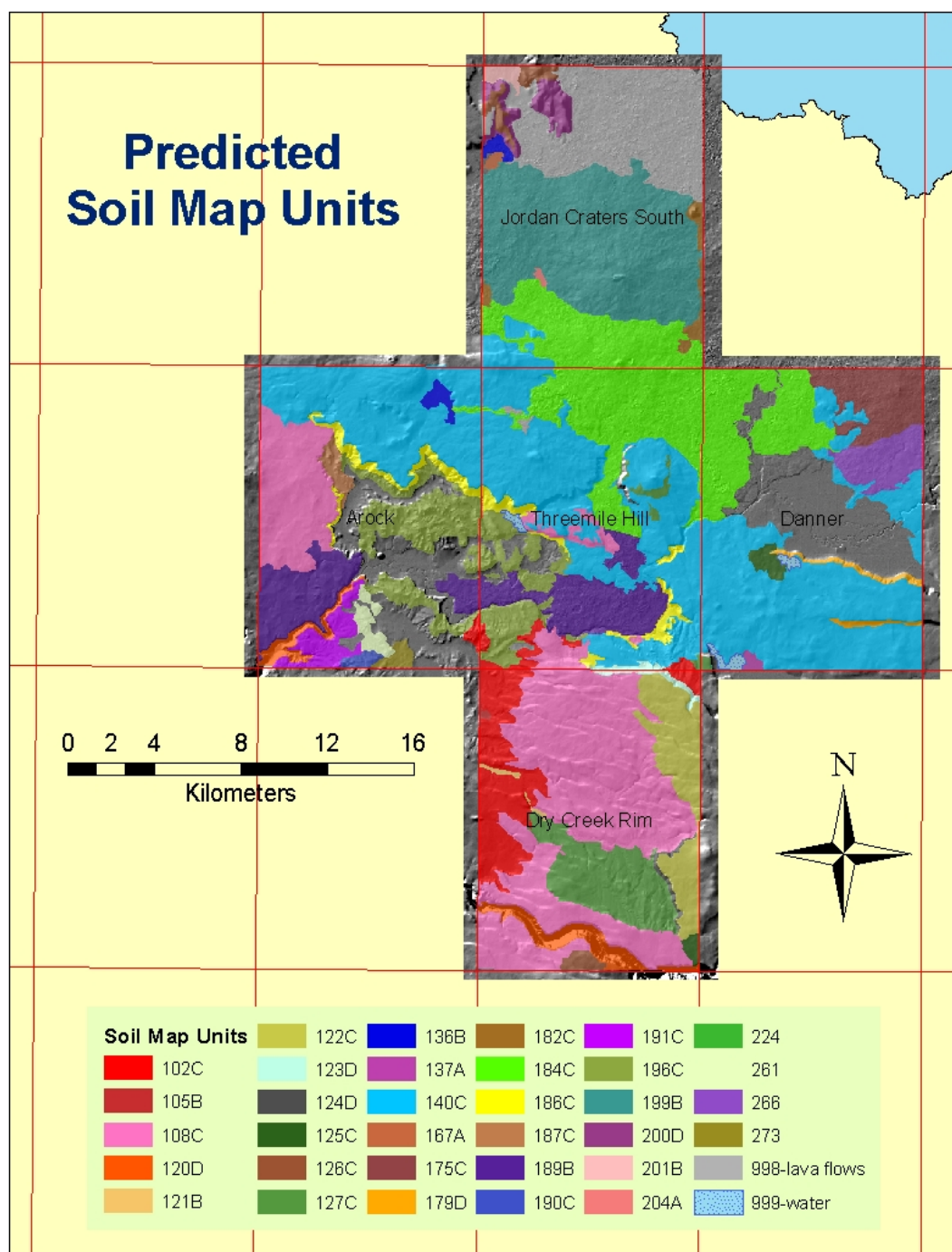
Figure 3.6-2: Predicted Soil Map Units for Whole Area – October 12, 2007 Predictive Map

Predicted vs. Actual
Map Units for Threemile Hill

Predicted

Actual

Predicted Map Units

■ 102C	■ 186C		
■ 108C	■ 189B		
■ 123D	■ 196C		
■ 127C	■ 998-lava flows		
■ 140C	▧ 999-water		
■ 184C			

N

Actual Map Units

■ 102C	■ 184C		
■ 108C	■ 186C		
■ 123D	■ 189B		
■ 127C	■ 196C		
■ 129A	■ 222		
■ 140C	■ 224		

0 1 2 4 6 8 Kilometers
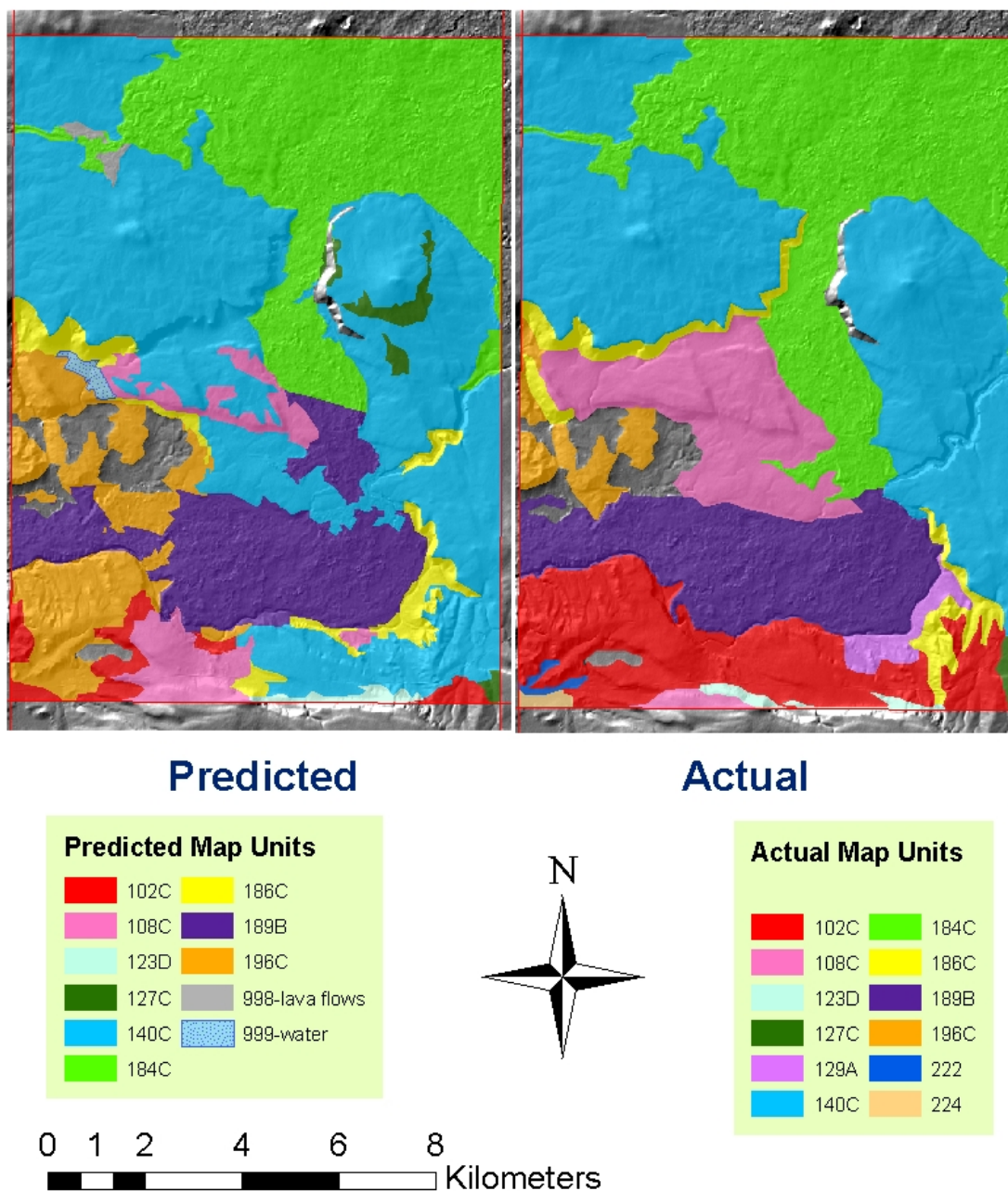
Figure 3.6-3: Comparison of Predicted and Actual Map Units for Threemile Hill – October 12, 2007 Predictive Map
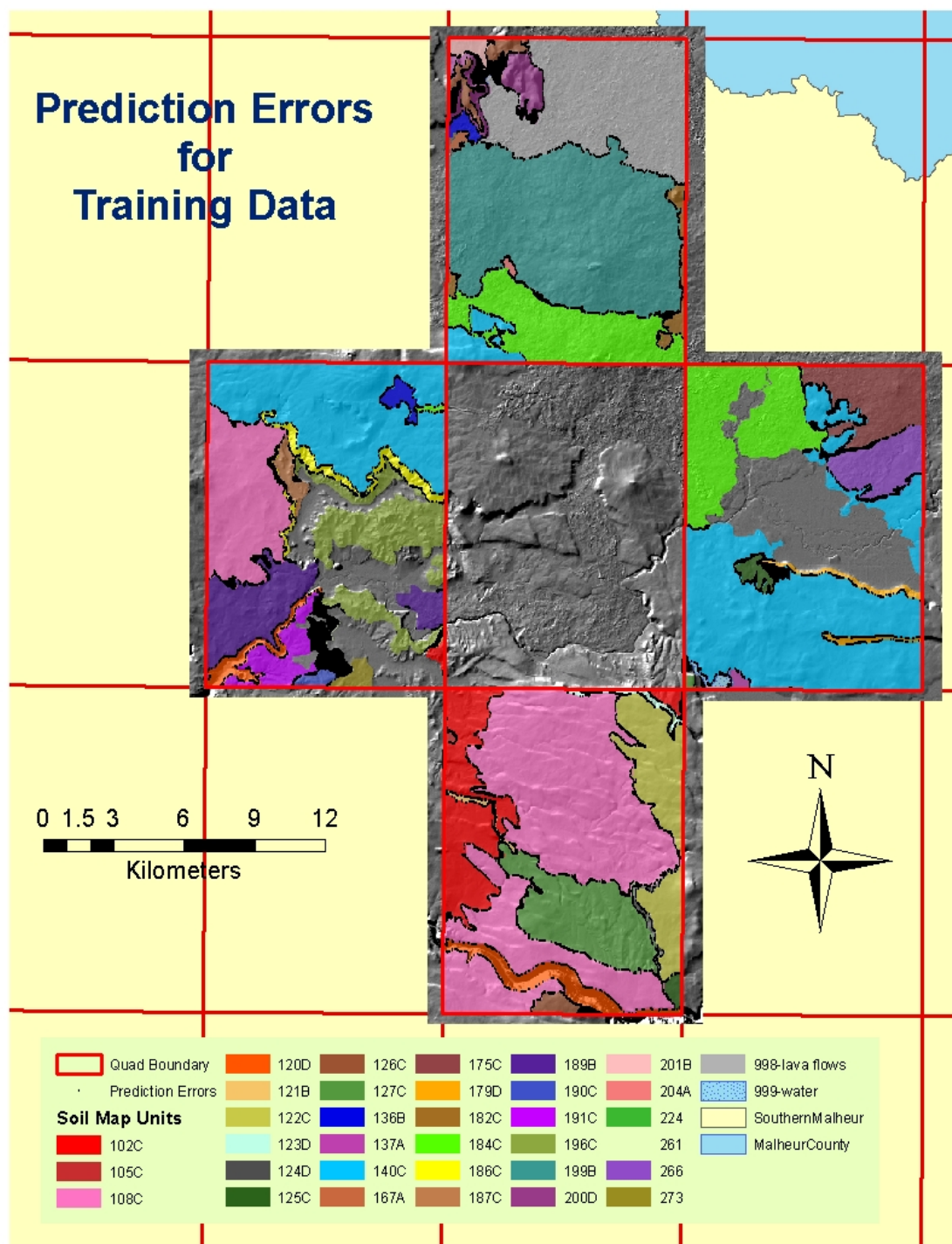
Figure 3.6-4: Prediction Errors for Training Data - October 12, 2007 Predictive Map
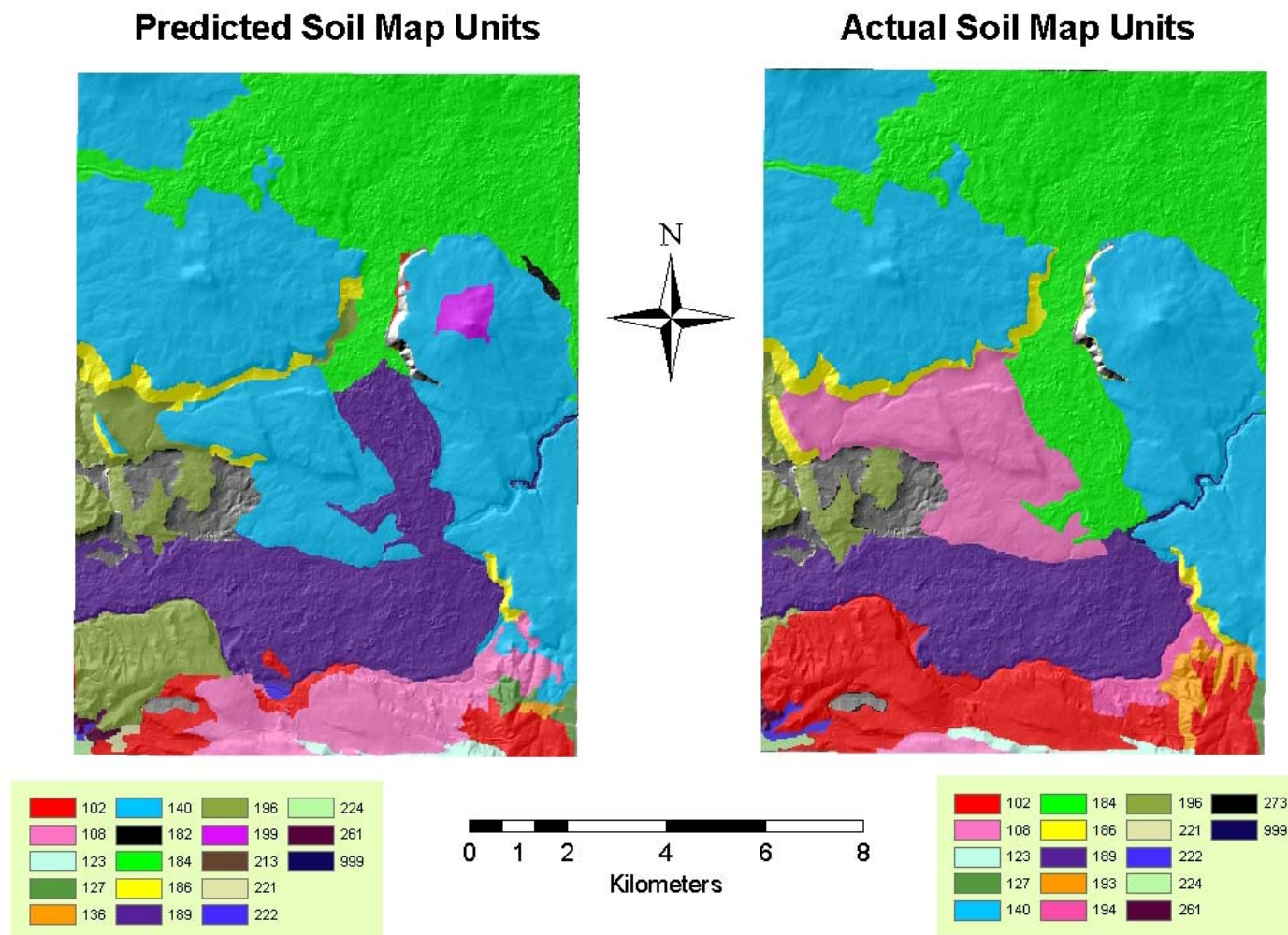
Figure 3.6-5: Comparison of Predicted and Actual Map Units for Threemile Hill – April 10, 2008 Predictive Map
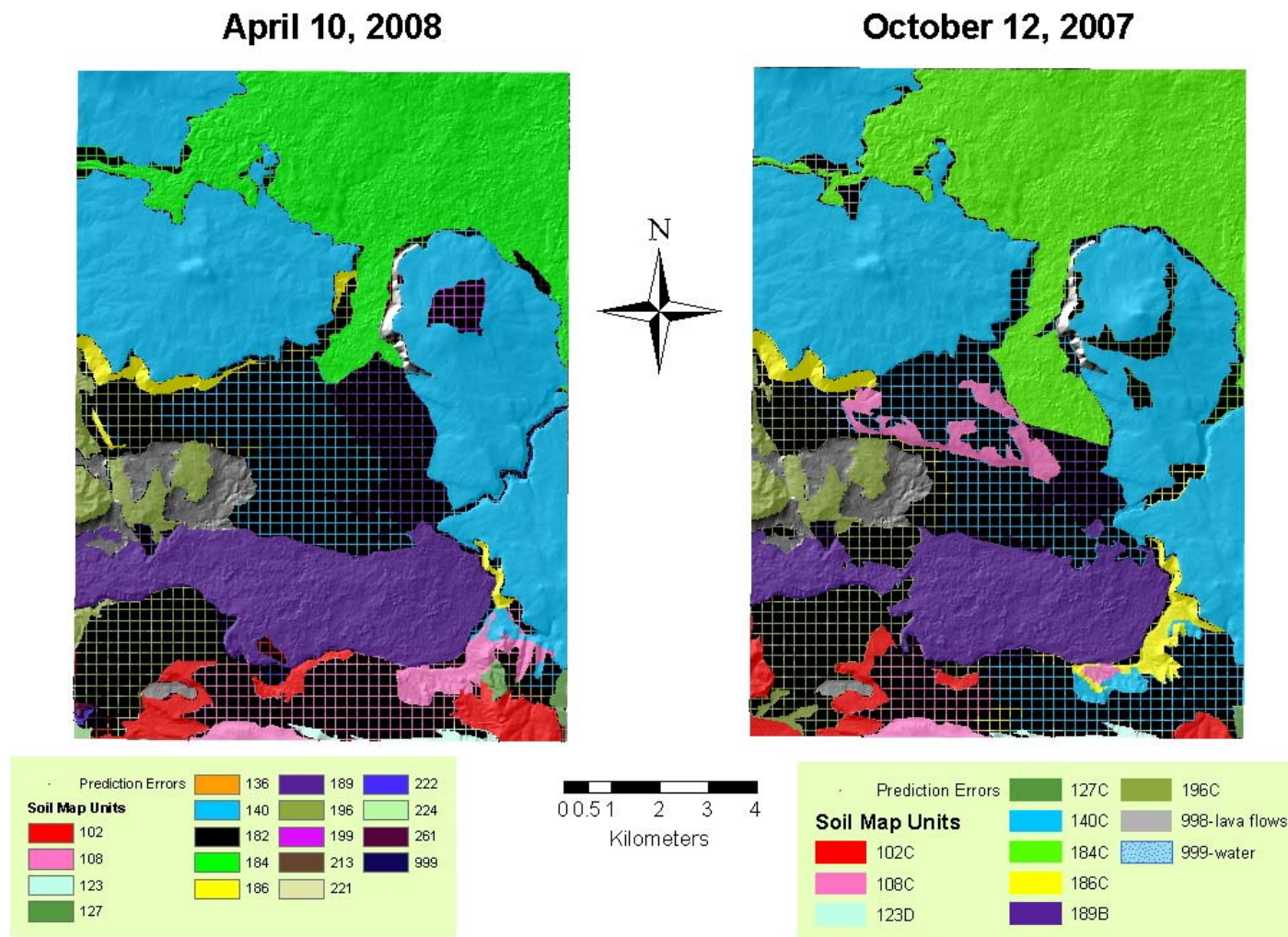
Figure 3.6-6: Comparison of Prediction Errors – April 10, 2008 vs. October 12, 2007 Maps

Table 3.1-1: Survey Area Land Ownership

Land Category	Acres
Non-Federal (Private and State)	450,459
Bureau of Land Management	2,603,820
Fort McDermitt Indian Reservation	18,829
Water	7,000
Total	**3,080,108**

Table 3.3-1: Summary of Independent Variables

Variable	Source	Resolution/ Scale	Data Type
Elevation	ifSAR Digital Terrain Model (DTM), Intermap Technologies, confidential data provided by NRCS	5m	continuous
Slope	Derived from ifSAR DTM using ArcGIS 9.2 Surface Analysis tools	5m	continuous
Aspect			continuous
Diffuse Solar Radiation (W-h/m2)	Derived from ifSAR DTM using ArcGIS 9.2 Solar Radiation tool	5m	continuous
Mean Annual Precipitation (1971-2000)	PRISM Group, Oregon State University (2007)	800 m	continuous
Minimum January Temperature (1971-2000)			continuous
Maximum July Temperature (1971-2000)			continuous
Average Minimum Temperature (1971-2000)			continuous
Average Maximum Temperature (1971-2000)			continuous
Bedrock Geology	Oregon Geologic Data Compilation (OGDC) v. 3 (Oregon Dept. of Geology and Mineral Industries, 2006)	varies	categorical
Gap Analysis Vegetation Cover	Kagan et al. (1999)	1:100,000	categorical
Historic (Pre-1938) Vegetation Cover	Oregon Natural Heritage Program (2002)	1:100,000	categorical
LandSat TM Bands 1-7	Landsat.org, Global Observatory for Ecoystem Services, Michigan State University	28.5 - 30 m	continuous
Vegetation Index	Derived from LandSat TM image, Band4/Band3	30	continuous
Iron Oxide Minerals	Derived from LandSat TM image, Band 3/Band1	30	continuous
Surficial Age and Surficial Lithology	Derived from surficial geology maps of Jordan Craters South, Arock, Threemile Hill, Danner and Dry Creek Rim quads (Noller, 2008)	1:24,000	categorical

**Several other independent variables were included in the DTA but were winnowed or ranked with importance < 10%. These include plan and profile surface curvature and total and direct solar radiation (derived from the ifSAR DTM); topographic position index (derived from the ifSAR DTM (Jenness, 2005); NDVI, TNDVI, and ferrous minerals (calculated from LandSat TM data); and brightness, greenness and wetness (derived from Tasseled Cap Transformation of LandSat TM data) (Kauth and Thomas, 1976).

Table 3.5-1: Prediction Approaches and Accuracy Summaries for Early Runs

Date	Approach	Map Products	Overall % Accuracy (from point data)			
			Order	Suborder	Great Group	Subgroup
June 16, 2007	B1	Downey Canyon, Antelope Reservoir, Jordan Valley	67	61	35	--
July 23, 2007	B2	Arock, Threemile Hill, Danner, Antelope Reservoir, Jordan Valley, Parsnip Peak, Whitehorse Butte	87	73	67	53
August 3, 2007	B2	Arock, Bogus Bench, Cow Lakes, Downey Canyon, Antelope Reservoir, Jordan Valley, Parsnip Peak	80	67	55	32
September 14, 2007	B2	Arock, Bogus Bench, Cow Lakes, Downey Canyon, Antelope Reservoir, Jordan Valley, Parsnip Peak	79	66	50	30
October 12, 2007	B3	Threemile Hill	91	83	83	35

Table 3.6-1: Accuracy Summaries for October 12, 2007 and April 10, 2007 Predictive Maps

Date	Approach	Confusion Matrix Analysis (Map Unit Level)			
		Training Area		Predicted Area	
		Overall % Accuracy	Kappa	Overall % Accuracy	Kappa
October 12, 2007	B3	98.5	0.98	74.0	0.67
April 10, 2008		98.6	0.98	74.5	0.68
order		–	–	99.3	0.92
suborder	B3	–	–	81.6	0.66
great group		–	–	80.6	0.74
subgroup		–	–	79.3	0.72

Table 3.6-2: Areal Summaries for Mapped Soils in Threemile Hill Area

Map Unit	Acres	% Area	Mean Delineation Area (Std. Dev.*)	Cumulative % Area
140C	36,317	22.8	5,188 (7,449)	22.8
108C	27,406	17.2	4,568 (3,934)	39.9
184C	19,611	12.3	6,537 (7,689)	52.2
199B	15,096	9.5	15,096 (-)	61.7
998	9,278	5.8	9,278 (-)	67.5
102C	8,890	5.6	2,963 (2,249)	73.0
189B	8,053	5.0	4,026 (887)	78.1
196C	5,380	3.4	897 (984)	81.5
127C	4,760	3.0	1,190 (1,936)	84.4
122C	4,156	2.6	4,156 (-)	87.0
175C	3,051	1.9	3,051 (-)	89.0
120D	2,288	1.4	1,144 (400)	90.4
186C	2,288	1.4	572 (538)	91.8
266B	1,995	1.2	1,995 (-)	93.1
200D	1,505	0.9	752 (198)	94.0
191C	1,501	0.9	500 (561)	95.0
126C	1,182	0.7	394 (20)	95.7
182C	779	0.5	260 (114)	96.2
261C	634	0.4	317 (215)	96.6
136B	613	0.4	153 (152)	97.0
179D	594	0.4	297 (74)	97.3
125C	560	0.4	280 (67)	97.7
187C	495	0.3	495 (-)	98.0
123D	484	0.3	242 (109)	98.3
201B	480	0.3	480 (-)	98.6
129A	407	0.3	407 (-)	98.9
273	356	0.2	356 (-)	99.1
124D	249	0.2	249 (-)	99.2
190C	246	0.2	246 (-)	99.4
999	204	0.1	34 (64)	99.5
121B	187	0.1	187 (-)	99.6
137A	151	0.1	50 (70)	99.7
167A	106	0.1	106 (-)	99.8
105B	90	0.1	90 (-)	99.9
222	84	0.1	84 (-)	99.9
224	82	0.1	41 (41)	100.0
204	68	0.0	68 (-)	100.0
Totals	159,625	100.0		

* Standard Deviation calculated as $\sqrt{\dfrac{\sum (x-\bar{x})^2}{n}}$

Table 3.6-3: Map Unit Descriptions for Threemile Hill Area

Map Unit	Name	Component	%	Classification	% Total Area
Actual Soil Map Units for Threemile Hill Area   **(Jordan Craters South, Arock, Danner, Dry Creek Rim, and Threemile Hill)**
(16 of 37 Classes, 95% of Total Area)
140C	Midraw very stony ashy loam	Midraw	85	Clayey, smectitic, mesic, shallow Abruptic Xeric Argidurids	22.8
108C	Sandhollow ashy sandy loam	Sandhollow	85	Loamy, mixed, superactive, mesic shallow Vitrixerandic Argidurids	17.2
184C	Rock outcrop-Clarksbutte complex	Rock outcrop	65	rock outcrop	12.3
		Clarksbutte	25	Fine-loamy, mixed, superactive, mesic Vitrixerandic Argidurids	
199B	Clarksbutte-Rock outcrop complex	Clarksbutte	80	Fine-loamy, mixed, superactive, mesic Vitrixerandic Argidurids	9.5
		Rock outcrop	15	rock outcrop	
998	Lava Flows	Lava	98	lava flows	5.8
102C	Muni gravelly ashy loam	Muni	90	Loamy, mixed, superactive, mesic shallow Haploxeralfic Argidurids	5.6
189B	Clarksbutte-Rock outcrop-Rimview complex	Clarksbutte	40	Fine-loamy, mixed, superactive, mesic Vitrixerandic Argidurids	5.0
		Rock outcrop	30	rock outcrop	
		Rimview	15	Fine-loamy, mixed, superactive, mesic Typic Haplargids	
196C	Sagehill-Muni dry complex	Sagehill	60	Coarse-loamy, mixed, superactive, mesic Durinodic Haplocambids	3.4
		Muni	25	Fine-loamy, mixed, superactive, mesic Typic Argidurids	
127C	Toney ashy silt loam	Toney	85	Fine, smectitic, mesic Xeric Paleargids	3.0

Table 3.6-3 Continued: Map Unit Descriptions for Threemile Hill Area

Map Unit	Name	Component	%	Classification	% Total Area
Actual Soil Map Units for Threemile Hill Area (Jordan Craters South, Arock, Danner, Dry Creek Rim, and Threemile Hill)					
(16 of 37 Classes, 95% of Total Area) Continued					
122C	Foleylake-Martinson complex	Foleylake	60	Fine, smectitic, frigid Abruptic Xeric Argidurids	2.6
		Martinson	25	Fine, smectitic, frigid Vitritorrandic Argixerolls	
175C	Cocklebur-Wisher-Rock outcrop complex	Cocklebur	40	Loamy, mixed, superactive, mesic, shallow Vitrixerandic Haplargids	1.9
		Wisher	30	Fine-silty, mixed superactive, mesic Vitritorrandic Argixerolls	
		Rock outcrop	20	rock outcrop	
120D	Rock outcrop-Rubbleland complex	Rock outcrop	80	rock outcrop	1.4
		Rubbleland	15	rubbleland	
186C	Felcher loam	Felcher	85	Fine-loamy, mixed, superactive, mesic Vitrixerandic Haplocambids	1.4
266B	Skinnerpit-Lava flows-Cheatroad complex	Skinnerpit	45	Coarse-loamy, mixed, superactive, mesic Xeric Haplocambids	1.3
		Lava	25	lava flows	
		Cheatroad	15	Coarse-loamy, mixed, superactive, mesic Xeric Haplodurids	
200D	Zymans-like-Barbermill complex	Zymans-like	50	Fine, smectitic, frigid Vitritorrandic Argixerolls	0.9
		Barbermill	35	Clayey, smectitic, mesic, shallow Aridic Argixerolls	
191C	Muni dry-Drewsey Dry complex	Muni, dry	70	Loamy, mixed, superactive, mesic shallow Vitrixerandic Argidurids	0.9
		Drewsey, dry	20	Fine-loamy, mixed, superactive, mesic Vitrixerandic Haplargids	

4. Predictive Mapping Implementation on the Malheur County, Oregon Initial Soil Survey

4.1. Interactions with NRCS and BLM Personnel

Although predictive mapping work was previously completed in southwestern Malheur County (Elnaggar, 2007), the areas examined in that study did not intersect with areas actively being mapped by NRCS during the 2006-2008 field seasons. Initial crew contact and field investigations for this study took place during the summer of 2006. Ride-alongs with NRCS soil scientists and the BLM ESI team leader during regular mapping activities were conducted to become familiar with the soils, topography and vegetation communities in the survey area. At that time, crew members were introduced to the proposed methodology and theory behind using decision-tree classifier software for predictive soils mapping for soil survey. Initial crew reactions varied, but skepticism and suspicion were evident in many of the team members' reactions. There were, however, survey team members who were quite enthusiastic about the use of predictive mapping. Despite some negative perceptions, all of the survey team members were accommodating and generous with their knowledge. Field excursions with the ESI team leader and soil scientists yielded valuable information regarding the soil-landscape models and relationships that were driving map unit differentiation. These observations aided in the selection of independent variables for the decision-tree analysis.

Causative factors behind negative perceptions varied between the NRCS and BLM team members. Many of the BLM ESI surveyors felt alienated by the process for the following two reasons: 1) since the focus of this project is on soils mapping, they believe that the ESI goals of this BLM-driven survey are being ignored, and 2) if ESI information

were incorporated into the prediction, they do not believe that a model could accurately predict soils-ESI map units containing so much complex information. Most of the NRCS soil scientists were less direct in stating their concerns, but the following two issues seemed of greatest importance: 1) fear of replacement of expert knowledge by a computer program that attempts to simulate the surveyor's thought process, and 2) fear that a technology barrier exists (in user knowledge, software usability, equipment availability, or compatibility with field procedures). ESI surveyors and soil scientists were equally concerned about how validation data would be obtained and map accuracy analyzed. Mapping accuracy and usability of soils maps is highly dependent upon consistent landscape interpretation and taxonomy application among the different scientists working on a team. The decision to use the survey team's pedon descriptions and map unit lines for validation data resulted in the team members having greater confidence in the accuracy statements.

The earliest predictive runs that yielded low accuracy (using Owyhee County survey data as training data to predict adjacent quads) fostered negative views of the predictive mapping process. However, two meetings (one in the summer and one in the fall of 2007) were instrumental in influencing perception and building confidence in potential usefulness. The Summer 2007 meeting, attended by Bob Graham, State Soil Conservationist, GIS analysts from the Vale, Oregon BLM office, and the members of the Malheur County survey crew, allowed for a brief formal introduction to the predictive mapping project (variable selection, decision tree analysis, and map generation) and an overview of work planned for the 2007 field season. Field work for the 2007 season was performed in concert with Alina Rice, Soil Survey Project Leader, who provided pedon

descriptions for accuracy analysis and numerous suggestions for model improvement.

The Fall 2007 meeting was held during the initial field review, and was attended by Tom McKay, MO3 Soil Data Quality Specialist, NRCS; Paul Blackburn, MLRA Leader, NRCS; Greg Kuehl, Snake River Area Basin Team Leader, NRCS; Bob Gillaspy, State Range Conservationist, NRCS; Ed Peterson, District Conservationist, NRCS; and the regular Malheur County survey crew. A summary of completed predictive maps along with accuracy analyses was presented in the field area. Sources of potential error and methodology evolution were explained. The final predictive map for the Threemile Hill area, which had substantially improved accuracy over previous runs, was presented along with a proposed checkerboard mapping structure (prediction mode B3) to facilitate field work planning. Reactions were generally positive, and serious discussion of using the process for generating pre-maps was initiated. At this time, it was decided that a soil scientist from the Malheur County survey would attend predictive-mapping training at Oregon State University. A description of that training, its goals, and the results is provided below.

4.2 Predictive-Mapping Training Session

In April 2008, Shanna Bernal-Fields, Student Soil Conservation Technician assigned to the Malheur County, Southern Part soil survey, attended a three-day predictive mapping training session at Oregon State University. The goal of the training session was to introduce her to the mapping process and simultaneously produce pre-maps for four quads to be mapped in the 2008 field season. The session was split into three parts: 1) acquiring and preparing digital data sets, 2) analyzing training data with decision tree

analysis software, and 3) generating and editing predictive maps. Conveying the information and generating the maps was a success, but discussion with Ms. Bernal-Fields over the course of the session revealed some major barriers to successful implementation at the NRCS office. The computing resources currently available are not sufficient to carry out the computationally-intensive decision-tree analyses. Time constraints are also limiting. Data preparation is a time-intensive process. A detail or permanent employee with duties specific to predictive mapping would be needed, and GIS support personnel for data acquisition and accuracy analysis might also be necessary. Finally, the learning curve may be large for some employees. A solid background in GIS and geospatial analysis is necessary to troubleshoot problems and ensure data integrity through multiple manipulations of data. Knowledge of field-mapping procedures and of the specific soil-forming environments in the survey area are also required.

After having some time to reflect on the predictive mapping training session, Ms. Bernal-Fields provided the following feedback on the experience (personal e-mail communication, May 2, 2008):

> "After the three-day course in predictive modeling I feel I have a fairly strong understanding of the basic principles behind this tool and a healthy view of its limitations. If our agency purchased all the required software and a computer that could produce the predictive maps in a timely manner I could learn to operate them with a bit more training. I speak specifically of *operating* the model because of the large amount of time that goes into the data acquisition and preparation. I would not have time to do it with my current workload nor would I feel comfortable performing the data transformations with my current GIS ability levels. If GIS support was provided and I was given a packet of the needed data layers I think I could manage running it in the office very soon. You provided some excellent documents last week, but I am more comfortable with the format of our agency's technical notes. Not all field scientists are GIS savvy, but this tool could be very handy if it arrived ready to go with technical guides.

I think this tool could be used for pre-mapping, as well as training new soil scientists in a given area. The predictive maps would give them a higher starting point for mapping--they could ground truth and modify as needed and gain confidence and familiarity with the local soils. Perhaps soil survey updates could even use it to take larger orders down as needed. If the accuracy of the model were sufficient, it could be used as a check of older mapping to gauge quality. Perhaps some GIS specialists could analyze the rule sets generated by the model and translate them into general rules of thumb concerning important environmental variables and meaningful breaks within those variables for the particular survey area."

Examining Ms. Bernal-Fields' comments in context of the TAM suggests that

perceived usefulness is not a major barrier, and that future efforts should focus on the

perceived ease of use. Specific consideration should be paid to the modes of discussion

and presentation. Presenting the procedures and supporting documentation in a familiar

format (as an agency technical note) will foster confidence and make the process seem

more approachable and understandable for the field scientists.

4.3 Proposed System for Predictive Mapping Implementation

Technology acceptance research has shown that adequate support services are

critical in early stages of adoption of new tools for performing work. An administrative

support position within the agency would conduct training sessions and serve as the point

of contact for issue resolution. For successful predictive mapping implementation on an

individual survey, certain personnel would be needed: 1) a soil scientist with field mapping

experience in the survey area and designated predictive/digital mapping job

responsibilities, and 2) a GIS analyst who would provide data acquisition, data preparation

and map digitization support. Background in GIS theory and basic proficiency in a

commercial GIS software package (including geospatial analysis and digitizing tools)

would be assumed. In the proposed system, the soil scientist, GIS analyst and soil survey

project leader would attend a week-long winter training session on predictive mapping.

Topics to be covered would include digital data sources, selection of appropriate

environmental variables, data accuracy and integrity analysis (including outlier detection

and descriptive statistics), issues of scale and raster cell size selection, preprocessing of

digital data sets, creating sampling grids for rule set construction, use of decision tree

analysis software (including theoretical background and selection of appropriate pruning

and boosting options), CART analysis for predictive map generation, and predictive map

preparation (including selection of minimum map unit size and level of classification to be

used for pre-maps). Full documentation of procedures, software and hardware

requirements, and agency map and data accuracy standards should be compiled in one

publication (an NRCS technical note) for distribution at the training session. During the

training session, the team would use field mapping from previous seasons (along with

other spatial data sets representing independent variables) to produce pre-maps for the

following field season.

Subsequent fieldwork planning according to the quad checkerboard approach

described earlier would provide the most efficient strategy for predictive mapping. Once

three to four adjacent quads have been mapped, that mapping can be used to produce a

predictive map for the quad of interest. With this approach, pre-maps can be produced

from field mapping within the same field season by digitizing the tentative quad maps as

they are produced. The soil survey project leader decides what digital data sets are needed

and requests those data (along with tentative digitized map lines) from the GIS analyst.

The designated soil scientist then produces the predictive maps under the supervision of the

soil survey project leader. As mapping progresses, the survey project leader and other field

scientists will find that their mental soil-landscape models evolve and that different variables serve as the best predictors of soil types in different areas. As a result, the predictive mapping process will also evolve with time. Selection of independent variable data sets and manipulation of those data sets will change. Expert knowledge of field conditions and what environmental variables reflect those conditions is essential to successful model construction. The analyst has a tremendous amount of control over the map characteristics and quality through optimal data selection and preparation.

It is recommended that predictive maps be prepared at the soil map unit level and then recoded to represent soil subgroups or great groups. The subgroup level of classification preserves the most information and retains accuracy suitable for pre-maps (~80% in this study). The soil survey project leader and ESI team leader could then use the existing map unit legend to develop and document decision trees for individual soil subgroups. Once the predicted subgroup is verified, the decision tree would be used to help select the specific map units to be considered. Figure 4.3-1 shows an example of a decision tree that could be used to differentiate Argidurids, a great group, into soil map units documented for the survey area.
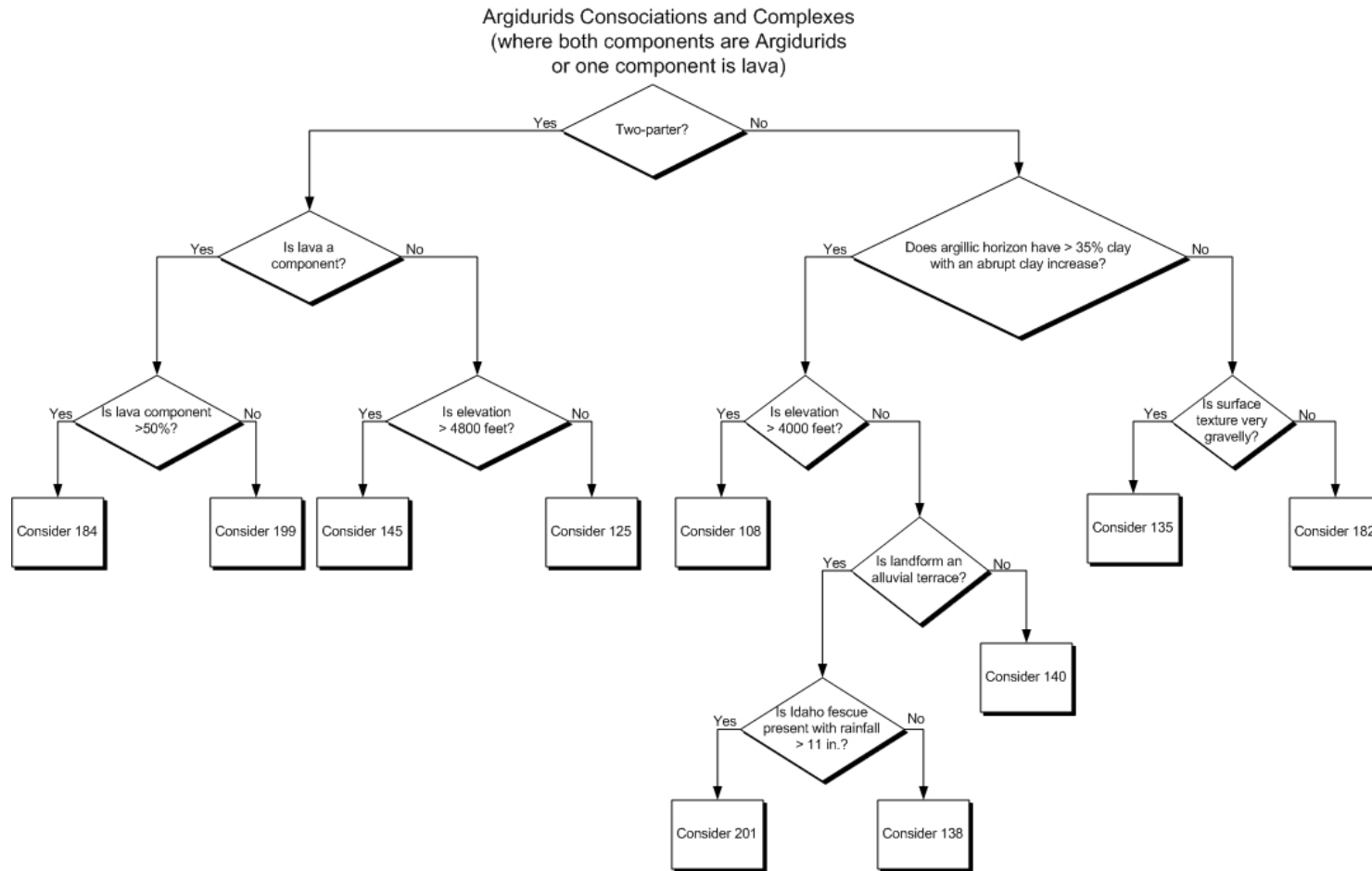
Figure 4.3-1: Decision Tree for Differentiating Argidurids into Soil Map Units

5. Conclusions

Soil-landscape relationships in the study area can be successfully extracted and quantified using decision-tree analysis. These relationships can then be applied to an area of interest to produce predictive soils maps. The soil scientists' landscape model for this area within the MCSPSS, principally involving vegetation, elevation, climatic zones, and landform/geology, is faithfully reproduced by the DTA. The suite of environmental variables that most accurately characterize those attributes often varies from area to area. The soil survey construct and mapping approach on the MCSPSS (combining ESI and soils information in unit delineations) results in climatic variables always being ranked with high importance. ESI units are based largely on precipitation amounts and seasonal temperature variations, so some rules that use climatic variables are predicting for the ESI component of a delineation as opposed to the soils component. Contrary to the soil scientist's common landscape model, some environmental variables, such as slope and aspect, are not shown by the DTA to be directly important in predicting for soil type. These variables may be indirectly accounted for in other datasets (precipitation data sets likely account for the influence of aspect) or may not be important in predicting soil type (slope is not an important predictor of soil type in this area). Because soils and soil-landscape relationships vary over short distances within the study area, rule sets built in one area may not be applicable in a geographically distant area. Carefully constraining the study area and selecting the most appropriate training data are critical for successful prediction. Error-matrix analysis for mapped areas shows where misclassification occurs, identifies to which classes misclassified cases are being assigned, and allows the analyst to

identify data sets needed to enhance characterization. Predictive soils mapping using DTA has a tremendous potential to improve efficiency and accuracy of a soil survey by taking advantage of the wealth of digital data that characterizes the environment. This approach also corrects problems inherent in conventional soils mapping that are related to 1) subjectivity, by forcing objective delineation, and 2) lack of soil-landscape model documentation, by producing expressive logs and charts.

Digital mapping techniques described here should be used to produce pre-maps for an active soil survey. Strategic planning of the mapping schedule (e.g. checkerboard mapping) and digitizing preliminary soils maps, as they become available, allows rapid turnaround on pre-maps during the same field season and increases efficiency by exactly targeting the number and locations of delineations that must be visited. Field scientists can concentrate more heavily on delineations of map units that are predicted with low confidence and make quicker verification stops at delineations of map units predicted with high confidence. Working closely with field scientists to fully document all landscape models and rubrics used to make field delineations will facilitate model improvement and may highlight existing digital data deficiencies.

Personnel, computer hardware, and software constraints currently prevent successful implementation of the procedure discussed in this thesis for the MCSPSS. Adding a field scientist with dedicated predictive soils mapping responsibilities, providing GIS support (perhaps from a central location supporting several survey crews), implementing the training procedure proposed herein, and investing a small sum in computer equipment would provide the capability. Fully documenting the process as an NRCS technical note and actively involving support personnel in early stages of adoption

will ensure that needs are met on the crew and individual levels. Many soil and range scientists within the NRCS and BLM are wary of (or even fear) adopting a computer modeling tool for field mapping. The agencies must ensure that the required tools are accessible and that the process is understandable to everyone involved. The effort associated with adoption must not be too great. Ultimately, control over the product (the map) must rest with the field crew. It is important to emphasize that while the DTA model attempts to replicate the soil scientist's soil landscape model, it is not a replacement for expert knowledge. Quite to the contrary, the field scientist's expert knowledge is needed at every step to evaluate and manipulate data, assess output, and make changes for model development. This codification of the field scientist's experienced and complex working model of the soil landscape ensures that it is available in full detail for generations to come. The well-trained and experienced field scientist will never be obsolete.

References

Alexander, E. B. & Nettleton, W. D. (1977). Post-Mazama Natrargids in Dixie Valley, Nevada. *Soil Science Society of America Journal*, *41*, 1210-1212.

Anderson, K., Wells, S., & Graham, R. (2002). Pedogenesis of Vesicular Horizons, Cima Volcanic Field, Mojave Desert, California. *Soil Science Society of America Journal*, *66*, 878-887.

Ashwill, M. (1983). Seven fossil floras in the rain shadow of the Cascade Mountains, Oregon. *Oregon Geology*, *45*(10), 107-111.

Belnap, J. (2003). The world at your feet: desert biological soil crusts. *Frontiers in Ecology and the Environment*, *1*(4), 181-189.

Bergeron, F., Rivard, S., & De Serre, L. (1990). Investigating the Support Role of the Information Center. *MIS Quarterly*, *14*(3), 247-259.

Bestland, E. A. (1997). Alluvial Terraces and Paleosols as Indicators of Early Oligocene Climate Change (John Day Formation, Oregon). *Journal of Sedimentary Research*, *67*(5), 840-855.

Bestland, E. A., Retallack, G. J., & Swisher III, C. C. (1997). Stepwise Climate Change Recorded in Eocene-Oligocene Paleosol Sequences from Central Oregon. *The Journal of Geology*, *105*, 153-172.

Bishop, E. M. (2003). *In Search of Ancient Oregon: A Geological and Natural History.* Portland, OR: Timber Press, Inc. 298 pp.

Blank, R. R., Cochran, B., & Fosberg, M. A. (1998). Duripans of Southwestern Idaho: Polygenesis during the Quaternary Deduced through Micromorphology. *Soil Science Society of America Journal*, *62*, 701-709.

Bondre, N.R. (2006). *Field and Geochemical Investigation of Basaltic Magmatism from the Western United States and Western India.* (Doctoral dissertation, Miami University, 2006). Retrieved from http://www.ohiolink.edu/etd/view.cgi?acc_num=miami1164916380. 263 pp.

Blank, R. R. & Fosberg, M. A. (1991). Duripans of Idaho, U.S.A: in situ alteration of eolian dust (loess) to an opal-A/X-ray amorphous phase. *Geoderma*, *48*, 131-149.

Bradbury, J. P., Colman, S. M., & Dean, W. E. (2004). Limnological and climatic environments at Upper Klamath Lake, Oregon during the past 45,000 years. *Journal of Paleolimnology*, *31*, 167-188.

Brandon, A. D. & Goles, G. G. (1988). A Miocene subcontinental plume in the Pacific Northwest: Geochemical evidence. *Earth and Planetary Science Letters*, *88*, 273-283.

Brown de Colstoun, E. C., Story, M. H., Thompson, C., Commisso, K., Smith, T. G., & Irons, J. R. (2003). National Park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. *Remote Sensing of Environment*, *85*, 316-327.

Brown, S. A. & Venkatesh, V. (2003). Bringing Non-Adopters Along: The Challenge Facing the PC Industry. *Communications of the Association of Computing Machinery*, *46*(4), 76-80.

Brueseke, M. E., Heizler, M. T., Hart, W. K., & Mertzman, S. (2007). Distribution and Geochronology of Oregon Plateau (U.S.A.) flood basalt volcanism: The Steens Basalt revisited. *Journal of Volcanology and Geothermal Research*, *161*, 187-214.

Bui, E. N., Loughhead, A., & Corner, R. (1999). Extracting soil-landscape rules from previous soil surveys. *Australian Journal of Soil Research*, *37*, 495-508.

Bui, E. N. & Moran, C. J. (2003). A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray-Darling basin of Australia. *Geoderma*, *111*, 21-44.

Camp, V. E. (1995). Mid-Miocene propagation of the Yellowstone mantle plume head beneath the Columbia River basalt source region. *Geology*, *23*, 435-438.

Camp, V. E., Ross, M. E., & Hanson, W. E. (2003). Genesis of flood basalts and Basin and Range volcanic rocks from Steens Mountain to the Malheur River Gorge, Oregon. *GSA Bulletin*, *115*(1), 105-128.

Carter, D. T., Ely, L. L., O'Connor, J. E., & Fenton, C. R. (2006). Late Pleistocene outburst flooding from pluvial Lake Alvord into the Owyhee River, Oregon. *Geomorphology*, *75*, 346-367.

Chadwick, O. A. & Davis, J. O. (1990). Soil-forming intervals caused by eolian sediment pulses in the Lahontan Basin, northwestern Nevada. *Geology*, *18*, 243-246.

Chadwick, O. A., Hendricks, D. M., & Nettleton, W. D. (1987a). Silica in Duric Soils: I. A Depositional Model. *Soil Science Society of America Journal*, *51*, 975-982.

Chadwick, O. A., Hendricks, D. M., & Nettleton, W. D. (1987b). Silica in Duric Soils: II. Mineralogy. *Soil Science Society of America Journal*, *51*, 982-985.

Chadwick, O. A., Hendricks, D. M., & Nettleton, W. D. (1989). Silicification of Holocene soils in northern Monitor Valley, Nevada. *Soil Science Society of America Journal*, *53*, 158-164.

Chadwick, O. A., Nettleton, W. D., & Staidl, G. J. (1995). Soil polygenesis as a functionof Quaternary climate change, northern Great Basin, USA. *Geoderma, 68*, 1-26.

Chaney, R. W. (1938). The Deschutes flora of eastern Oregon. *Publications of the Carnegie Institution of Washington*, *476*, 187-216.

Chaney, R. W. & Axelrod, D. I. (1959). Miocene floras of the Columbia Plateau. *Carnegie Institute of Washington Publication*, *617*, 237.

Chitwood, L. A. (1994). Inflated basaltic lava--examples of processes and landforms from central and southeast Oregon. *Oregon Geology*, *56*(1), 11-21.

Cohen, A. S., Palacios-Fest, M. R., Negrini, R. M., Wigand, P. E., & Erbes, D. B. (2000). A paleoclimate record for the past 250,000 years from Summer Lake, Oregon, USA: II. Sedimentology, paleontology and geochemistry. *Journal of Paleolimnology*, *24*, 151-182.

Cohen, J. (1960). A coefficient of agreement at nominal scales. *Educational and Psychological Measurement*, *20,* 37-46.

Colman, S. M., Bradbury, J. P., & Rosenbaum, J. G. (2004). Paleolimnology and paleoclimate studies in Upper Klamath Lake, Oregon. *Journal of Paleolimnology*, *31*, 129-138.

Cooke, R., Warren, A., & Goudie, A. (1993). Surface particle concentrations: stone pavements. In *Desert Geomorphology*. (pp. 68-76). London, England: UCL Press Unlimited.

Cooke, R. U. (1970). Stone Pavements in Deserts. *Annals of the Association of American Geographers*, *60*(3), 560-577.

Cox, G. W. & Hunt, J. (1990). Nature and origin of stone stripes on the Columbia Plateau. *Landscape Ecology*, *5*(1), 53-64.

Cronquist, A. (1978). The biota of the intermountain region in geohistorical context. *Great Basin Naturalist Memoirs*, *2*, 3-15.

Cummings, M. L., Evans, J. G., Ferns, M. L., & Lees, K. R. (2000). Stratigraphic and structural evolution of the middle Miocene synvolcanic Oregon-Idaho graben. *GSA Bulletin*, *112*(5), 668-682.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319-339.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, *35*, 982-1002.

Dokuchaev, V. V. (1883). *Russian Chernozem.* St. Petersburg, Russia.

Earth Satellite Corporation (2003). *CART software user's guide.* U.S. Geological Survey - National Land Cover Database (NLCD).

Elnaggar, A.A. (2007). *Development of predictive mapping techniques for soil survey and salinity mapping.* (Doctoral dissertation, Oregon State University, Corvallis, Oregon). Retrieved from http://hdl.handle.net/1957/5754. 148 pp.

Elnaggar, A. A. & Noller, J. S. (2008a). Assessing the Consistency of Conventional Soil Survey Data: Switching from Conventional to Digital Soil Mapping Techniques. *Soil Science Society of America Journal*, submitted.

Elnaggar, A. A. & Noller, J. S. (2008b). Spatial Data Mining and Soil-Landscape Modeling Applied to Initial Soil Survey. *Geoderma*, submitted.

Ferns, M. L. (1997). Field trip guide to the eastern margin of the Oregon-Idaho graben and the middle Miocene calderas of the Lake Owyhee volcanic field. *Oregon Geology*, *59*(1), 9-20.

Franklin, J. (1995). Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, *19*(4), 474-499.

Franklin, J. (1998). Predicting the Distribution of Shrub Species in Southern California from Climate and Terrain-Derived Variables. *Journal of Vegetation Science*, *9*(5), 733-748.

Fritz, S. C. (1996). Paleolimnological Records of Climatic Change in North America. *Limnology and Oceanography*, *41*(5), 882-889.

Fu, P. & Rich, P.M. (1999). *The solar analyst 1.0 user manual.* Helios Environmental Modeling Institute, LLC. http://www.fs.fed.us/informs/solaranalyst/solar_analyst_users_guide.pdf

Geist, D. & Richards, M. (1993). Origin of the Columbia Plateau and Snake River plain: Deflection of the Yellowstone plume. *Geology*, *21*, 789-792.

Gile, L. H., Peterson, F. F., & Grossman, R. B. (1966). Morphological and Genetic Sequences of Carbonate Accumulations in Desert Soils. *Soil Science*, *101*(5), 347-360.

Goldthwait, R. P. (1976). Frost Sorted Patterned Ground: A Review. *Quaternary Research*, *6*, 27-35.

Gordon, I. (1985). The Paleocene Denning Spring flora of north-central Oregon. *Oregon Geology*, *47*(10), 115-118.

Guisan, A. & Zimmerman, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*, 147-186.

Harden, J. W., Taylor, E. M., Reheis, M. C., & McFadden, L. D. (1991). Calcic, Gypsic, and Siliceous Soil Chronosequences in Arid and Semiarid Environments. In W. D. Nettleton (Ed.), *Occurrence, Characteristics, and Genesis of Carbonate, Gypsum, and Silica Accumulations in Soils: SSSA Special Publication Number 26*. (pp. 1-16). Madison, WI: Soil Science Society of America, Inc.

Hart, W. K. & Mertzman, S. A. (1983). Late Cenozoic stratigraphy of the Jordan Valley area, southeastern Oregon. *Oregon Geology*, *45*(2), 15-19.

Hash, Sarah J. (2007). *Using See5 Decision-Tree Classifier Software to Generate Rule Sets for Predictive Soils Mapping*. Unpublished manuscript, Oregon State University, Corvallis, OR.

Henderson, B. L., Bui, E. N., Moran, C. J., & Simon, D. A. P. (2005). Australia-wide predictions of soil properties using decision trees. *Geoderma*, *124*, 383-398.

Hooper, P. R. (1997). The Columbia River flood basalt province: Current status. In J. J. Mahoney & M. F. Coffin (Eds.), *Large Igneous Provinces: Continental, Oceanic and Planetary Flood Volcanism: American Geophysical Union Geophysical Monograph 100*. (pp. 1-28). Washington, DC: American Geophysical Union.

Hooper, P. R., Binger, G. B., & Lees, K. R. (2002). Age of the Steens and Columbia River flood basalts and their relationship to extension-related calc-alkalic volcanism in eastern Oregon. *GSA Bulletin*, *114*(1), 43-50.

ITT Industries (2006). ENVI Version 4.3. http://www.RSInc.com/envi.

Jenness, J. (2005). Topographic position index (tpi_jen.avx) extension for ArcView 3.x. Flagstaff, AZ: Jenness Enterprises. http://www.jennessent.com/arcview/tpi.htm.

Jenny, H. (1941). *Factors of Soil Formation: A System of Quantitative Pedology*. New York, NY: McGraw-Hill. 324 pp.

Jessup, R. W. (1960). The stony tableland soils of the southeastern portion of the Australian Arid Zone and their evolutionary history. *Australian Journal of Soil Science*, *11*, 188-196.

Johansen, J. R. (1993). Cryptogamic Crusts of Semiarid and Arid Lands of North America. *Journal of Phycology*, *29*, 140-147.

Kagan, J. S., Hak, J. C., Csuti, B., Kiilsgaard, C. W., & Gaines, E. P. (1999). *Oregon Gap Analysis Project Final Report: A Geographical Approach to Planning for Biological Diversity*. Portland, OR: Oregon Natural Heritage Program. 168 pp.

Kauth, R.J. & Thomas, G.S. (1976). The tasseled cap: a graphical description of the spectral-temporal development of agricultural crops as seen by LANDSAT. *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, 4B41-4B51. West Lafayette, IN: Purdue University.

Kittleman, L. R. (1973). *Guide to the Geology of the Owyhee Region of Oregon* (Bulletin No. 21). Eugene, OR: University of Oregon, Museum of Natural History. 61 pp.

Knapp, P. A. (1996). Cheatgrass (*Bromus Tectorum L*) dominance in the Great Basin Desert. *Global Environmental Change*, *6*(1), 37-52.

Lagacherie, P. & Holmes, S. (1997). Addressing geographical data errors in a classification tree for soil unit prediction. *International Journal of Geographical Information Science*, *11*(2), 183-198.

Landsat.org. (1989). LandSat TM Path 42, Row 30. Acquisition date July 25, 1989. East Lansing, MI: Michigan State University, Global Observatory for Ecosystem Services. http://www.landsat.org.

Lees, B. G. & Ritman, K. (1991). Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environmental Management*, *15*(6), 823-831.

Leica Geosystems (2008). *ERDAS Field Guide*. Norcross, GA: Leica Geosystems Geospatial Imaging, LLC.

Luoto, M. & Hjort, J. (2005). Evaluation of current statistical approaches for predictive geomorphological mapping. *Geomorphology*, *67*, 299-315.

Malde, H. E. (1961). Patterned Ground of Possible Solifluction Origin at Low Altitude in the Western Snake River Plain, Idaho. *Short papers in the geologic and hydrologic sciences: U.S. Geological Survey Professional Paper 424-B*. (pp. B170-B173).

Malde, H. E. (1964). Patterned Ground in the Western Snake River Plain, Idaho, and Its Possible Cold-Climate Origin. *GSA Bulletin*, *75*, 191-208.

Malone, M. (in preparation). *Predictive Mapping for the Delineation of Landtype Association Units in the Fremont National Forest, Oregon.* (M.S. Thesis, Oregon State University, Corvallis, Oregon).

Manchester, S. R. (1981). Fossil plants of the Eocene Clarno Nut Beds. *Oregon Geology*, *43*(6), 75-81.

Mather, P.M. (2004). *Computer processing of remotely-sensed images—an introduction. 3rd Ed.* Chichester, England: John Wiley and Sons Ltd. 347 pp.

Mathieson, K. (1991). Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior. *Information Systems Research*, *2*(3), 173-191.

McBratney, A. B., Medoca Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, *117*(1-2), 3-52.

McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S., & Shatar, T. M. (2000). An overview of pedometric techniques for use in soil survey. *Geoderma*, *97*, 293-327.

McDonald, E. V. (1994). *The Relative Influence of Climatic Change, Desert Dust, and Lithological Control on Soil-Geomorphic Processes and Hydrology of Calcic Soils formed on Quaternary Alluvial-Fan Deposits in the Mojave Desert, California.* (Doctoral Dissertation, University of New Mexico, Albuquerque, NM). 383 pp.

McFadden, L. D., Eppes, M. C., Gillespie, A. R., & Hallet, B. (2005). Physical weathering in arid landscapes due to diurnal variation in the direction of solar heating. *GSA Bulletin*, *117*(1/2), 161-173.

McFadden, L. D., Wells, S. G., & Dohrenwend, J. C. (1986). Influences of Quaternary Climatic Changes on Processes of Soil Development on Desert Loess Deposits of the Cima Volcanic Field, California. *Catena*, *13*, 361-389.

McFadden, L. D., McDonald, E. V., Wells, S. G., Anderson, K., Quade, J., & Forman, S. L. (1998). The vesicular layer and carbonate collars of desert soils and pavements: formation, age and relation to climate change. *Geomorphology*, *24*, 101-145.

McFadden, L. D., Wells, S. G., & Jercinovich, M. J. (1987). Influences of eolian and pedogenic processes on the origin and evolution of desert pavements. *Geology*, *15*, 504-508.

McKenzie, N. J. & Ryan, P. J. (1999). Spatial prediction of soil properties using environmental correlation. *Geoderma*, *89*, 67-94.

Mehringer, J., Peter J. (1987). *Late Holocene Environments on the Northern Periphery of the Great Basin: Final Report to Bureau of Land Management, Oregon State Office, Portland for Contract YA551-CT5-340075*. Pullman, WA: Washington State University.

Mehringer, J., Peter J. (1985). Late-Quaternary Pollen Records from the Interior Pacific Northwest and Northern Great Basin of the United States. In J. Bryant, Vaughn M & R. G. Holloway (Eds.), *Pollen Records of Late-Quaternary North American Sediments*. (pp. 167-189). Dallas, TX: American Association of Stratigraphic Palynologists Foundation.

Mehringer, J., Peter J. & Wigand, P. E. (1990). Comparison of Late Holocene Environments from Woodrat Middens and Pollen: Diamond Craters, Oregon. In J. L. Betancourt, T. R. Van Devender, & P. S. Martin (Eds.), *Packrat Middens: The Last 40,000 Years of Biotic Change*. (pp. 294-325). Tucson, AZ: The University of Arizona Press.

Meyer, H. W. & Manchester, S. R. (1997). *The Oligocene Bridge Creek Flora of the John Day Formation, Oregon* (141). Berkeley and Los Angeles, CA: University of California Press. 364 pp.

Meyer, W. H. (1934). Growth in Selectively Cut Ponderosa Pine Forests of the Pacific Northwest. *U.S. Dept. of Ag. Tech. Bull. 407*. Washington, DC: U.S. Department of Agriculture. 64 pp.

Miller, J. & Franklin, J. (2002). Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*, *157*(2-3), 227-247.

Moran, C. J. & Bui, E. N. (2002). Spatial data mining for enhanced soil map modelling. *International Journal of Geographical Information Science*, *16*(6), 533-549.

Negrini, R. M., Erbes, D. B., Faber, K., Herrera, A. M., Roberts, A. P., Cohen, A. S., et al. (2000). A paleoclimate record for the past 250,000 years from Summer Lake, Oregon, USA: I. Chronology and magnetic proxies for lake level. *Journal of Paleolimnology*, *24*, 125-149.

Nettleton, W. D., Witty, J. E., Nelson, R. E., & Hawley, J. W. (1975). Genesis of Argillic Horizons in Soils of Desert Areas of the Southwestern United States. *Soil Science Society of America Journal*, *39*, 919-926.

Noller, J.S. (2008). Surficial geology maps of Jordan Craters South, Arock, Threemile Hill, Danner and Dry Creek Rim 7.5' quadrangles, Malheur County, Oregon. 1:24,000 scale. Unpublished maps, Oregon State University, Corvallis, Oregon.

Norton, J. B., Monaco, T. A., Norton, J. M., Johnson, D. A., & Jones, T. A. (2002). *Cheatgrass Invasion Alters Soil Morphology and Organic Matter Dynamics in Big Sagebrush-Steppe Rangelands*. Paper presented at the Seed and Soil Dynamics in Shrubland Ecosystems, Ogden, UT.

Norton, J. B., Monaco, T. A., Norton, J. M., Johnson, D. A., & Jones, T. A. (2004). Soil morphology and organic matter dynamics under cheatgrass and sagebrush-steppe plant communities. *Journal of Arid Environments*, *57*, 445-466.

Nowak, C. L., Nowak, R. S., Tausch, R. J., & Wigand, P. E. (1994a). Tree and Shrub Dynamics in Northwestern Great Basin Woodland and Shrub Steppe During the Late-Pleistocene and Holocene. *American Journal of Botany*, *81*(3), 265-277.

Nowak, C. L., Nowak, R. S., Tausch, R. J., & Wigand, P. E. (1994b). A 30,000 year record of vegetation dynamics at a semi-arid locale in the Great Basin. *Journal of Vegetation Science*, *5*, 579-590.

Oregon Dept. of Geology and Mineral Industries (2006). *Oregon Geologic Data Compilation (OGDC) v. 3*. http://www.oregongeology.com/sub/ogdc/.

Oregon Natural Heritage Program (2002). *Oregon Historic Vegetation*. http://gis.oregon.gov/DAS/EISPD/GEO/alphalist.shtml.

Otto, B. R. & Hutchison, D. A. (1977). The Geology of Jordan Craters, Malheur County, Oregon. *Ore Bin*, *39*(8), 125-140.

Pace, J. (2007). *Soil Survey Technical Note – Predictive Soil Mapping, Wheeler County Soil Survey as a Case Study*. (M.Ag. Project Report, Oregon State University, Corvallis, Oregon).

Peterson, F. F. (1980). Holocene Desert Soil Formation under Sodium Salt Influence in a Playa-Margin Environment. *Quaternary Research*, *13*, 172-186.

Ponzetti, J. M. & McCune, B. P. (2001). Biotic Soil Crusts on Oregon's Shrub Steppe: Community Composition in Relation to Soil Chemistry, Climate and Livestock Activity. *The Bryologist*, *104*(2), 212-225.

PRISM Group, Oregon State University (2007). 800m Temperature and Precipitation Normals. http://www.prismclimate.org. Created June 25, 2007.

Qi, F. (2004). Knowledge Discovery from Area-Class Resource Maps: Data Preprocessing for Noise Reduction. *Transactions in GIS*, *8*(3), 297-308.

Qi, F. & Zhu, A.-X. (2003). Knowledge discovery from soil maps using inductive learning. *International Journal of Geographical Information Science*, *17*(8), 771-795.

Quinlan, J. R. (2007). See5: An Informal Tutorial. http://www.rulequest.com/see5-win.html.

Raven, K.A. (2004). *An investigation of soil, vegetation and mycorrhizal characteristics associated with native grass re-establishment in crested wheatgrass seedings.* (M.S. Thesis, Oregon State University, Corvallis, Oregon). 89 pp.

Reheis, M. (1999). Highest Pluvial-Lake Shorelines and Pleistocene Climate of the Western Great Basin. *Quaternary Research*, *52*, 196-205.

Reheis, M. C., Sarna-Wojcicki, A. M., Reynolds, R. L., Repenning, C. A., & Mifflin, M. D. (2002). Pliocene to Middle Pleistocene Lakes in the Western Great Basin: Ages and Connections. In R. Hershler, D. B. Madsen, & D. R. Currey (Eds.), *Great Basin Aquatic Systems History*. (pp. 53-108). Washington, D.C.: Smithsonian Institution Press.

Reheis, M. C., Goodmacher, J. C., Harden, J. W., McFadden, L. D., Rockwell, T. K., Shroba, R. R., et al. (1995). Quaternary soils and dust deposition in southern Nevada and California. *GSA Bulletin*, *107*, 1003-1022.

Retallack, G. (1981). Preliminary observations on fossil soils in the Clarno Formation (Eocene to early Oligocene) near Clarno, Oregon. *Oregon Geology*, *43*(11), 147-150.

Retallack, G. J. (1991). A field guide to mid-Tertiary paleosols and paleoclimatic changes in the high desert of central Oregon - Part 2. *Oregon Geology*, *53*(4), 75-80.

Retallack, G. J. (2004a). Late Oligocene bunch grassland and early Miocene sod grassland paleosols from central Oregon, USA. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *207*, 203-237.

Retallack, G. J. (2004b). Late Miocene climate and life on land in Oregon within a context of Neogene global change. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *214*, 97-123.

Retallack, G. J. (2007). Cenozoic Paleoclimate on Land in North America. *The Journal of Geology*, *115*, 271-294.

Retallack, G. J., Tanaka, S., & Tate, T. (2002). Late Miocene advent of tall grassland paleosols in Oregon. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *183*, 329-354.

Retallack, G. J., Wynn, J. G., & Fremd, T. J. (2004). Glacial-interglacial-scale paleoclimate change without large ice sheets in the Oligocene of central Oregon. *Geology, 32*(4), 297-300.

Rogers, E. M. (1995). Innovativeness and Adopter Categories. In *Diffusion of Innovations, 4th ed*. (pp. 252-280). New York, NY: The Free Press.

Savory, A. (1988). *Holistic Resource Management*. Covelo, California: Island Press. 590 pp.

Schwarz, T. (1997). Lateritic bauxite in central Germany and implications for Miocene palaeoclimate. *Palaeogeography, Palaeoclimatology, Palaeoecology, 129*, 37-50.

Scull, P., Franklin, J., & Chadwick, O. A. (2005). The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modelling, 181*, 1-15.

Scull, P., Franklin, J., Chadwick, O. A., & McArthur, D. (2003). Predictive soil mapping: a review. *Progress in Physical Geography, 27*(2), 171-197.

Sheldon, N. D., Retallack, G. J., & Tanaka, S. (2002). Geochemical climofunctions from North American soils and application to paleosols across the Eocene-Oligocene boundary in Oregon. *Journal of Geology, 110*, 687-696.

Shoemaker, K. A. & Hart, W. K. (2002). Temporal Controls on Basalt Genesis and Evolution on the Owyhee Plateau, Idaho and Oregon. In B. Bonnichsen, C. M. White, & M. McCurry (Eds.), *Tectonic and Magmatic Evolution of the Snake River Plain Volcanic Province: Idaho Geological Survey Bulletin 30*. (pp. 313-328). Moscow, ID: Idaho Geological Survey.

Simard, M., De Grandi, G., Saatchi, S., & Mayaux, P. (2002). Mapping tropical coastal vegetation using JERS-1 and ERS-1 radar data with a decision tree classifier. *International Journal of Remote Sensing, 23*(7), 1461-1474.

Soil Survey Staff (2006a). *Keys to Soil Taxonomy, 10th ed.* Washington, DC: U.S. Government Printing Office.

Soil Survey Staff (2006b). *Land resource regions and major land resource areas of the United States. Agriculture Handbook 296.* Washington, D.C.: United States Department of Agriculture, Soil Conservation Service.

Southard, R. J. & Southard, A. R. (1985). Genesis of Cambic Horizons in Two Northern Utah Aridisols. *Soil Science Society of America Journal, 49*, 167-171.

Stehman, S.V. (1999). Alternative Measures for Comparing Thematic Map Accuracy. In K. Lowell and A Jaton (Eds.), *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*. (pp. 45-51). Chelsea, MI: Ann Arbor Press.

Sullivan, L. A. & Koppi, A. J. (1991). Morphology and Genesis of Silt and Clay Coatings in the Vesicular Layer of a Desert Loam Soil. *Australian Journal of Soil Research, 29*, 579-586.

Takahashi, E., Nakajima, K., & Wright, T. L. (1998). Origin of the Columbia River basalts: Melting model of a heterogeneous plume head. *Earth and Planetary Science Letters*, *162*, 63-80.

Thompson, R. S. (1990). Late Quaternary Vegetation and Climate in the Great Basin. In J. L. Betancourt, T. R. Van Devender, & P. S. Martin (Eds.), *Packrat Middens: The Last 40,000 Years of Biotic Change*. (pp. 200-239). Tucson, AZ: The University of Arizona Press.

Thompson, R. S. (1991). Pliocene environments and climates in the western United States. *Quaternary Science Reviews*, *10*, 115-132.

Thompson, R. S. (1996). Pliocene and early Pleistocene environments and climates of the western Snake River Plain, Idaho. *Marine Micropaleontology*, *27*, 141-156.

Thuiller, W., Araujo, M., & Lavorel, S. (2004). Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, *14*, 669-680.

Tukey, J.W. (1977). Box and Whisker Plots. In *Exploratory Data Analysis*. (pp. 39-43). Reading, MA: Addison-Wesley.

U.S. Department of Agriculture, N. C. S. S. (2005). Memorandum of Understanding Relative to the the Making of a Soil Survey for Malheur County, Oregon, Southern Part OR644., 9 pp.

U.S. Department of Agriculture, N. R. C. S. (2007). National Soil Survey Handbook, title 430-IV. http://soils.usda.gov/technical/handbook

U.S. Department of the Interior, B. o. L. M. (2001). *Ecological Site Inventory, Technical Reference 1734-7*. Denver, CO: National Business Center Printed Materials Distribution Services.

Valentine, G. A. & Harrington, C. D. (2006). Clast size controls and longevity of Pleistocene desert pavements at Lathrop Wells and Red Cone Volcanoes, southern Nevada. *Geology*, *34*(7), 533-536.

Vayssieres, M. P., Plant, R. E., & Allen-Diaz, B. H. (2000). Classification Trees: An Alternative Non-Parametric Approach for Predicting Species Distributions. *Journal of Vegetation Science*, *11*(5), 679-694.

Venkatesh, V. & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly*, *24*(1), 115-139.

Walker, G. W. & MacLeod, N. S. (1991). Geologic Map of Oregon. 1:500,000 scale, 2 sheets. Reston, VA: U.S. Geologic Survey.

Wells, P. V. (1983). Paleobiography of montane islands in the Great Basin since the last glaciopluvial. *Ecological Monographs*, *53*, 341-382.

Wells, S. G., Dohrenwend, J. C., McFadden, L. D., Turin, B. D., & Mahrer, K. D. (1985). Late Cenozoic landscape evolution on lava flows of the Cima volcanic field, Mojave Desert, California. *GSA Bulletin*, *96*, 1518-1529.

Wells, S. G., McFadden, L. D., & Dohrenwend, J. C. (1987). Influence of Late Quaternary climatic changes on geomorphic and pedogenic processes on a desert piedmont, Eastern Mojave Desert, California. *Quaternary Research*, *27*, 130-146.

West, N. E. (1990). Structure and Function of Microphytic Soil Crusts in Wildland Ecosystems of Arid to Semi-arid Regions. *Advances in Ecological Research*, *20*, 179-222.

Wigand, P. E. (1987). Diamond Pond, Harney County, Oregon: Vegetation History and Water Table in the Eastern Oregon Desert. *Great Basin Naturalist*, *47*(3), 427-458.

Williams, S. H. & Zimbelman, J. R. (1994). Desert Pavement Evolution: An Example of the Role of Sheetflood. *The Journal of Geology*, *102*, 243-248.

Wilson, R. E. & Young, J. A. (2008). Managing Invasive Noxious Range Weeds in the Great Basin. *University of Nevada Cooperative Extension Fact Sheet 96-12*.

Woolard, E. W. (1937). Climatic Cycles in Eastern Oregon as Indicated by Tree Rings. *Monthly Weather Review*, *65*(5), 175-188.

Yaalon, D. H. (1974). Note on some geomorphic effects of temperature changes on desert surfaces. *Zeitschrift fur Geomorphologie N.F*, *21*, 29-34.

Appendices

Appendix I: Pre-Pleistocene Vegetation and Climatic History

Paleocene, Eocene and Oligocene

The very first lands of modern-day Oregon, the Klamath Mountains, did not join the North American continent until the early Cretaceous (~144 Ma). The volcanic arc that would become the Blue Mountains completed the first Oregon coastline by the middle Cretaceous (~100 Ma). The palms and other tropical flora of Oregon's coastline and the ferns of the mountain foothills were likely eradicated at the end of the Cretaceous, along with 70% of all plant and animal species on earth, during catastrophic meteorite impacts at the K-T Boundary (Bishop, 2003). The oldest known flora in the state, and the only one representing Paleocene times, was collected and described by Gordon (1985) near Pilot Rock in north-central Oregon. Most of the plants described are dicotyledonous woody taxa from the Rutaceae (citrus) family, Betulaceae (birch) family, and Lauraceae (laurel) family, pteridophytes such as *Dryopteris* sp. (wood fern), and gymnosperms such as *Taxodium* sp. (bald cypress) and *Glyptostrobus* sp. (swamp cypress). This vegetation assemblage suggests a mildly subtropical climate, although conditions would continue to warm into the Eocene. Middle Eocene fossil plants in the Clarno Nut Beds (part of the John Day Fossil Beds National Monument) include species from the Palmae (palm) family, Juglandaceae (walnut) family, Menispermaceae (moonseed) family, Lauraceae (avocado) family, Magnoliaceae (magnolia) family, and Moraceae (fig) family, among others, and indicate tropical rain forest conditions (Manchester, 1981). The climate was likely much like modern-day lowlands of Central America. The highly-weathered kaolinite-rich paleosols in the Clarno Nut Beds are further indicators of a warm, wet climate (Retallack, 1981). Fossils dating to the late Eocene reflect a transition to a more temperate vegetation

assemblage of grassland and woodland (Retallack, 1991). The late Eocene and the

Eocene-Oligocene transition are the earliest periods commonly examined by paleoclimate

and paleovegetation studies in eastern Oregon.

Bestland (1997) used geomorphic concepts about what drives fluvial aggradation

and incision along with examinations of paleosols in dated geologic materials of central

Oregon's John Day Formation to look at late Eocene to middle Oligocene climatic change.

The Big Basin Member of the John Day Formation contains a kaolinite-rich Ultisols-like

paleosol that is truncated at approximately 33.2 Ma (interpreted as the Eocene-Oligocene

transition) and overlain by thick alluvial deposits. A smectite-rich, Alfisols-like paleosol

developed in the upper part of these sediments. This sequence suggests an abrupt climatic

shift from an Eocene subtropical humid climate to an Oligocene temperate humid climate

with a period of vegetation disequilibrium and resulting massive soil erosion at the

transition. A stack of red, well-developed paleosols and decreased alluvial deposition in

the 32.6 to 32 Ma time span suggest an early Oligocene warming period where recovering

vegetation suppressed additional erosion. Meyer and Manchester (1997) suggest a mean

annual precipitation (MAP) of about 1000-1500 mm and a mean annual temperature (MAT)

of 3 to 9° C at ~32 Ma based on the Bridge Creek flora of the John Day Formation.

Bestland et al. (1997) found similar paleosol sequences in the Clarno and John Day

Formations. They also discovered Inceptisols-like paleosols with calcic horizons in

mid-Oligocene sediments (30-28 Ma), suggesting another period of cooling and decreasing

precipitation. Bestland (1997) suggests that ~30 Ma marks another major climatic shift, as

evidenced by the transition from non-calcareous smectitic paleosols to calcareous zeolitic

paleosols. Geochemical analysis of Eocene-Oligocene paleosols in the Clarno and John

Day Formations by Sheldon et al. (2002) provides refined paleoprecipitation and temperature estimates for local climate, and these figures are consistent with general and regionally-applicable observations made in earlier research based on depth to calcic horizons and plant fossils. Another study looking at the fossil record and pedogenic features in the paleosols of the middle John Day Formation (Retallack et al., 2004) found that vegetation communities of the Oligocene alternated between semiarid sagebrush steppe and subhumid wooded grassland. This study also found that Oligocene climate change was likely cyclic with Milankovitch-obliquity cycles (41,000-year-intervals). Other studies by Retallack (2007; 2004a), give credence to the Milankovitch-obliquity theory. Most cycles were represented by one paleosol with a deep calcic horizon containing large carbonate nodules (MAP = 490 ± 50 mm) buried by two paleosols with a shallower calcic horizon containing smaller carbonate nodules (MAP = 366 ± 36 mm) (MAP estimates are from a pedotransfer function based on depth to calcic horizon). The soils with deeper calcic horizons also contained dung beetle (*Pallichnus*) and earthworm (*Edaphichnium*) fossils, indicative of grassland ecosystems, while the soils with shallow calcic horizons contained cicada (*Taenidium*) fossils, which are found in modern semiarid sagebrush ecosystems (Retallack et al., 2004). Retallack (2004a; 2007), presents evidence based on soil morphological features (such as presence of fine root traces and crumb soil structure) that bunchgrasses first appeared in the late Oligocene (~30 Ma) during a punctuated warmer, wetter period. This appearance, along with the appearances of short sod grassland in the middle Miocene and tall sod grassland in the late Miocene, was followed by longer-term cooling and drying. This evidence stands in stark contrast with common beliefs that grasses expanded during cool and dry periods (Retallack, 2007).

Although local volcanic and tectonic phenomena (such as the rising Cascades and Coast

Range mountains) likely influenced Oligocene paleoclimate to some extent, results from

all of these studies are well-correlated to global climate shifts that are documented in

marine oxygen isotope records and sea-level changes (Bestland et al., 1997; Bestland,

1997).

Miocene

The oldest sod grassland paleosols in Oregon appeared in the early Miocene (ca.

19 Ma) and indicate a MAP of less than 400 mm (Retallack, 2004a).   The Miocene

Mascall Formation of the John Day area contains fossils of cool-temperate deciduous

angiosperms and conifers, animal fossils, and thin calcareous paleosols that indicate dry,

open grassy woodland vegetation similar to that of modern-day southern Indiana and Ohio

(Retallack, 1991). With an estimated MAT of 9 to 10° C, upper elevations would likely

have experienced wintertime snow (Retallack, 1991). Evidence for an early to middle

Miocene global warm-wet trend can be found in the lateritic paleosols present as far north

as the northwestern U.S., Germany, and Japan, and as far south as southern Australia

(Schwarz, 1997). Woodlands were dominant in Oregon during this time, and Retallack

(2007) used paleosol pedogenic features to derive mid-Miocene (16 Ma) paleoclimate

estimates that agreed with those found by Chaney and Axelrod (1959) using leaf

physiognomy in the Mascall flora (ca. 1270 mm MAP and ca. 17°C MAT). Red, clayey

paleosols with coarse, sparse root traces found between flows of the Picture Gorge basalt

group support this paleoclimate estimate (Retallack, 1991). Climate warmed to a global

thermal maximum in the mid-Miocene (at 16 Ma), but this temperature increase was short

lived; fossil plants indicate significant cooling by 15 Ma and vegetation communities

progressed toward the sagebrush steppe assemblages that would be re-emerge by late

Miocene (Retallack, 2004b). Fossil plants in mid-Miocene (ca. 12 Ma) paleosols near

Unity and Juntura, Oregon indicate grassy live oak (*Quercus pollardiana*) woodland

savannah while depth to pedogenic carbonate and clay chemical composition indicate a

MAT of 13°C and MAP of ca. 879 mm (2004b).   Paleosols in the Rattlesnake Formation

show a major shift from woodland to tall grassland (subhumid to semiarid climate) that

occurred ca. 7.3 Ma, followed by another shift to desert shrubland ca. 7.2 to 7.1 Ma

(semiarid climate similar to that of present-day) (Retallack et al., 2002). Fossil flora, along

with soil morphological features such as concentrically banded rhizoconcretions and

abundant silts and salts, indicate that a Mediterranean (summer dry) climate was present in

eastern Oregon by the late Miocene (Retallack, 2004b). However, Ashwill (1983) points

out that the late Miocene Vibbert flora (described near Gateway, Oregon) contains

abundant specimens of *Populus* (poplar), *Acer* (maple), *Platanus* (sycamore) and *Quercus*

(oak), all of which point to a climate significantly less arid than present-day. Retallack's

paleopedology studies (2007; 2004a), suggest a brief warm wet spike occurred in the late

Miocene (ca. 7 Ma) and ushered in the first appearance of tall sod grassland vegetation.

These conditions would likely have persisted long enough to support the hardwood

vegetation found in the Vibbert flora, even if the overall trend of the late Miocene was

toward a cooler, drier climate.

Pliocene

The earliest stages of the Pliocene were cool and dry with a semiarid climate

similar to that of today. Plant fossils in the Rattlesnake and Deschutes formations of

north-central Oregon support this assertion (Retallack, 1991). When he described the

Deschutes flora, Chaney (1938) concluded that at ca. 4.3 Ma the climate was cool and semiarid with approximately 130 mm more rainfall than today. Retallack (2007) suggests a global warm-wet spike occurred ca. 4 Ma, as evidenced by base-depleted soils with deep calcic horizons dating to this time frame in the John Day area. Work by Robert Thompson (1991; 1996), shows that part of southeastern Oregon and the western Snake River Plain in Idaho were blanketed in conifer forests during the middle Pliocene, and large graben valleys in the area were filled with deep lakes. Conditions continued to warm (although interspersed with punctuated cool dry periods) throughout most of the Pliocene. Palynological studies on the western Snake River Plain indicate that warmer, wetter periods were dominated by coniferous forests (*Pinus* and *Juniperus* species) while cooler, drier periods were characterized by steppe vegetation (Chenopodiaceae, Poaceae, *Artemisia*, *Amaranthus* and *Ambrosia* species) (Thompson, 1991; Thompson, 1996). Changing global air and ocean circulation patterns ushered in a cooling period in the late Pliocene (ca. 2 Ma) that eventually led to the Ice Age (Thompson, 1991).

Appendix II: MCSPSS Rubric for Distinguishing Between Mollisols and Aridisols and Assigning Appropriate Subgroups

This model is only applicable for slopes less than 15%

MESIC Temperature Regime:

Aridisols (Typic, generally **Vitrandic** subgroup):

1) Shadscale vegetation
2) Elevation < 4000 ft.
3) MLRA 24
4) Precipitation 6 to 10"

Aridisols (bordering on Xeric, generally **Vitrixerandic** subgroup):

1) Sagebrush vegetation
2) Elevation 4000-4500 ft.
3) MLRA 25 (occasional MLRA 24)
4) Precipitation 8 to 11"

(Some frigid Aridisols occur on north slopes. Floke, Aritolla, and Rockstop series are exceptions to the precipitation and elevation rules due to skeletal conditions.)

FRIGID Temperature Regime:

Mollisols (Aridic bordering on Xeric, generally **Vitritorrandic** subgroup):

1) Idaho fescue present
2) Elevation 4500 to 4700
3) MLRA 23 and 25
4) Precipitation 11-13"

(Soils sometimes key out as Aridisols in this area due to the thin mollic epipedon. These soils, with the exception of the three listed above, have been correlated to Mollisols)

Mollisols (xeric, generally **Vitrandic** subgroup):

1) Mountain big sagebrush and Idaho fescue, occasional bitterbrush
2) Elevation > 4700 ft
3) MLRA 25
4) Precipitation > 13"
5) sometimes Pachic

Appendix III: Box Plots, Histograms, Summary Statistics, Outlier Maps and Discussion for October 12, 2007 Predictive Run
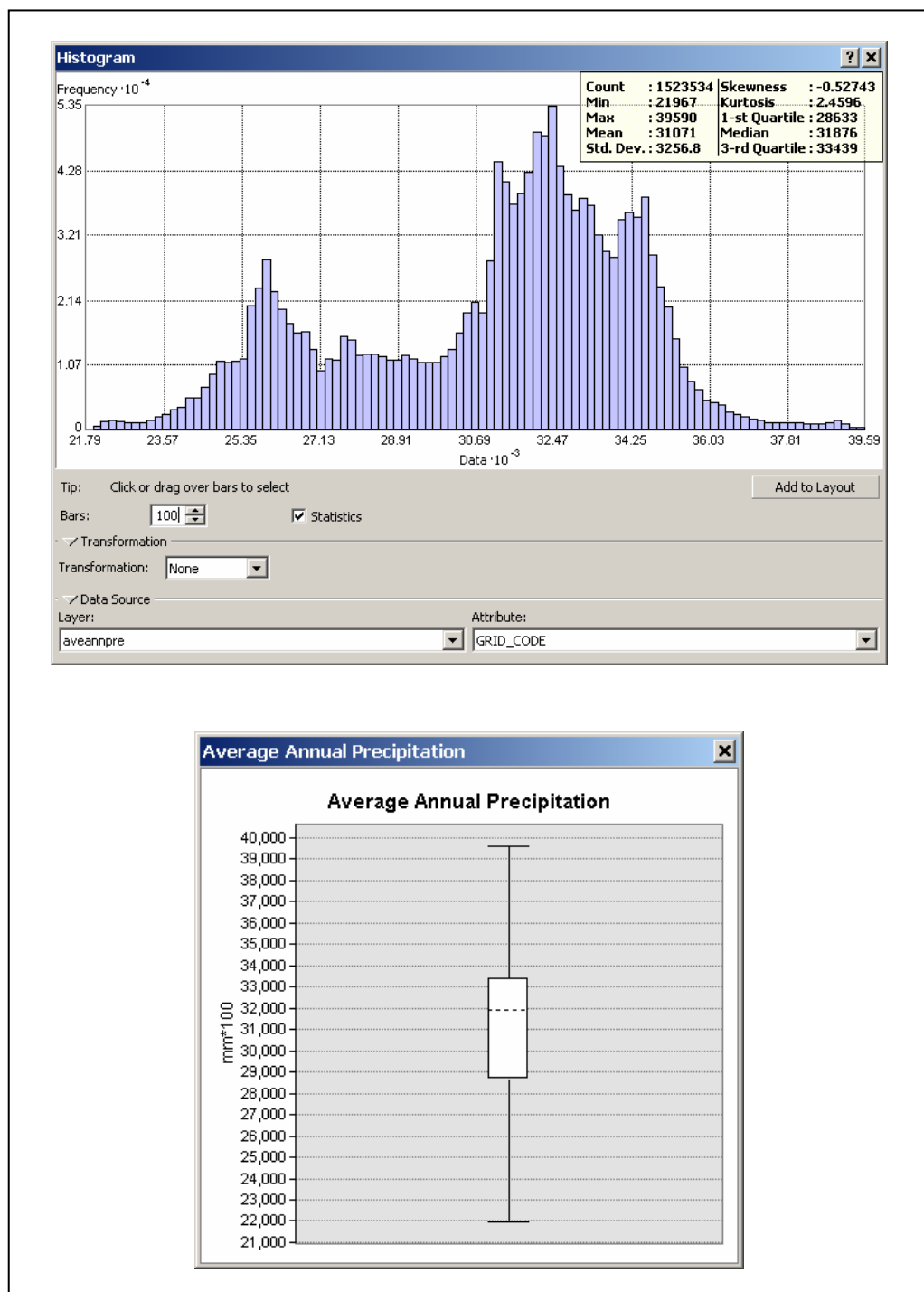
Figure A3-1: Histogram and Box Plot for Mean Annual Precipitation

Histogram

Frequency ·10⁻³

Count : 1523534	Skewness : -0.15182
Min : -19.9	Kurtosis : 2.3811
Max : 241.8	1-st Quartile : 45.6
Mean : 89.36	Median : 96.4
Std. Dev. : 58.184	3-rd Quartile : 129.9

Data ·10⁻¹

Tip: Click or drag over bars to select

Bars: 100   ☑ Statistics

Transformation
Transformation: None

Data Source
Layer: avemintemp   Attribute: GRID_CODE

Average Minimum Temperature
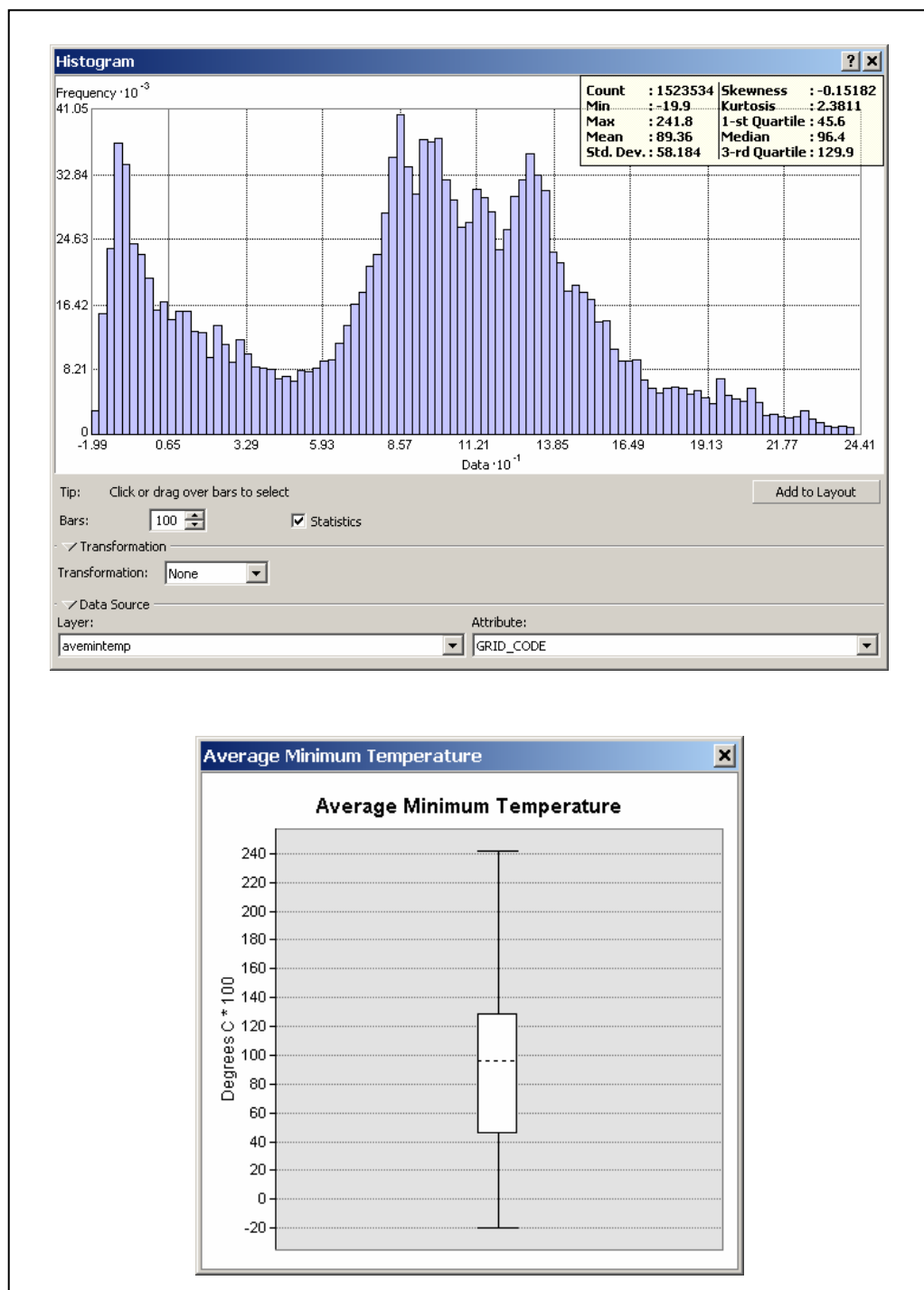
Figure A3-2: Histogram and Box Plot for Average Minimum Temperature
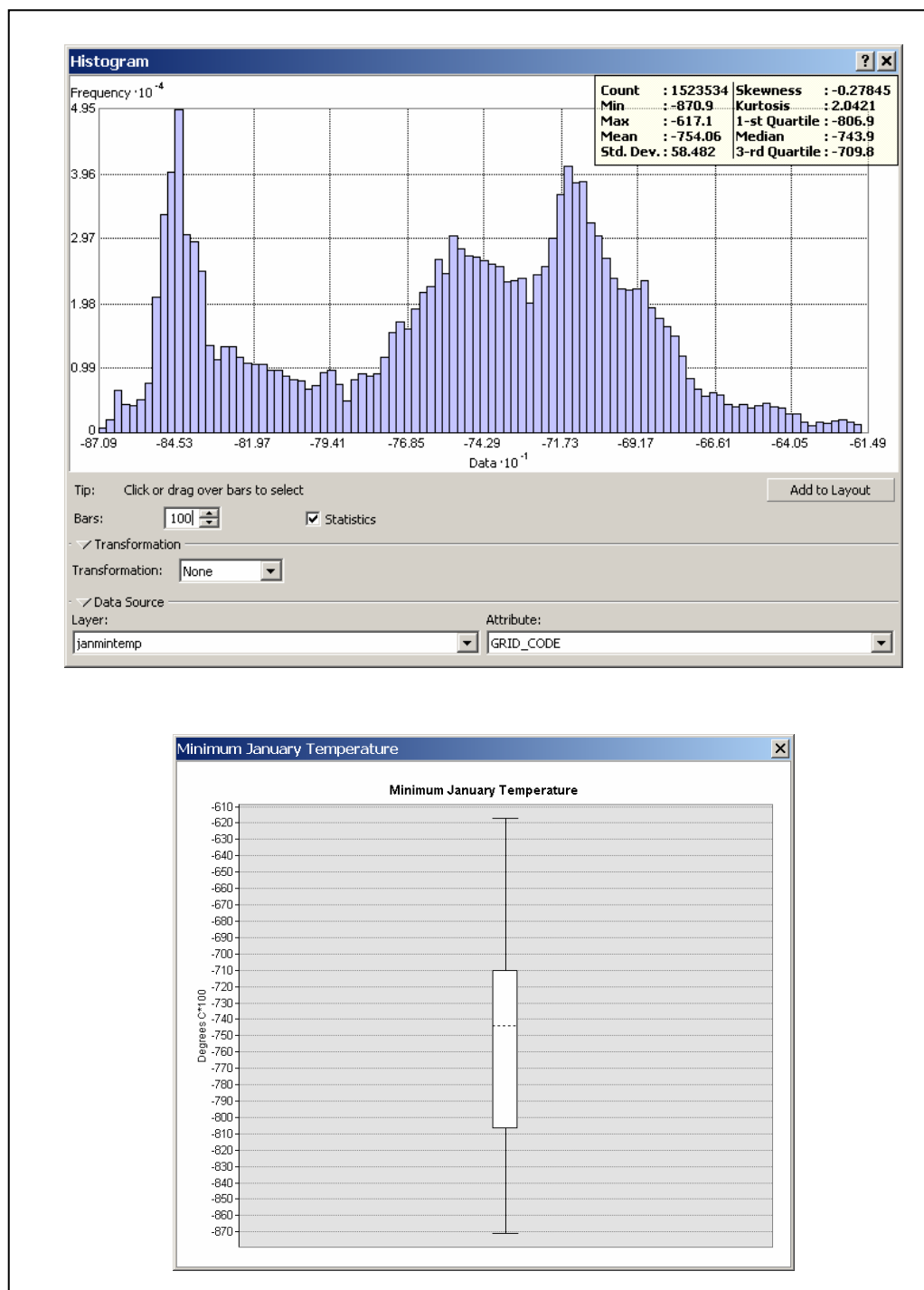
Figure A3-3: Histogram and Box Plot for Minimum January Temperature
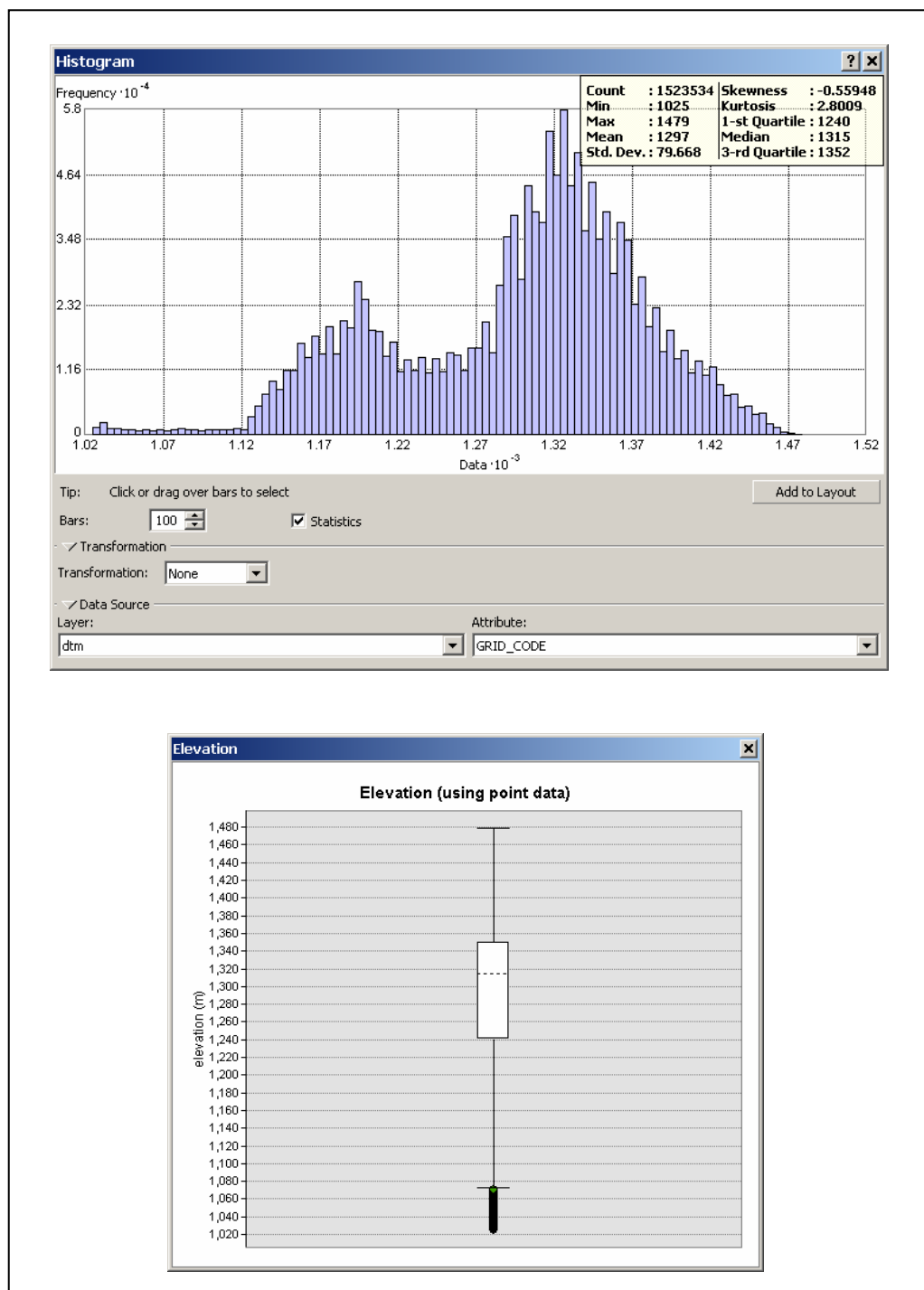
Figure A3-4: Histogram and Box Plot for Elevation
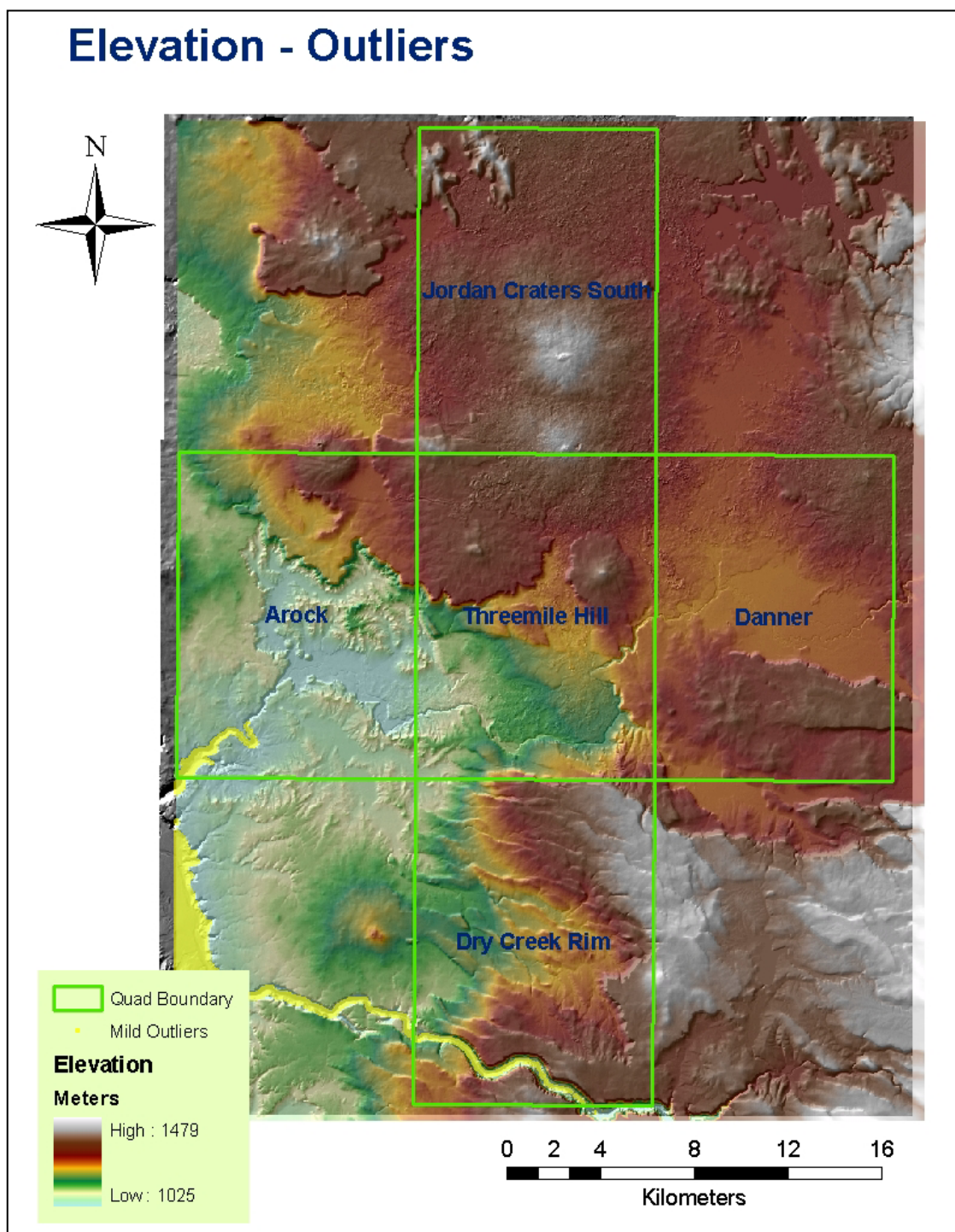
Figure A3-5: Outlier Map for Elevation
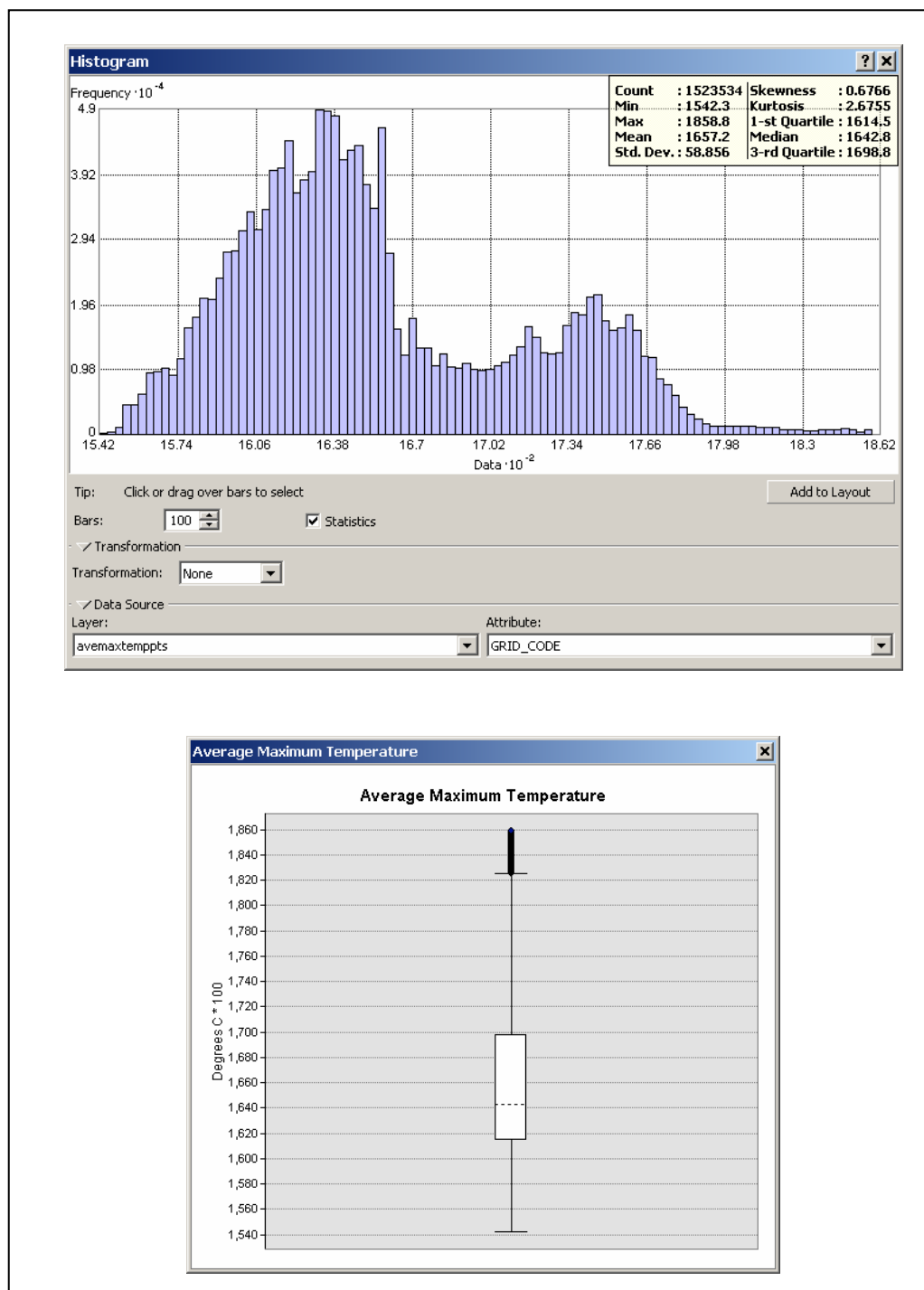
Figure A3-6: Histogram and Box Plot for Average Maximum Temperature
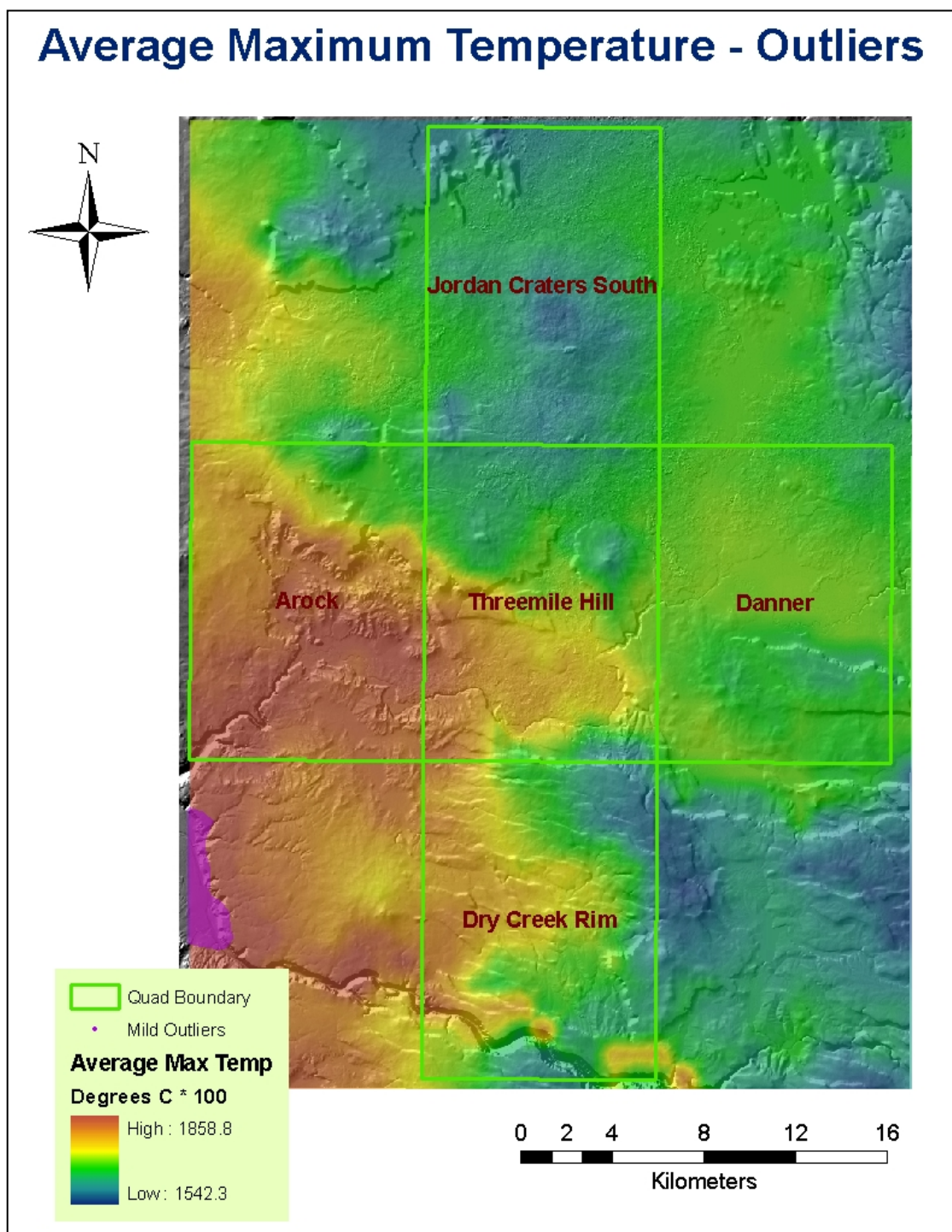
Figure A3-7: Outlier Map for Average Maximum Temperature
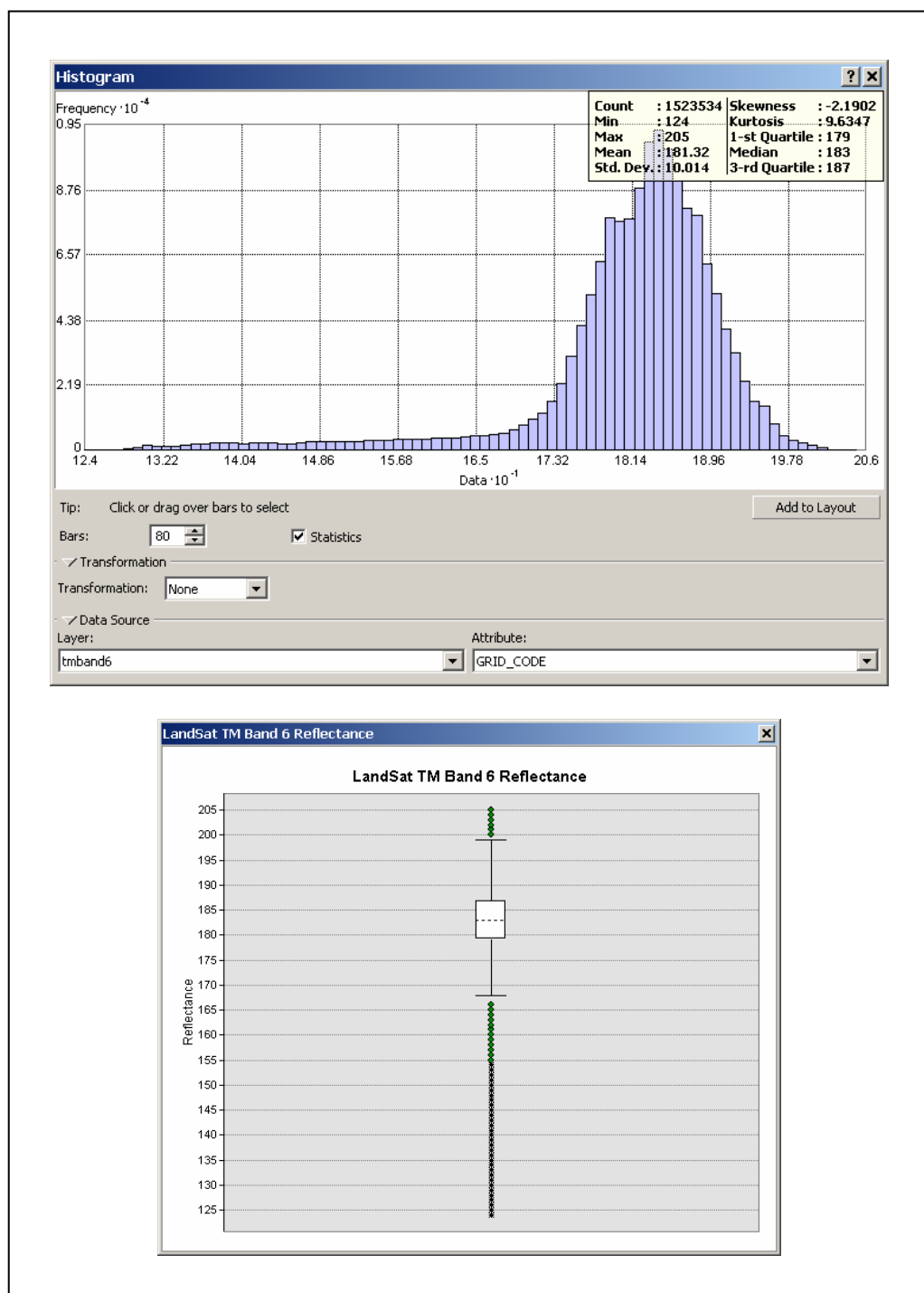
Figure A3-8: Histogram and Box Plot for LandSat TM Band 6 Reflectance
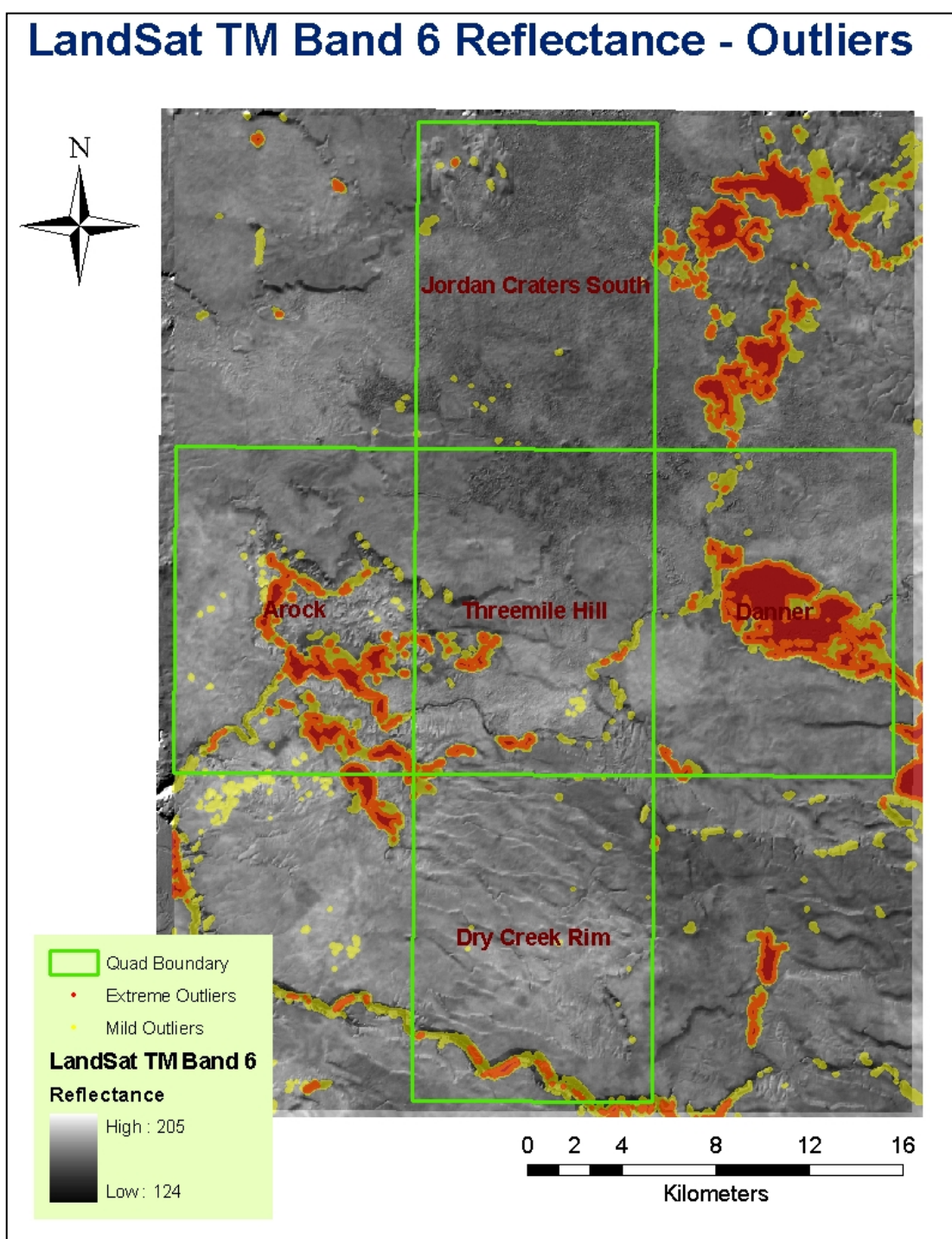
Figure A3-9: Outlier Map for LandSat TM Band 6 Reflectance
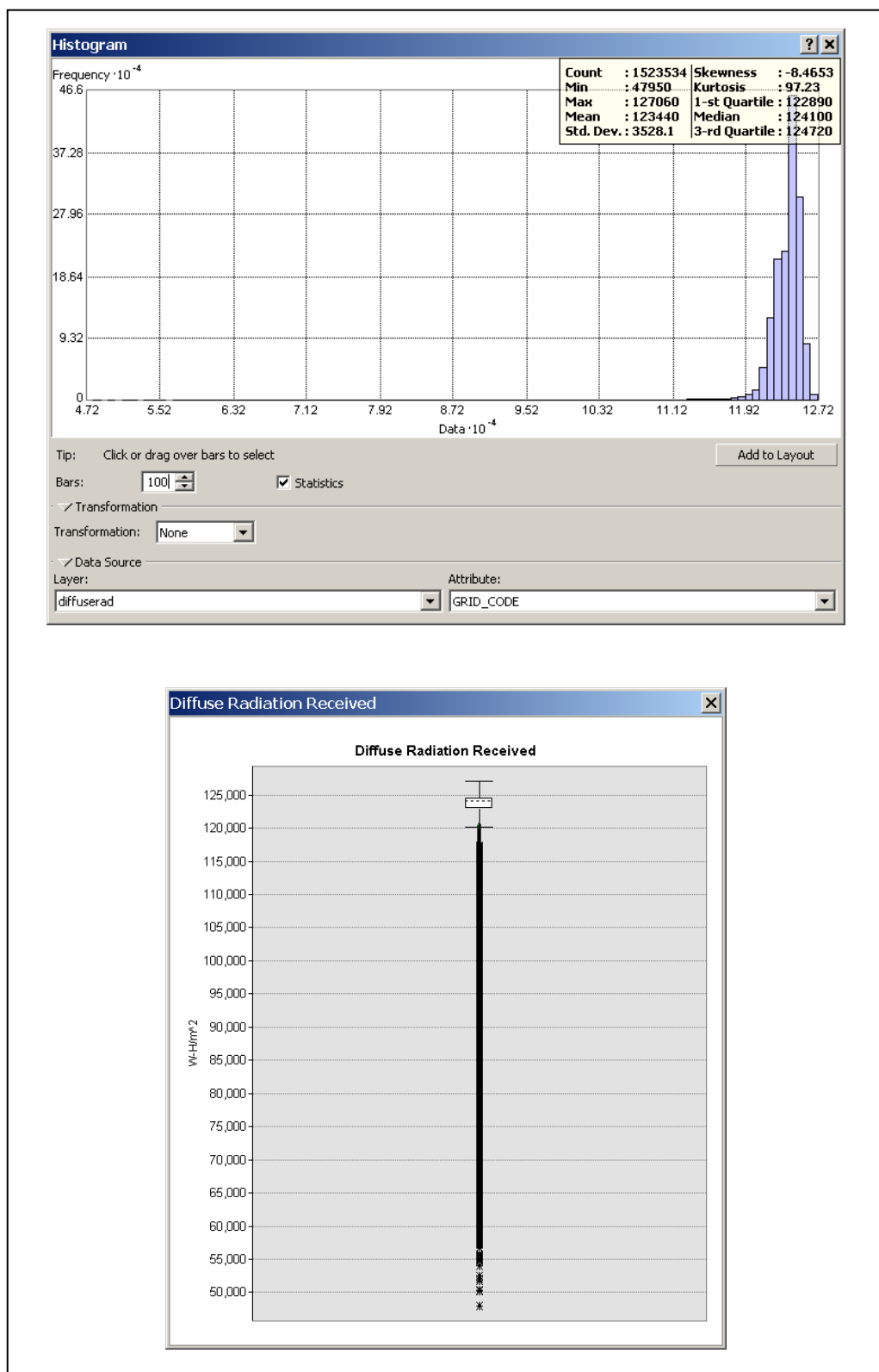
Figure A3-10: Histogram and Box Plot for Diffuse Radiation Received
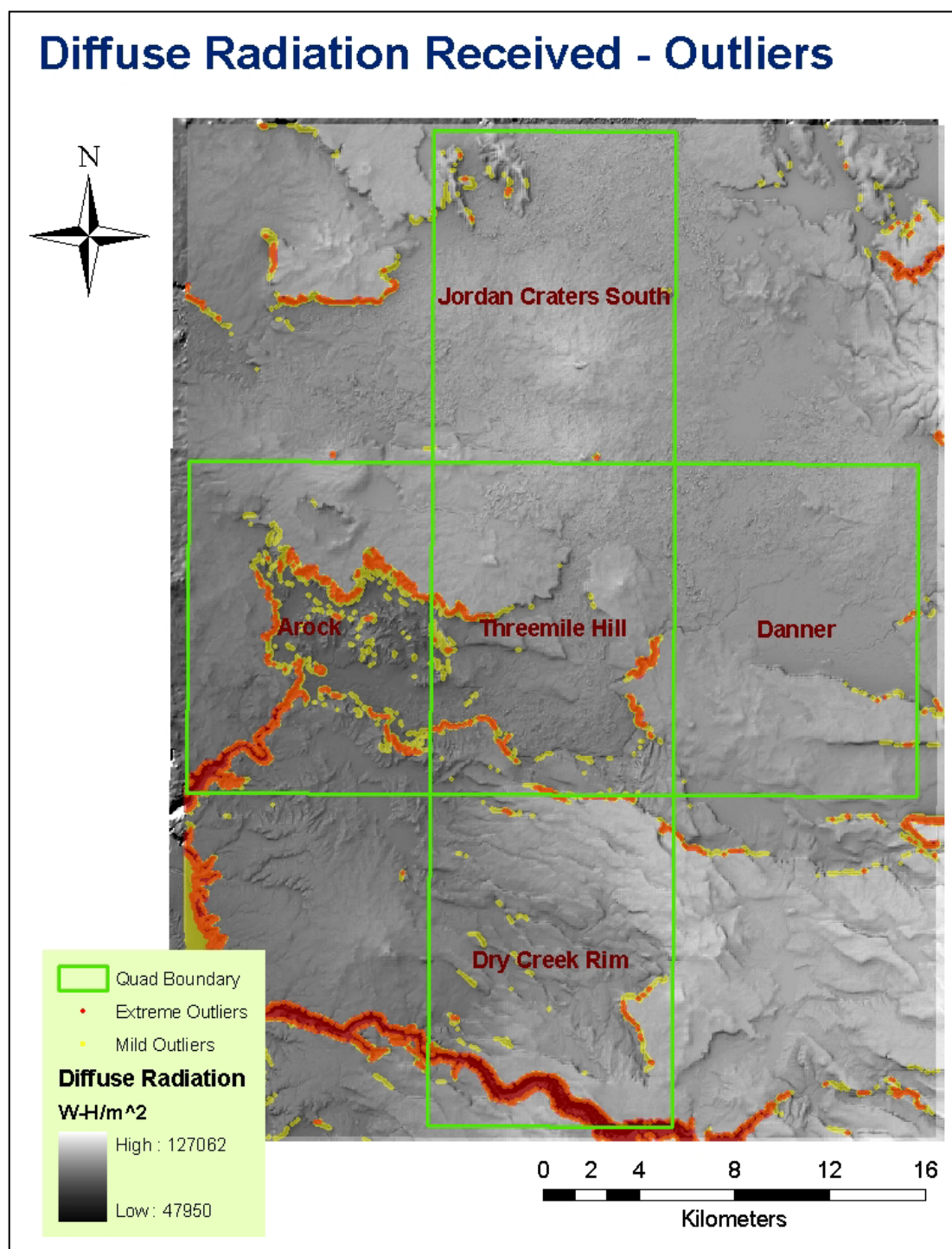
Figure A3-11: Outlier Map for Diffuse Radiation Received
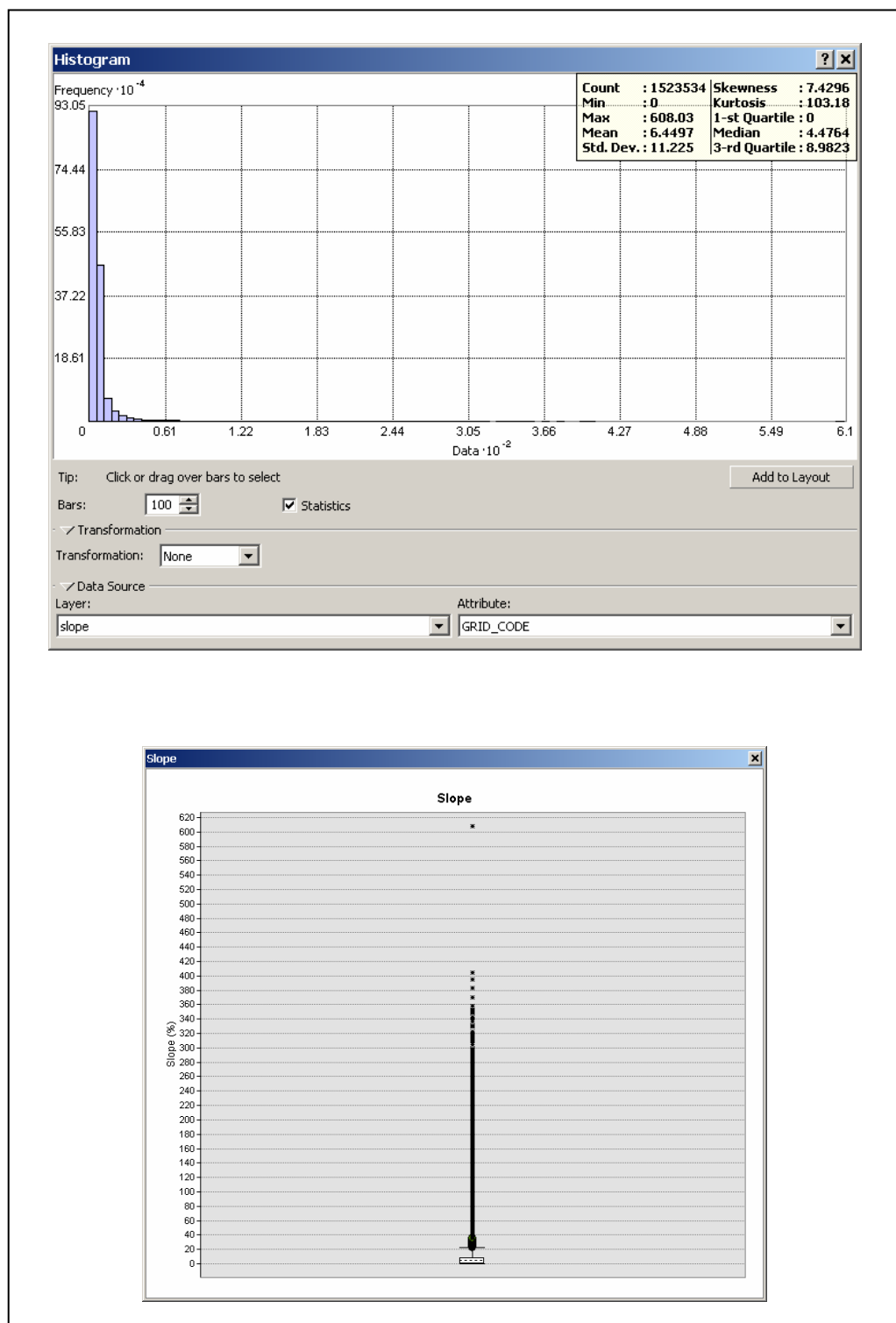
Figure A3-12: Histogram and Box Plot for Slope
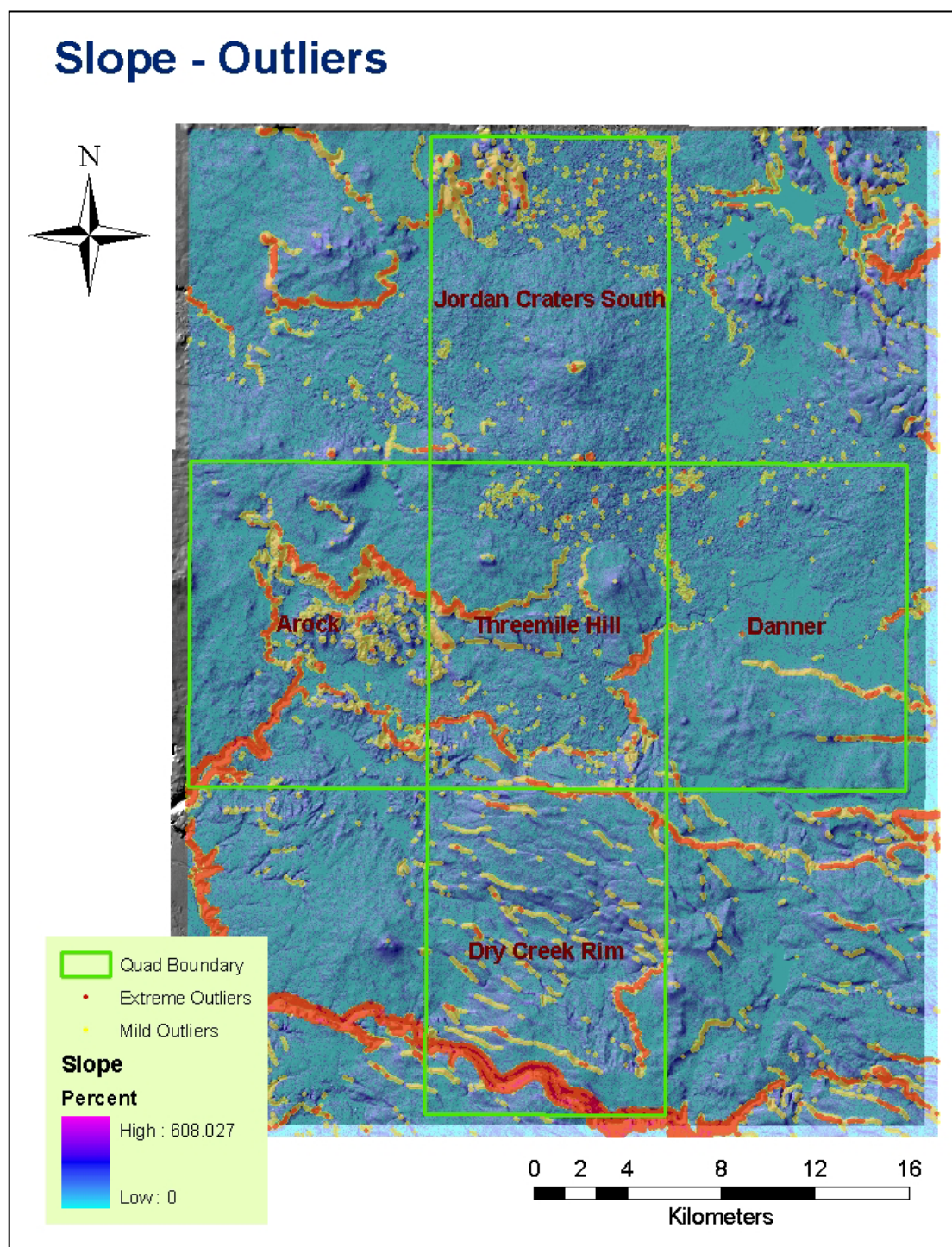
Figure A3-13: Outlier Map for Slope
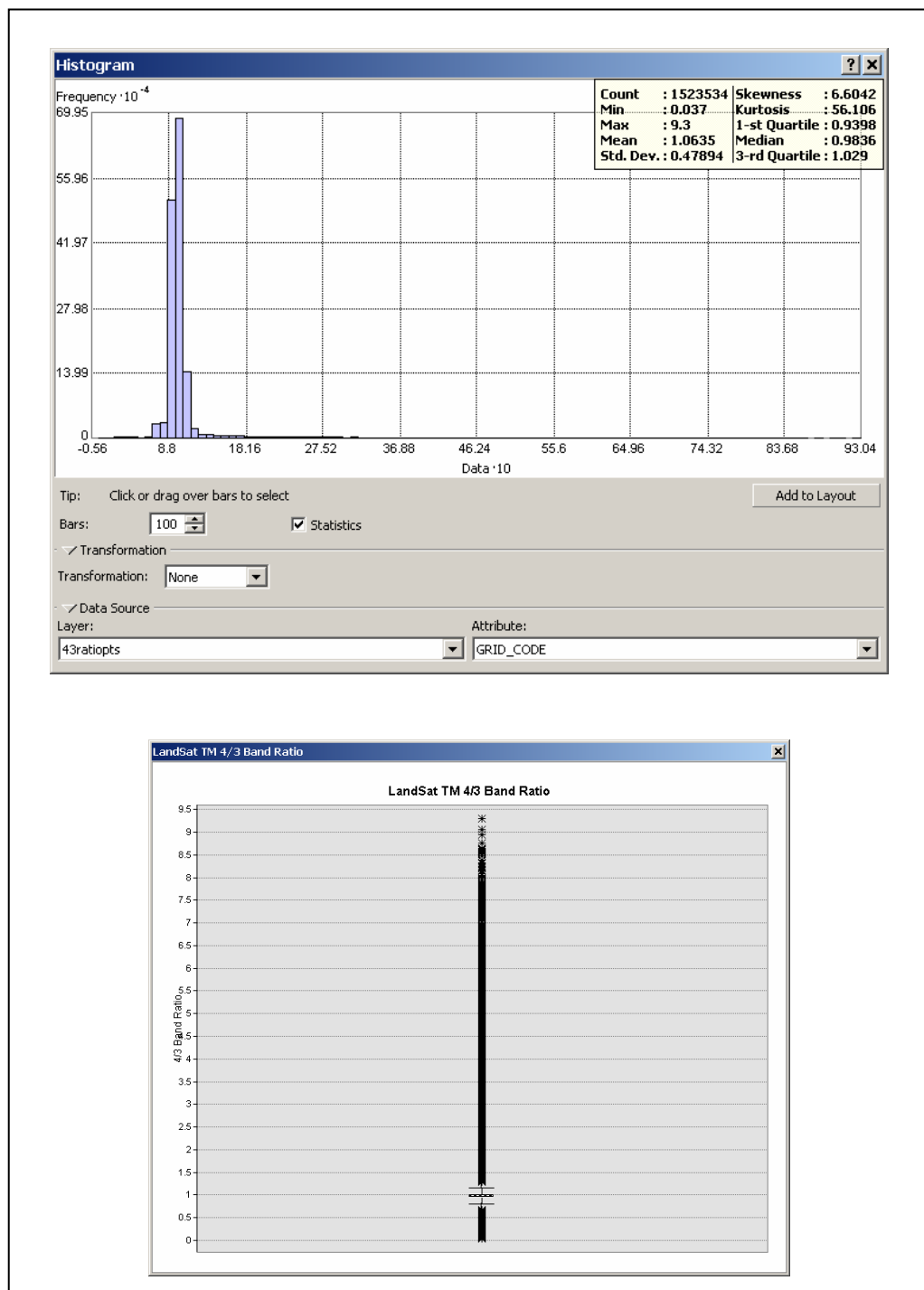
Figure A3-14: Histogram and Box Plot for LandSat TM 4/3 Band Ratio
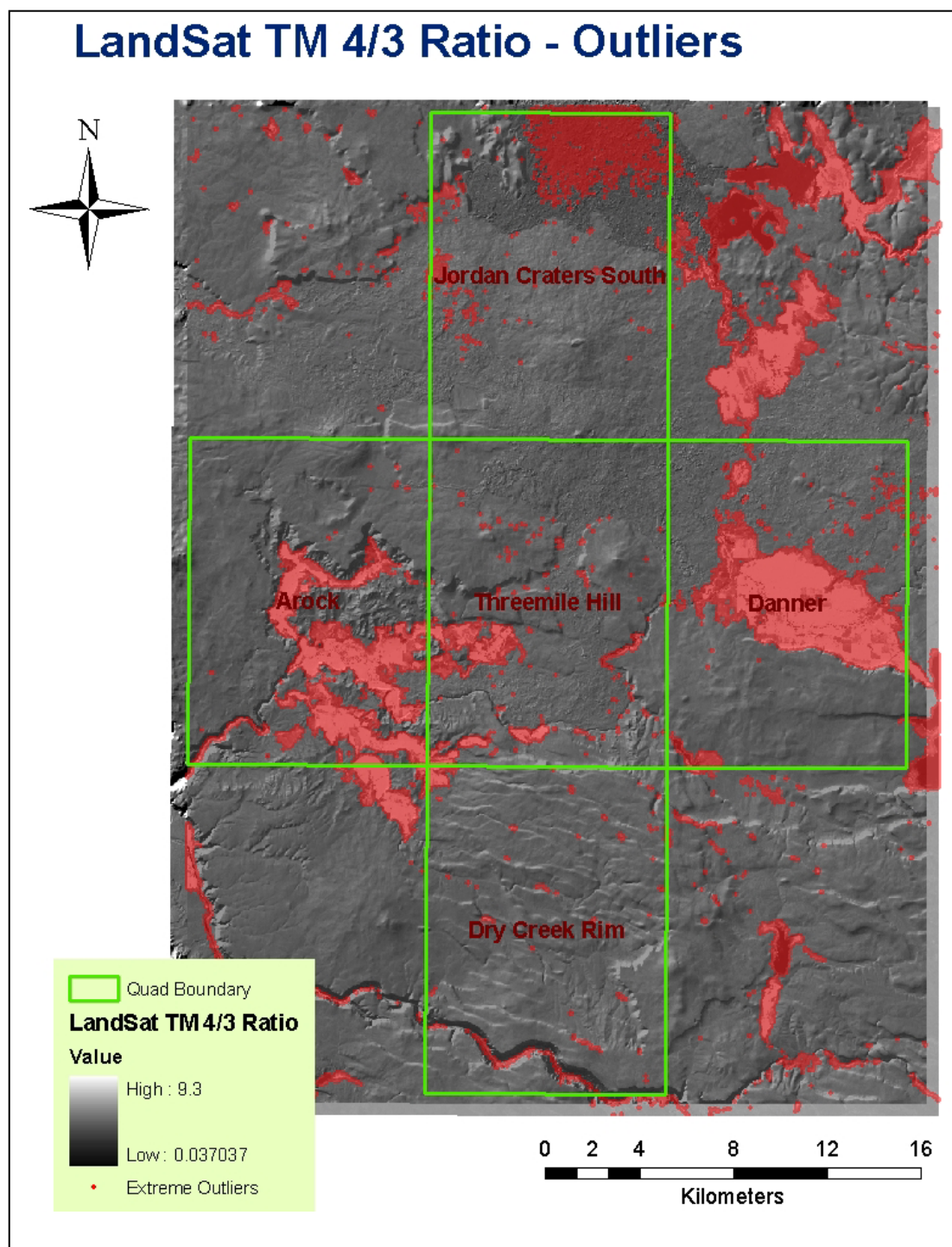
Figure A3-15: Outlier Map for LandSat TM 4/3 Band Ratio

Discussion

Outlier Detection and Descriptive Statistics for October 12, 2007 Predictive Run

Box-plot analysis shows that several of the input variables used for prediction have outliers. The elevation data set has several outliers below 1075 m. Mapping their locations on top of the elevation and hillshade layers shows that all outliers occur in the Owyhee River canyon, which is several hundred meters deep in this region. The outliers appear to be legitimate data points, so no data manipulation is necessary. Outliers in the Average Maximum Temperature data set are found in a broad, low-elevation river valley where temperatures are expected to be warmer. The LandSat TM Band 6 Reflectance data set has many outliers, both mild and extreme. Band 6 measures reflectance in the 10.4-12.5μm range (far infrared part of the electromagnetic spectrum) and is useful for assessing soil moisture and vegetation cover differences. Mild outliers with higher values (greater than 200) occur on lower-elevation summits or ridges where higher temperatures and landscape position make them particularly droughty. These drier regions reflect more far-infrared energy and have higher values. The rest of the mild outliers and all of the extreme outliers have values below 167 and 154 reflectance units respectively, and all occur either along the Owyhee River, on irrigated farmland, or over water bodies. Wetter soils and green vegetation absorb far infrared light, so the smaller reflectance values are expected. The diffuse radiation data set has a few mild and many extreme outliers in the low range. Direct radiation is comprised of energy that comes from sunbeams directly striking the Earth's surface, while diffuse radiation is that which is scattered from the direct solar beam by particles in the Earth's atmosphere. It is unclear why diffuse radiation (instead of direct radiation) proved to be an important predictor in the model. Since most of the study area is

made up of broad, gently sloping lava plateaus it makes sense that most of it would receive relatively high amounts of radiation. Mapping the outliers shows that all occur either in the Owyhee River canyon or along highly-sloping scarps at edges of distinct geologically-controlled landscape features where we would expect less of the sun's energy to reach the surface. Outliers in the slope data set occur in roughly the same locations as the diffuse radiation outliers, and all look reasonable. Outliers with very large values (>100%) nearly all occur within the Owyhee River canyon. The only area that appears suspect is a single pixel with a value of 610 that actually occurs outside of the study area boundary. Even if it were within the study area, the likelihood of a single pixel negatively influencing the rule sets or map unit delineation is almost nil. The LandSat TM 4/3 Band Ratio data set also shows several outliers. This band ratio (near infrared / red) is a useful vegetation index since living vegetation is the only terrestrial material that both absorbs red light and reflects near infrared light (Jensen, 2005). Lighter tones (higher reflectance values) indicate that more healthy vegetation is present. Outliers occur primarily on irrigated farmland (high values), but some also occur on Antelope Reservoir, Upper and Lower Cow Lakes, along the Owyhee River, and on parts of the Jordan Craters lava flow (low values). With the exception of the single data point in the slope data set, all outliers appear to be legitimate and meaningful data points that might have significant importance for predicting certain soil types or characteristics.

Descriptive statistics readily apparent from the box plots are consistent with those calculated from the histogram using the Geostatistical Analyst in ArcMap. All statistics appear reasonable in context of their extent and the landscape characteristics. As expected, data sets with outliers have high kurtosis values (>2.5).

Appendix IV: Predicted Soils Maps, Sampling Points, Accuracy Summaries and Soil Map Unit Descriptions for Early Runs
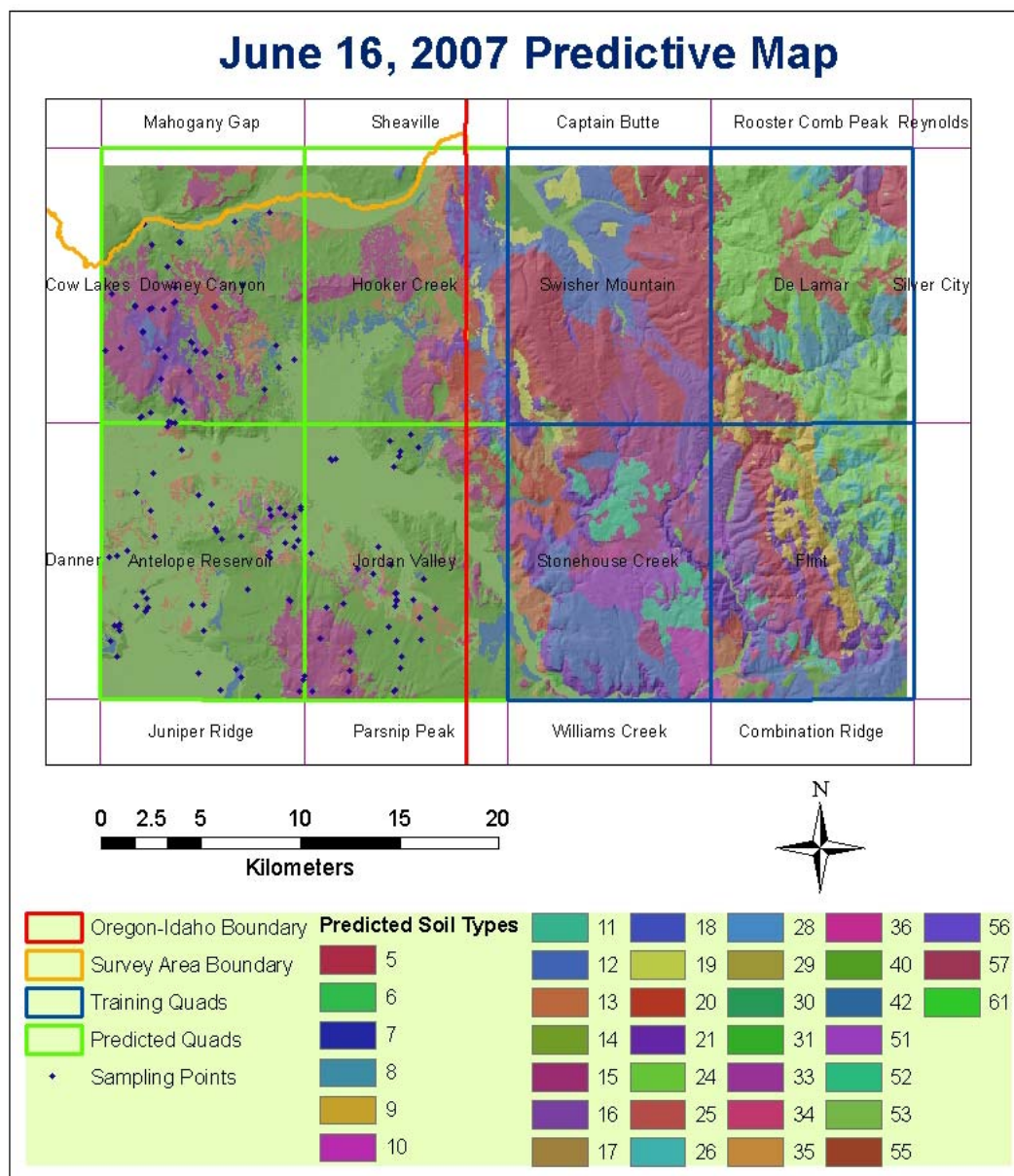
Figure A4-1: June 16, 2007 Predictive Map

Table A4-1: Accuracy Summary for June 16, 2007 Predictive Map

	Order	Suborder	Great Group	Subgroup	Map Unit
Jordan Valley	74%	59%	33%	Not considered	Not Considered
Antelope Reservoir	59%	57%	34%	Not considered	Not considered
Downey Canyon	68%	66%	36%	Not considered	Not considered
Overall	67%	61%	35%	Not considered	Not considered

Table A4-2: Soil Map Units for June 16, 2007 Predictive Map

Soil Type Code	Classification (Great Group)
5	Argixerolls
6	Argixerolls, Camborthids
7	Argixerolls, Cryaqualfs, Durixerolls
8	Argixerolls, Cryoborolls, Haploxerolls
9	Argixerolls, Cryoborrols, rock outcrop
10	Argixerolls, Durargids
11	Argixerolls, Durixeralfs
12	Argixerolls, Durixerolls
13	Argixerolls, Haplargids
14	Argixerolls, Haplargids, Palexerolls
15	Argixerolls, Haploxeralfs
16	Argixerolls, Haploxerolls
17	Argixerolls, Haploxerolls, barren lacustrine deposits
18	Argixerolls, Haploxerolls, Haplargids
19	Argixerolls, Paleargids, Palexerolls
20	Argixerolls, Palexerolls
21	Argixerolls, rock outcrop
24	Cryoborolls
25	Cryoborolls, Haplargids
26	Cryoborolls, Haplargids, Haploxeralfs
28	Cryoborolls, Haploxeralfs, rock outcrop
29	Cryoborolls, rock outcrop
30	Cryorthents, Haploxerolls
31	Durargids, Haplargids
33	Durixeralfs, Durixerolls
34	Durixeralfs, Durixerolls, Haploxeralfs
35	Durixeralfs, Haplargids
36	Durixeralfs, Haploxeralfs
40	Durixerolls
42	Endoaquolls, Haploxerolls
51	Haploxeralfs, Torriorthents
52	Haploxeralfs, Torriorthents, rock outcrop
53	Haploxerolls
55	Haploxerolls, rock outcrop
56	Lake and stream deposits of sand, gravel, cobbles and stones
57	Open excavations and waste rock piles
61	Water
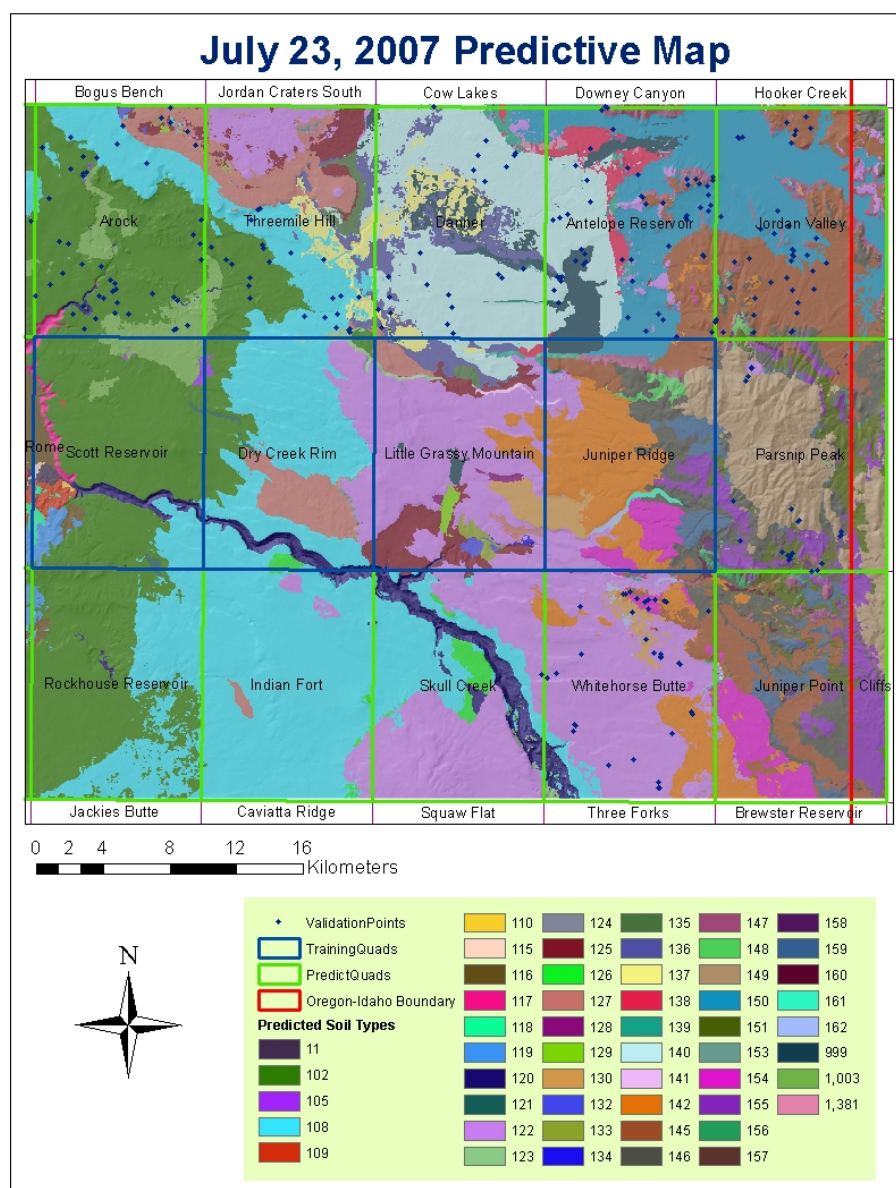
Figure A4-2: July 23, 2007 Predictive Map

Table A4-3: Accuracy Summary for July 23, 2007 Predictive Map

	Order	Suborder	Great Group	Subgroup	Map Unit
Whitehorse Butte	83%	83%	74%	71%	17%
Parsnip Peak	100%	100%	87%	87%	4%
Jordan Valley	72%	59%	51%	49%	0%
Antelope Reservoir	91%	89%	84%	82%	18%
Danner	86%	71%	67%	62%	60%
Threemile Hill	91%	46%	46%	27%	18%
Arock	88%	56%	54%	2%	0%
Overall	87%	73%	67%	53%	14%
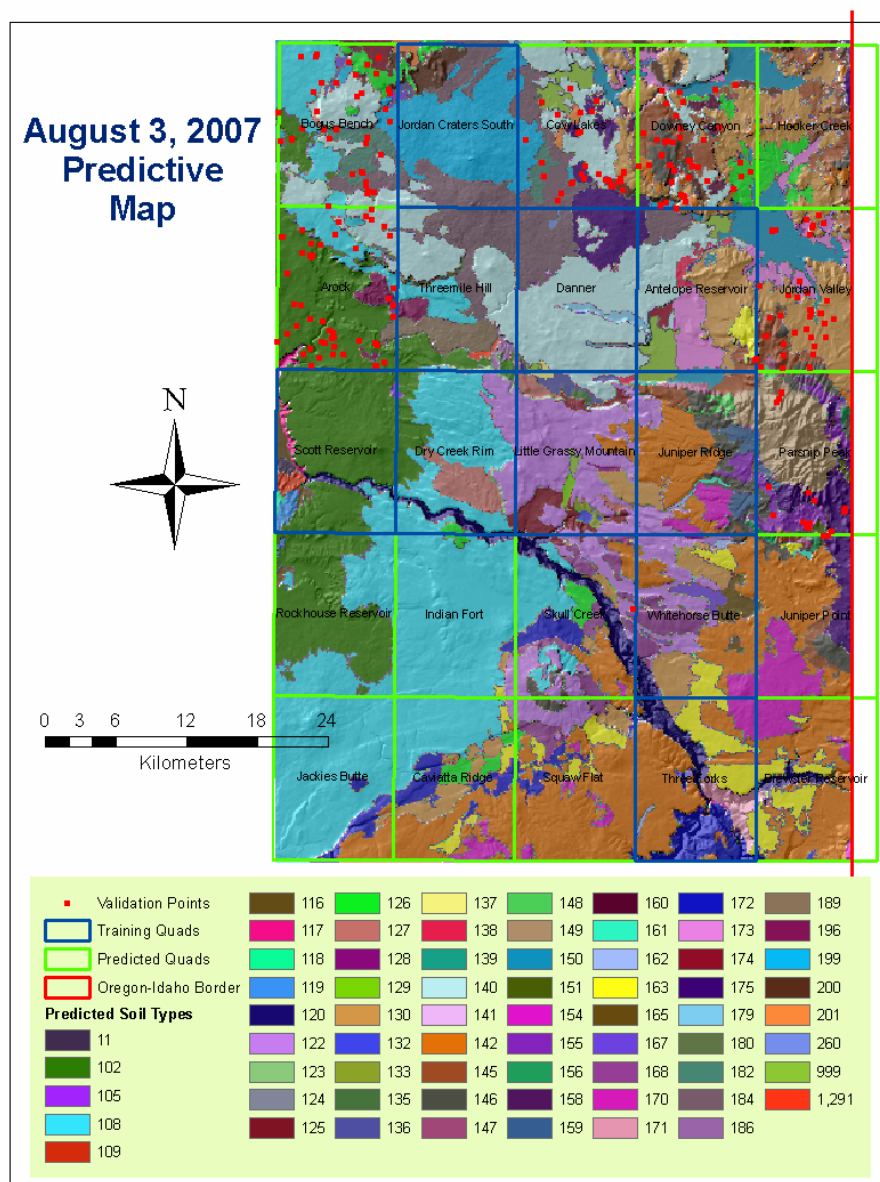
Figure A4-3: August 3, 2007 Predictive Map

Table A4-4: Accuracy Summary for August 3, 2007 Predictive Map

	Order	Suborder	Great Group	Subgroup	Map Unit
Arock	98%	66%	64%	14%	8%
Bogus Bench	85%	71%	66%	39%	20%
Cow Lakes	58%	49%	49%	32%	19%
Downey Canyon	79%	67%	30%	25%	0%
Jordan Valley	69%	56%	54%	44%	17%
Antelope Reservoir	71%	69%	44%	25%	20%
Parsnip Peak	100%	100%	87%	65%	0%
Overall	80%	67%	55%	32%	13%

Table A4-5: Soil Map Units for July 23, 2007 and August 3, 2007 Predictive Runs

Soil Type Code	Classification (Subgroup)
102	Durinodic Xeric Haplargids
105	Cambidic Durixerolls
108	Abruptic Xeric Argidurids
109	Xeric Haplargids, Vitrixerandic Paleargids
110	Durinodic Xeric Haplargids
115	Xeric Torrifluvents
116	Xeric Haplocambids, Vitrandic Argidurids
117	Vitrixerandic Haplargids
118	Durinodic Haplocalcids
119	Durinodic Xeric Paleargids
120	Rock outcrop, Rubbleland
121	Xeric Haplocambids
122	Vitritorrandic Argixerolls
123	Xeric Haplargids, rock outcrop
124	Vitrixerandic Haplargids, rock outcrop
125	Abruptic Xeric Argidurids
126	Vitrandic Palexerolls
127	Xeric Haplargids
128	Aridic Argixerolls
129	Fluventic Haploxerolls, Vitritorrandic Haploxerolls
130	Vitritorrandic Durixerolls
132	Cumulic Haploxerolls, Xerertic Argialbolls
133	Xerollic Paleargids
134	Typic Haplargids, Typic Palexerolls
135	Xeric Argidurids
136	Vitritorrandic Argixerolls
137	Aquandic Palexeralfs
138	Abruptic Xeric Argidurids
139	Vitrandic Argixerolls, Xeric Haplargids, rock outcrop
140	Abruptic Xeric Argidurids
141	Vitrandic Palexerolls, rock outcrop
142	Vitritorrandic Argixerolls
145	Xeric Argidurids
146	Vitrandic Argixerolls
147	Vitrandic Argixerolls, rock outcrop
148	Lithic Haploxerolls, Vitrandic Argixerolls
149	Vitrandic Argixerolls
150	Abruptic Xeric Argidurids, Durinodic Xeric Paleargids, Aridic Argixerolls
151	Vitrandic Argixerolls

Table A4-5 Continued: Soil Map Units for July 23, 2007 and August 3, 2007
Predictive Runs

Soil Type Code	Classification (Subgroup)
153	Vertic Argiaquolls
154	Vitritorrandic Argixerolls, Abruptic Xeric Argidurids
155	Abruptic Argiduridic Durixerolls
156	Aridic Argixerolls, Vitrandic Argixerolls, rock outcrop
157	Lithic Haploxerolls
158	Vitritorrandic Argixerolls, Abruptic Xeric Argidurids
159	Vitrandic Argixerolls, Abruptic Xeric Argidurids
160	Vitrandic Argixerolls, Aquandic Endoaquolls
161	Vitrandic Haploxerolls
162	Vitrandic Haploxerolls
163	Vitritorrandic Argixerolls
165	Vitritorrandic Argixerolls, Aquic Paleargids
167	Xeric Epiaquerts, Aquandic Palexeralfs
168	Vitritorrandic Argixerolls
170	Cumulic Haploxerolls, Vitritorrandic Haploxerolls
171	Vitrandic Palexerolls
172	Aridic Argixerolls, rock outcrop, Vitrandic Argixerolls
173	Vitritorrandic Argixerolls
174	Durinodic Xeric Paleargids, Playa, Aquandic Endoaquolls
175	Vitrixerandic Argidurids, Vitritorrandic Argixerolls, lava flows
179	Vitrixerandic Haplocambids, Vitrandic Palexerolls
180	Vitrixerandic Haplargids, Aquandic Palexeralfs
182	Vitrixerandic Argidurids
184	Lava flows, Vitrixerandic Argidurids
186	Vitrixerandic Haplocambids
189	Vitrixerandic Argidurids, lava flows, Typic Haplargids
196	Durinodic Xeric Haplocambids, Durinodic Xeric Haplargids
199	Abruptic Xeric Argidurids, lava flows
200	Vitritorrandic Argixerolls, Aridic Argixerolls
201	Abruptic Xeric Argidurids
260	Vitritorrandic Argixerolls, Cumulic Haploxerolls, Aquandic Palexeralfs
999	Water
1003	Private land - not yet mapped
1291	Fluventic Haploxerolls, Vitritorrandic Haploxerolls
1381	Abruptic Xeric Argidurids
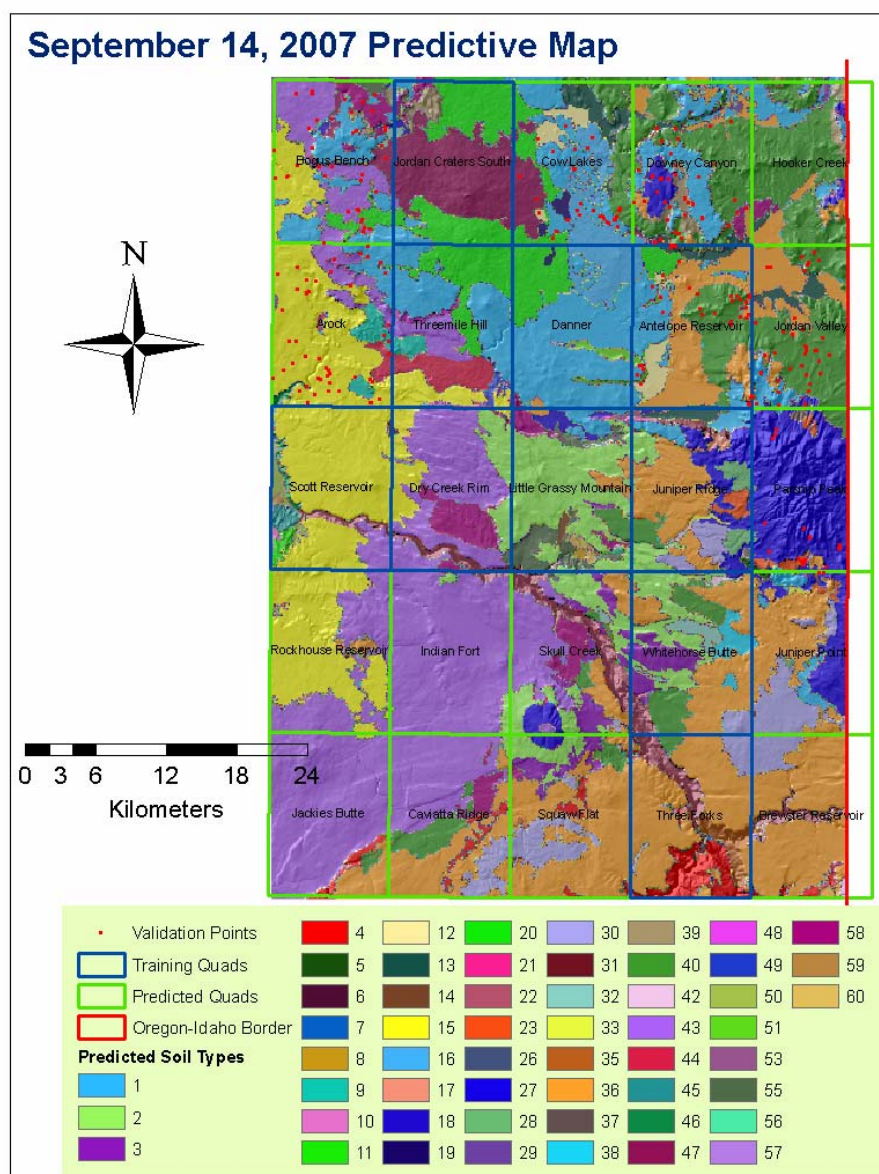
Figure A4-4: September 14, 2007 Predictive Map

Table A4-6: Accuracy Summary for September 14, 2007 Predictive Map

	Order	Suborder	Great Group	Subgroup	Map Unit
Arock	92%	67%	65%	12%	Not Considered
Bogus Bench	83%	63%	56%	29%	Not Considered
Cow Lakes	71%	47%	47%	29%	Not Considered
Downey Canyon	73%	61%	30%	30%	Not Considered
Jordan Valley	72%	62%	28%	26%	Not Considered
Antelope Reservoir	68%	68%	50%	32%	Not Considered
Parsnip Peak	100%	100%	83%	83%	Not Considered
Overall	79%	65%	50%	30%	Not Considered

Table A4-7: Soil Map Units for September 14, 2007 Predictive Run

Soil Type Code	Classification (Subgroup)
1	Abruptic Xeric Argidurids
2	Abruptic Xeric Argidurids, Vitritorrandic Argixerolls
3	Aridic Argixerolls
4	Aridic Argixerolls, rock outcrop, Vitritorrandic Argixerolls
5	Aridic Argixerolls, Vitrandic Argixerolls, Rock Outcrop
6	Cambic Durixerolls
7	Cumulic Haploxerolls, Vitritorrandic Haploxerolls
8	Cumulic Haploxerolls, Xerertic Argialbolls
9	Durinodic Haplocambids, Typic Argidurids
10	Durinodic Xeric Haplocambids
11	Durinodic Xeric Paleargids
12	Durinodic Xeric Paleargids, Playa, Aquandic Endoaquolls
13	Durinodic Xeric Paleargids, Vitritorrandic Argixerolls, Xeric Argidurids
14	Fluventic Haploxerolls, Cumulic Vitritorrandic Haploxerolls
15	Haploxeralfic Argidurids
16	Lithic Haplargids, Vitritorrandic Argixerolls, rock outcrop
17	Lithic Haploxerolls, Vitrandic Argixerolls
18	Pachic Argixerolls
19	Pachic Argixerolls, Vitritorrandic Argixerolls
20	Lava Flows, Vitrixerandic Argidurids
21	Rock outcrop, Pachic Argixerolls, Xeric Haplargids
22	Rock outcrop, rubbleland
23	Shallow Vitrandic Durixerolls
26	Vertic Argixerolls, rock outcrop
27	Vitrandic Argixerolls
28	Vitrandic Argixerolls, Abruptic Xeric Argidurids
29	Vitrandic Argixerolls, Cumulic Endoaquolls
30	Vitrandic Argixerolls, Lithic Xeric Haplargids
31	Vitrandic Argixerolls, rock outcrop
32	Vitrandic Haploxeralfs
33	Vitrandic Haploxeralfs (Cumulic and Fluventic)
35	Vitrandic Palexerolls
36	Vitritorrandic Argixerolls
37	Vitritorrandic Argixerolls, Abruptic Xeric Argidurids
38	Vitritorrandic Argixerolls, Aquic Paleargids
39	Vitritorrandic Argixerolls, Aridic Argixerolls
40	Vitritorrandic Durixerolls
42	Vitritorrandic Haploxerolls
43	Vitrixerandic Argidurids

Table A4-7 Continued: Soil Map Units for September 14, 2007 Predictive Run

Soil Type Code	Classification (Subgroup)
44	Vitrixerandic Argidurids, rock outcrop, Haploxeralfic Haplargids
45	Vitrixerandic Argidurids, Xeric Haplargids
46	Vitrixerandic Haplargids
47	Vitrixerandic Haplargids, rock outcrop
48	Vitrixerandic Haplargids, Vitrixerandic Argidurids
49	Vitrixerandic Haplocambids, rock outcrop
50	Vitrixerandic Haplocambids, Haploxeralfic Argidurids
51	Vitrixerandic Haplocambids, Vitritorrandic Argixerolls
53	Abruptic Xeric Argidurids
55	Xeric Haplargids, Durinodic Xeric Haplargids
56	Xeric Haplargids, Vitrandic (Aridic) Palexerolls
57	Xeric Haplocambids
58	Xeric Paleargids
59	Xeric Torrifluvents
60	Xerollic Paleargids

Appendix V: Map Unit Descriptions for the Threemile Hill Area

Table A5-1: Map Unit Descriptions for the Threemile Hill Area

Map Unit	Name	%	Classification
Soil Map Unit Summary, Threemile Hill Area			
102C	Muni gravelly ashy loam, 1 to 15% slopes	90	Loamy, mixed, superactive, mesic shallow Haploxeralfic Argidurids
105B	Garnet loamy sand, 1 to 6% slopes	85	Coarse-loamy, mixed, superactive, mesic Cambidic Durixerolls
108C	Sandhollow ashy sandy loam, 2 to 15% slopes	85	Loamy, mixed, superactive, mesic shallow Vitrixerandic Argidurids
120D	Rock outcrop-Rubbleland complex, steep	80	rock outcrop
		15	rubbleland
121B	Drewsey ashy silt loam, 1 to 4% slopes	85	Coarse-loamy, mixed, superactive, mesic Xeric Haplocambids
122C	Foleylake-Martinson complex, 2 to 15% slopes	60	Fine, smectitic, frigid Abruptic Xeric Argidurids
		25	Fine, smectitic, frigid Vitritorrandic Argixerolls
123D	Morfittash-rock outcrop complex, 15 to 50% slopes	70	Fine-loamy, mixed, superactive, mesic Vitrixerandic Haplargids
		20	rock outcrop
124D	Hardtrigger-Rock outcrop complex, 15 to 50% slopes	65	Fine, smectitic, mesic Vitrixerandic Haplargids
		20	rock outcrop
125C	Suncold-Catchell complex, 2 to 15% slopes	55	Clayey-skeletal, smectitic, mesic, shallow Abruptic Argidurids
		35	Fine, smectitic, mesic Abruptic Argidurids
126C	Bigflat stony ashy silt loam, 2 to 15% slopes	85	Fine, smectitic, frigid, Xeric Paleargids

Table A5-1 Continued: Map Unit Descriptions for the Threemile Hill Area

Map Unit	Name	%	Classification
127C	Toney ashy silt loam, 2 to 15% slopes	85	Fine, smectitic, mesic Xeric Paleargids
129A	Beetville-like-Koosharem complex, 1 to 4% slopes	70	Fine-loamy, mixed, superactive, frigid Fluventic Haploxerolls
		20	Fine-loamy, mixed, superactive, frigid Vitritorrandic Haploxerolls
136B	Wisher ashy silt loam, 1 to 5% slopes	85	Fine-silty, mixed , superactive, mesic Pachic Argixerolls
137A	Upcreek-like ashy silt loam, <2% slopes	90	Fine-silty, mixed, superactive, mesic Vitritorrandic Argixerolls
140C	Midraw very stony ashy loam, 2 to 15% slopes	85	Clayey, smectitic, mesic, shallow Abruptic Xeric Argidurids
167A	Jacares-Silverash complex, <3% slopes	75	Fine, smectitic, frigid Vitrandic Haploxeralfs
		15	Fine, smectitic, frigid Vitrandic Haploxeralfs
175C	Cocklebur-Wisher-Rock outcrop complex, 2 to 15% slopes	40	Loamy, mixed, superactive, mesic, shallow Vitrixerandic Haplargids
		30	Fine-silty, mixed superactive, mesic Vitritorrandic Argixerolls
		20	rock outcrop
179D	Felcher-Patron complex, 15 to 35% slopes	50	Loamy-skeletal, mixed, superactive, mesic Vitrixerandic Haplocambids
		35	Fine, smectitic, mesic Vitritorrandic Argixerolls
182C	Clarksbutte cobbly ashy silt loam, 2 to 15% slopes	90	Fine, loamy, mixed, superactive, mesic Vitrixerandic Argidurids
184C	Rock outcrop-Clarksbutte complex, 2 to 15% slopes	65	rock outcrop
		25	Fine-loamy, mixed, superactive, mesic Vitrixerandic Argidurids
186C	Felcher loam, 15 to 50% slopes	85	Fine-loamy, mixed, superactive, mesic Vitrixerandic Haplocambids
187C	Arock-Drewsey complex, 2 to 15% slopes	50	Coarse-loamy, mixed, superactive, mesic Vitrixerandic Haplocambids
		40	Coarse-loamy, mixed, superactive, mesic Vitrixerandic Haplocambids

Table A5-1 Continued: Map Unit Descriptions for the Threemile Hill Area

Map Unit	Name	%	Classification
189B	Clarksbutte-Rock outcrop-Rimview complex, 1 to 8% slopes	40	Fine-loamy, mixed, superactive, mesic Vitrixerandic Argidurids
		30	rock outcrop
		15	Fine-loamy, mixed, superactive, mesic Typic Haplargids
190C	Tribeca fine sandy loam, 10 to 25% slopes	85	Coarse-loamy, mixed, superactive, mesic Typic Haplargids
191C	Muni dry-Drewsey Dry complex, 2 to 15% slopes	70	Loamy, mixed, superactive, mesic shallow Vitrixerandic Argidurids
		20	Fine-loamy, mixed, superactive, mesic Vitrixerandic Haplargids
196C	Sagehill-Muni dry complex, 10-25% slopes	60	Coarse-loamy, mixed, superactive, mesic Durinodic Haplocambids
		25	Fine-loamy, mixed, superactive, mesic Typic Argidurids
199B	Clarksbutte-Rock outcrop complex, 1 to 8% slopes	80	Fine-loamy, mixed, superactive, mesic Vitrixerandic Argidurids
		15	rock outcrop
200D	Zymans-like-Barbermill complex, 15 to 40% slopes	50	Fine, smectitic, frigid Vitritorrandic Argixerolls
		35	Clayey, smectitic, mesic, shallow Aridic Argixerolls
201B	Craterlak ashy silt loam, 1 to 8% slopes	85	Fine, smectitic, mesic Xeric Argidurids
204A	Silverash-like ashy loam, <2% slopes	90	Fine, smectitic, frigid Typic Palexeralfs
224	Wisher silty clay loam, 0 to 2 percent slopes	80	Fine-loamy, mixed, superactive, mesic Vitritorrandic Argixerolls
261C	Crunchie-Tribeca complex, 2-15% slopes	70	Fine-loamy, mixed, superactive, mesic Durinodic Haplargids
		20	Fine-loamy, mixed, superactive, mesic typic Xeric Haplargids
266B	Skinnerpit-Lava flows-Cheatroad complex, 2-15% slopes	45	Coarse-loamy, mixed, superactive, mesic Xeric Haplocambids
		25	lava flows
		15	Coarse-loamy, mixed, superactive, mesic Xeric Haplodurids

Table A5-1 Continued: Map Unit Descriptions for the Threemile Hill Area

Map Unit	Name	%	Classification
273	Orovada-Zevadez,dry-Nevador complex, 4 to 18 percent slopes	35	Coarse-loamy, mixed, superactive, mesic Durinodic Xeric Haplocambids
		30	Fine-loamy, mixed, superactive, mesic Durinodic Xeric Haplargids
		20	Fine-loamy, mixed, superactive, mesic Durinodic Xeric Haplargids
998	lava flows	98	lava flows
999	water	100	water

Appendix VI: Directions for Generating Confusion Matrices with ENVI;
Confusion Matrices for October 12, 2007 and April 10, 2008 Predictive Maps

Directions for Generating Confusion Matrices with ENVI 4.3

1. Convert your maps for the training area (mapped soils and predicted soils, hereafter referred to as MAPPED and PREDICTED) to TIFF files (.tif).
2. Start ENVI, and Load MAPPED (File → Open External File → Generic Formats → TIFF/GeoTIFF).
3. Click on Band1 and "Load Band" to display the map.
4. Use Tools → Color Mapping → ENVI Color Tables to find a color combination that highlights most of your map units. As you try to locate different units, you may have to play with different color schemes and the "Stretch Bottom" and "Stretch Top" options to get certain units to show up.
5. Display MAPPED in ArcMap to use for reference.
6. In ENVI, in your Band1 window, open the ROI tool (Tools → Region of Interest → ROI Tool).
7. Before you do anything else, for "Window", select Off. This option allows you to choose which window to use to delineate a region of interest (ROI), so if a window is activated and you click in it (ex. to navigate), it will start drawing an ROI.
8. You will want to create an ROI for every map unit type. Click on "New Region" to generate one region for each soil class. You can then click on the "ROI Name" to rename each ROI with your soil class/type. I prefer to create new regions and name them for all of my map units before I start delineating ROIs. (**You only need ONE ROI for each soil class—not one within every mapping unit of that class).
9. First, look at MAPPED in ArcMap. Use the identify tool to locate a mapping unit for your first soil class. Visually locate the same unit in your ENVI window. (**Note: You have three windows in ENVI for looking at your data. The "Scroll" window shows the whole dataset. A red square in that window delineates what is shown, closer-up, in the "Image" window. A red square in the "Image" window delineates what is shown in the "Zoom" window.
10. To delineate an ROI, first make sure the correct ROI name is highlighted in the ROI Tool window. Then, select the window you wish to use to delineate the ROI. You can use the Image window for large units, but in most cases you will want to use the Zoom window. Use the left mouse button to place vertices for your ROI. Right-click to complete the polygon, and then right-click again to accept the ROI placement. As long as the ROI is completely within your unit of interest, it can be any shape you wish. Keep the size to a minimum! ENVI only needs a few pixels to classify the image in this case, and the smaller the polygon the shorter the processing time will be. When you finish the ROI, switch your window to "OFF" so you can navigate to the next soil unit. When you're

finished, you can save your ROI's. In the ROI tool window, click File →
Save ROIs.

11. Repeat this process for every map unit type on the map. This should be a quick
process.

12. When finished delineating the ROI's, you will classify your image. In the main
menu bar, go to Classification → Supervised → Parallelepiped. Select
MAPPED as the input and click OK. In the "Select Classes from Regions" box,
click "Select All Items". Output Result to File, and click "Choose" to navigate
to the appropriate folder and name your file. You can click the toggle to turn
"Output Rule Image?" to "No". It may take the classification several minutes
to run, depending on the size of your area and the size of your ROI's.

13. After the classification is finished, load the band and look at it—the classified
unit boundaries should match exactly with your original MAPPED boundaries.

14. Close ENVI and then start a new session. Load your PREDICTED .tif image
and display. Do the exact same thing with the PREDICTED map—define
ROI's and then run the classifier.

15. Once both images are classified, the pixel values can be compared to generate
the error matrix. To do this, in the main menu bar, click on Classification →
Post-Classification → Confusion Matrix → Using Ground Truth Image. For
"Select Input File" select your CLASSIFIED PREDICTED image. If it isn't
currently loaded in ENVI, you can click Open → New File and navigate to it
(open the file with the correct name, but with no extension). Click OK. For the
"Ground Truth Input File" select your CLASSIFIED MAPPED image. Click
OK. A window opens called "Match Classes Parameters". Here, you just have
to pair the equivalent classes (easy if you named the map units the same way on
both maps). Select the "ground truth class" that matches the "classification
image class" and then click the "Add Combination" button. Match up all of the
classes. Click OK. In the Confusion Matrix Parameters window, you can
choose to generate an Error Image, which will show you the location of all the
incorrectly classified pixels. Click the "choose" button to navigate to a folder
and name the output file. Click OK, and the Confusion Matrix is generated.

Notes

-If there are discrepancies between an original map and a classified map, there was
probably an ROI with some noisy pixels of a different class included. You should use a
cleaned/filtered version of your predicted map for the confusion matrix, and be careful
when delineating the ROI's.

-All map units have to be matched to an analogous unit on the other map to be included in
the confusion matrix. If map units with a small areal extent are present on one image but
not the other, you may choose to either not delineate an ROI for that area (if it's only a few
pixels), or to mask the area on both images as NoData (if it is a bit larger and may
significantly impact the statistics).

Table A6-1: Confusion Matrix for Training Data, October 12, 2007 Run

```
Confusion Matrix: H:\Project\ErrorMatrix\PredictClassify

Overall Accuracy = (1399029/1420580)  98.4829%
Kappa Coefficient = 0.9755
```

Ground Truth (Percent)

Class	Unclass	102	105	108	120
Unclass	99.93	0.06	0.00	0.09	0.53
102	0.02	97.51	41.67	0.46	0.00
105	0.00	0.02	58.33	0.00	0.00
108	0.01	1.60	0.00	97.03	1.73
120	0.00	0.00	0.00	0.41	96.29
121	0.00	0.34	0.00	0.02	0.00
122	0.00	0.00	0.00	0.45	0.00
123	0.00	0.03	0.00	0.11	0.00
124	0.00	0.00	0.00	0.03	0.00
125	0.00	0.00	0.00	0.05	0.00
126	0.01	0.00	0.00	0.07	0.00
127	0.00	0.29	0.00	0.65	0.00
136	0.00	0.00	0.00	0.00	0.00
137	0.00	0.00	0.00	0.00	0.00
140	0.00	0.00	0.00	0.12	0.00
167	0.00	0.00	0.00	0.00	0.00
175	0.01	0.00	0.00	0.00	0.00
179	0.00	0.00	0.00	0.00	0.00
182	0.00	0.00	0.00	0.00	0.00
184	0.01	0.00	0.00	0.00	0.00
186	0.00	0.00	0.00	0.07	0.00
187	0.00	0.00	0.00	0.01	0.00
189	0.00	0.00	0.00	0.43	0.25
190	0.00	0.00	0.00	0.00	0.00
191	0.00	0.00	0.00	0.00	0.89
196	0.00	0.15	0.00	0.00	0.30
199	0.01	0.00	0.00	0.00	0.00
200	0.00	0.00	0.00	0.00	0.00
201	0.00	0.00	0.00	0.00	0.00
261	0.00	0.00	0.00	0.00	0.00
266	0.00	0.00	0.00	0.00	0.00
204	0.00	0.00	0.00	0.00	0.00
273	0.00	0.00	0.00	0.00	0.00
998	0.00	0.00	0.00	0.00	0.00
999	0.00	0.00	0.00	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Table A6-1 Continued: Confusion Matrix for Training Data,
October 12, 2007 Run

Ground Truth (Percent)

Class	121	122	123	124	125
Unclass	0.00	0.35	0.00	0.00	1.51
102	29.55	0.00	6.22	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00
108	0.60	0.36	1.43	4.01	0.16
120	0.00	0.00	0.00	0.00	0.20
121	65.91	0.00	0.00	0.00	0.00
122	0.00	98.25	5.81	11.50	0.79
123	0.00	0.84	86.54	0.00	0.00
124	0.00	0.12	0.00	83.69	0.00
125	0.00	0.08	0.00	0.00	95.91
126	0.00	0.00	0.00	0.00	0.00
127	3.95	0.00	0.00	0.80	0.00
136	0.00	0.00	0.00	0.00	0.00
137	0.00	0.00	0.00	0.00	0.00
140	0.00	0.00	0.00	0.00	1.11
167	0.00	0.00	0.00	0.00	0.00
175	0.00	0.00	0.00	0.00	0.00
179	0.00	0.00	0.00	0.00	0.08
182	0.00	0.00	0.00	0.00	0.00
184	0.00	0.00	0.00	0.00	0.00
186	0.00	0.00	0.00	0.00	0.00
187	0.00	0.00	0.00	0.00	0.00
189	0.00	0.00	0.00	0.00	0.00
190	0.00	0.00	0.00	0.00	0.00
191	0.00	0.00	0.00	0.00	0.00
196	0.00	0.00	0.00	0.00	0.00
199	0.00	0.00	0.00	0.00	0.00
200	0.00	0.00	0.00	0.00	0.00
201	0.00	0.00	0.00	0.00	0.00
261	0.00	0.00	0.00	0.00	0.00
266	0.00	0.00	0.00	0.00	0.00
204	0.00	0.00	0.00	0.00	0.00
273	0.00	0.00	0.00	0.00	0.00
998	0.00	0.00	0.00	0.00	0.00
999	0.00	0.00	0.00	0.00	0.24
Total	100.00	100.00	100.00	100.00	100.00

Table A6-1 Continued: Confusion Matrix for Training Data,
October 12, 2007 Run

Ground Truth (Percent)

Class	126	127	136	137	140
Unclass	0.00	0.08	0.00	0.15	0.09
102	0.00	0.52	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00
108	0.45	0.84	0.00	0.00	0.01
120	1.08	0.00	0.00	0.00	0.00
121	0.00	0.07	0.00	0.00	0.00
122	0.00	0.03	0.00	0.00	0.00
123	0.00	0.09	0.00	0.00	0.00
124	0.00	0.64	0.00	0.00	0.00
125	0.00	0.00	0.00	0.00	0.36
126	80.02	0.00	0.00	0.00	0.00
127	0.00	96.71	0.00	0.00	0.00
136	0.00	0.00	91.10	0.00	0.11
137	0.00	0.00	0.00	95.17	0.01
140	0.00	1.00	4.20	4.69	96.89
167	0.00	0.00	1.34	0.00	0.00
175	0.00	0.00	0.00	0.00	0.43
179	0.00	0.00	0.00	0.00	0.32
182	0.00	0.00	0.00	0.00	0.00
184	0.00	0.00	0.11	0.00	0.65
186	0.00	0.00	0.00	0.00	0.19
187	0.00	0.00	0.00	0.00	0.00
189	0.00	0.00	0.00	0.00	0.00
190	0.00	0.00	0.00	0.00	0.00
191	0.00	0.00	0.00	0.00	0.00
196	0.00	0.00	0.00	0.00	0.00
199	0.00	0.00	0.22	0.00	0.00
200	5.00	0.00	0.58	0.00	0.00
201	0.21	0.00	0.00	0.00	0.00
261	0.00	0.00	0.00	0.00	0.00
266	0.00	0.00	0.00	0.00	0.25
204	0.00	0.00	0.00	0.00	0.00
273	0.00	0.00	0.00	0.00	0.00
998	13.24	0.00	0.33	0.00	0.19
999	0.00	0.02	2.13	0.00	0.50
Total	100.00	100.00	100.00	100.00	100.00

Table A6-1 Continued: Confusion Matrix for Training Data,
October 12, 2007 Run

Ground Truth (Percent)

Class	167	175	179	182	184
Unclass	0.00	0.00	0.00	1.85	0.04
102	0.00	0.00	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00
108	0.00	0.00	0.00	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00
121	0.00	0.00	0.00	0.00	0.00
122	0.00	0.00	0.00	0.00	0.00
123	0.00	0.00	0.00	0.00	0.00
124	0.00	0.00	0.00	0.00	0.00
125	0.00	0.00	0.75	0.00	0.00
126	0.00	0.00	0.00	0.00	0.00
127	0.00	0.00	0.00	0.00	0.00
136	0.21	0.00	0.00	0.00	0.02
137	0.00	0.00	0.00	0.00	0.00
140	0.00	0.66	4.71	0.00	0.35
167	98.95	0.00	0.00	0.00	0.00
175	0.00	97.49	0.00	0.00	0.07
179	0.00	0.00	94.46	0.00	0.00
182	0.00	0.00	0.00	82.88	0.05
184	0.00	0.26	0.00	4.38	99.36
186	0.00	0.00	0.00	0.00	0.00
187	0.00	0.00	0.00	0.00	0.00
189	0.00	0.00	0.00	0.00	0.00
190	0.00	0.00	0.00	0.00	0.00
191	0.00	0.00	0.00	0.00	0.00
196	0.00	0.00	0.00	0.00	0.00
199	0.84	0.00	0.00	10.44	0.11
200	0.00	0.00	0.00	0.00	0.00
201	0.00	0.00	0.00	0.00	0.00
261	0.00	0.00	0.00	0.00	0.00
266	0.00	1.58	0.00	0.00	0.00
204	0.00	0.00	0.00	0.00	0.00
273	0.00	0.00	0.00	0.00	0.00
998	0.00	0.00	0.00	0.45	0.00
999	0.00	0.00	0.07	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Table A6-1 Continued: Confusion Matrix for Training Data,
October 12, 2007 Run

Ground Truth (Percent)

Class	186	187	189	190	191
Unclass	0.00	0.00	0.00	1.54	4.57
102	0.00	0.00	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00
108	5.00	1.21	0.98	0.00	0.00
120	0.00	0.00	0.87	0.00	2.02
121	0.00	0.00	0.00	0.00	0.00
122	0.00	0.00	0.00	0.00	0.00
123	0.00	0.00	0.00	0.00	0.00
124	0.00	0.00	0.00	0.00	0.00
125	0.00	0.00	0.00	0.00	0.00
126	0.00	0.00	0.00	0.00	0.00
127	0.00	0.00	0.00	0.00	0.00
136	0.00	0.00	0.00	0.00	0.00
137	0.00	0.00	0.00	0.00	0.00
140	1.15	0.00	0.00	0.00	0.00
167	0.00	0.00	0.00	0.00	0.00
175	0.00	0.00	0.00	0.00	0.00
179	0.00	0.00	0.00	0.00	0.00
182	0.00	0.00	0.00	0.00	0.00
184	0.00	0.00	0.00	0.00	0.00
186	85.81	1.35	0.00	0.00	0.00
187	2.60	96.45	0.00	0.00	0.00
189	0.62	0.00	97.67	0.00	0.00
190	0.00	0.00	0.00	95.03	1.03
191	0.00	0.00	0.00	1.99	92.12
196	4.83	0.99	0.48	0.00	0.25
199	0.00	0.00	0.00	0.00	0.00
200	0.00	0.00	0.00	0.00	0.00
201	0.00	0.00	0.00	0.00	0.00
261	0.00	0.00	0.00	0.00	0.00
266	0.00	0.00	0.00	0.00	0.00
204	0.00	0.00	0.00	0.00	0.00
273	0.00	0.00	0.00	1.45	0.00
998	0.00	0.00	0.00	0.00	0.00
999	0.00	0.00	0.00	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Table A6-1 Continued: Confusion Matrix for Training Data,
October 12, 2007 Run

Ground Truth (Percent)

Class	196	199	200	201	261
Unclass	0.23	0.00	0.00	0.00	96.95
102	0.79	0.00	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00
108	0.00	0.00	0.00	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00
121	0.00	0.00	0.00	0.00	0.00
122	0.00	0.00	0.00	0.00	0.00
123	0.00	0.00	0.00	0.00	0.00
124	0.00	0.00	0.00	0.00	0.00
125	0.00	0.00	0.00	0.00	0.00
126	0.00	0.00	2.81	3.76	0.00
127	0.00	0.00	0.00	0.00	0.00
136	0.00	0.00	0.80	0.00	0.00
137	0.00	0.00	0.00	0.00	0.00
140	0.00	0.00	0.00	0.00	0.00
167	0.00	0.05	0.00	0.00	0.00
175	0.00	0.00	0.00	0.00	0.00
179	0.00	0.00	0.00	0.00	0.00
182	0.00	0.13	0.00	0.00	0.00
184	0.00	0.82	0.00	0.00	0.00
186	0.41	0.00	0.00	0.00	0.00
187	0.35	0.00	0.00	0.00	0.00
189	0.00	0.00	0.00	0.00	0.00
190	0.00	0.00	0.00	0.00	1.69
191	0.00	0.00	0.00	0.00	0.74
196	98.21	0.00	0.00	0.00	0.00
199	0.00	98.51	0.00	0.00	0.00
200	0.00	0.00	83.58	2.32	0.00
201	0.00	0.00	0.10	70.41	0.00
261	0.00	0.00	0.00	0.00	0.00
266	0.00	0.00	0.00	0.00	0.00
204	0.00	0.06	0.00	0.00	0.00
273	0.00	0.00	0.00	0.00	0.63
998	0.00	0.43	12.71	23.50	0.00
999	0.00	0.00	0.00	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Table A6-1 Continued: Confusion Matrix for Training Data,
October 12, 2007 Run

```
                    Ground Truth (Percent)
```

Class	266	204	273	998	999
Unclass	0.00	0.00	3.19	0.14	0.98
102	0.00	0.00	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00
108	0.00	0.00	0.00	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00
121	0.00	0.00	0.00	0.00	0.00
122	0.00	0.00	0.00	0.00	0.00
123	0.00	0.00	0.00	0.00	0.00
124	0.00	0.00	0.00	0.00	0.00
125	0.00	0.00	0.00	0.00	0.00
126	0.00	0.00	0.00	0.03	0.00
127	0.00	0.00	0.00	0.00	0.00
136	0.00	0.00	0.00	0.00	0.00
137	0.00	0.00	0.00	0.00	4.68
140	3.18	0.00	0.00	0.00	1.85
167	0.00	0.00	0.00	0.00	0.00
175	0.28	0.00	0.00	0.00	0.00
179	0.00	0.00	0.00	0.00	0.00
182	0.00	0.00	0.00	0.00	0.00
184	0.00	7.84	0.00	0.00	0.00
186	0.00	0.00	0.00	0.00	0.00
187	0.00	0.00	0.00	0.00	0.00
189	0.00	0.00	0.00	0.00	0.00
190	0.00	0.00	0.25	0.00	0.00
191	0.00	0.00	0.00	0.00	0.00
196	0.00	0.00	0.00	0.00	0.00
199	0.00	6.54	0.00	0.18	0.00
200	0.00	0.00	0.00	0.04	0.00
201	0.00	0.00	0.00	0.00	0.00
261	0.00	0.00	0.00	0.00	0.00
266	96.54	0.00	0.00	0.00	0.00
204	0.00	85.62	0.00	0.00	0.00
273	0.00	0.00	96.56	0.00	0.00
998	0.00	0.00	0.00	99.60	13.82
999	0.00	0.00	0.00	0.00	78.67
Total	100.00	100.00	100.00	100.00	100.00

Table A6-1 Continued: Confusion Matrix for Training Data,
October 12, 2007 Run

```
Class Areal Percentages
```

Class	Total
Unclass	60.42
102	1.59
105	0.02
108	7.50
120	0.75
121	0.05
122	1.34
123	0.11
124	0.08
125	0.21
126	0.33
127	1.49
136	0.19
137	0.05
140	7.84
167	0.04
175	0.98
179	0.20
182	0.22
184	3.64
186	0.34
187	0.17
189	1.20
190	0.08
191	0.45
196	1.44
199	4.75
200	0.42
201	0.11
261	0.00
266	0.64
204	0.02
273	0.11
998	3.12
999	0.10
Total	100.00

Table A6-1 Continued: Confusion Matrix for Training Data,
October 12, 2007 Run

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Unclass	0.43	0.07	3726/858306	625/855205
102	6.25	2.49	1410/22549	540/21679
105	1.65	41.67	4/242	170/408
108	1.36	2.97	1452/106599	3216/108363
120	7.42	3.71	794/10703	382/10291
121	17.14	34.09	114/665	85/836
122	3.81	1.75	727/19090	327/18690
123	18.94	13.46	296/1563	197/1464
124	16.61	16.31	187/1126	183/1122
125	17.10	4.09	498/2913	103/2518
126	8.62	19.98	400/4642	1059/5301
127	3.80	3.29	805/21187	693/21075
136	6.88	8.90	186/2704	246/2764
137	8.06	4.83	57/707	33/683
140	1.15	3.11	1275/111351	3530/113606
167	12.75	1.05	69/541	5/477
175	4.27	2.51	597/13968	344/13715
179	12.60	5.54	364/2889	148/2673
182	4.58	17.12	140/3055	602/3517
184	3.15	0.64	1631/51697	320/50386
186	8.38	14.19	408/4870	738/5200
187	9.26	3.55	219/2365	79/2225
189	3.29	2.33	563/17109	394/16940
190	10.40	4.97	122/1173	55/1106
191	2.12	7.88	135/6370	533/6768
196	2.13	1.79	435/20462	364/20391
199	0.86	1.49	578/67472	1014/67908
200	6.23	16.42	376/6032	1111/6767
201	2.45	29.59	38/1554	637/2153
261	0.00	100.00	0/0	2848/2848
266	5.42	3.46	496/9157	310/8971
204	14.10	14.38	43/305	44/306
273	2.16	3.44	34/1576	55/1597
998	6.17	0.40	2734/44277	165/41708
999	46.88	21.33	638/1361	196/919

Table A6-1 Continued: Confusion Matrix for Training Data,
October 12, 2007 Run

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Unclass	99.93	99.57	854580/855205	854580/858306
102	97.51	93.75	21139/21679	21139/22549
105	58.33	98.35	238/408	238/242
108	97.03	98.64	105147/108363	105147/106599
120	96.29	92.58	9909/10291	9909/10703
121	65.91	82.86	551/836	551/665
122	98.25	96.19	18363/18690	18363/19090
123	86.54	81.06	1267/1464	1267/1563
124	83.69	83.39	939/1122	939/1126
125	95.91	82.90	2415/2518	2415/2913
126	80.02	91.38	4242/5301	4242/4642
127	96.71	96.20	20382/21075	20382/21187
136	91.10	93.12	2518/2764	2518/2704
137	95.17	91.94	650/683	650/707
140	96.89	98.85	110076/113606	110076/111351
167	98.95	87.25	472/477	472/541
175	97.49	95.73	13371/13715	13371/13968
179	94.46	87.40	2525/2673	2525/2889
182	82.88	95.42	2915/3517	2915/3055
184	99.36	96.85	50066/50386	50066/51697
186	85.81	91.62	4462/5200	4462/4870
187	96.45	90.74	2146/2225	2146/2365
189	97.67	96.71	16546/16940	16546/17109
190	95.03	89.60	1051/1106	1051/1173
191	92.12	97.88	6235/6768	6235/6370
196	98.21	97.87	20027/20391	20027/20462
199	98.51	99.14	66894/67908	66894/67472
200	83.58	93.77	5656/6767	5656/6032
201	70.41	97.55	1516/2153	1516/1554
261	0.00	0.00	0/2848	0/0
266	96.54	94.58	8661/8971	8661/9157
204	85.62	85.90	262/306	262/305
273	96.56	97.84	1542/1597	1542/1576
998	99.60	93.83	41543/41708	41543/44277
999	78.67	53.12	723/919	723/1361

Table A6-2: Confusion Matrix for Threemile Hill Quad, October 12, 2007 Run

```
Confusion Matrix: H:\Project\ErrorMatrix\ThreemileErrorMatrix\ThreeMiPred

Overall Accuracy = (114727/155124)  73.9583%
Kappa Coefficient = 0.6693
```

Ground Truth (Percent)

Class	Unclass	102	108	123	127
Unclass	99.02	0.05	0.02	0.00	0.00
102	0.06	15.66	0.03	3.96	0.00
108	0.00	14.68	24.62	35.80	0.00
123	0.00	1.89	0.00	49.54	0.00
127	0.02	0.55	0.00	0.00	30.48
140	0.27	28.73	53.75	0.79	69.52
186	0.00	2.80	4.31	9.91	0.00
189	0.06	2.38	3.44	0.00	0.00
196	0.49	33.26	13.69	0.00	0.00
184	0.08	0.00	0.13	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Ground Truth (Percent)

Class	140	186	189	196	184
Unclass	0.01	0.00	0.04	0.45	0.02
102	0.00	0.14	0.00	0.00	0.00
108	0.00	0.14	0.02	0.00	0.06
123	0.00	0.00	0.00	0.00	0.00
127	2.83	0.00	0.00	0.00	0.14
140	93.51	44.57	4.13	0.00	4.04
186	1.32	42.98	1.50	4.16	0.00
189	0.03	0.24	84.79	6.26	6.45
196	0.00	11.08	9.53	89.12	0.00
184	2.30	0.87	0.00	0.00	89.29
Total	100.00	100.00	100.00	100.00	100.00

Class Areal Percentages

Class	Total
Unclass	4.06
102	1.87
108	4.21
123	0.46
127	1.07
140	41.32
186	2.89
189	12.85
196	8.92
184	22.34
Total	100.00

Table A6-2 Continued: Confusion Matrix for Threemile Hill Quad,
October 12, 2007 Run

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Unclass	0.79	0.98	50/6303	62/6315
102	1.55	84.34	45/2904	15395/18254
108	45.69	75.38	2982/6527	10852/14397
123	47.92	50.46	345/720	382/757
127	93.55	69.52	1553/1660	244/351
140	27.81	6.49	17824/64101	3214/49491
186	51.34	57.02	2304/4488	2898/5082
189	18.13	15.21	3615/19941	2929/19255
196	75.71	10.88	10471/13831	410/3770
184	3.49	10.71	1208/34649	4011/37452

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Unclass	99.02	99.21	6253/6315	6253/6303
102	15.66	98.45	2859/18254	2859/2904
108	24.62	54.31	3545/14397	3545/6527
123	49.54	52.08	375/757	375/720
127	30.48	6.45	107/351	107/1660
140	93.51	72.19	46277/49491	46277/64101
186	42.98	48.66	2184/5082	2184/4488
189	84.79	81.87	16326/19255	16326/19941
196	89.12	24.29	3360/3770	3360/13831
184	89.29	96.51	33441/37452	33441/34649

Table A6-3: Confusion Matrix for Training Data, April 10, 2008 Run

```
Confusion Matrix: H:\ThreemileReRun\ErrorMatrix\TrainingArea\PredClassify

Overall Accuracy = (1399979/1419539)  98.6221%
Kappa Coefficient = 0.9785
```

Ground Truth (Percent)

Class	Unclass	102	105	108	120
Unclass	99.98	0.00	0.00	0.00	0.00
102	0.00	97.59	13.50	0.40	0.00
105	0.00	0.10	85.19	0.00	0.00
108	0.00	0.87	0.00	98.56	0.85
120	0.00	0.00	0.00	0.21	97.33
122	0.00	0.35	0.00	0.14	0.00
123	0.00	0.06	0.00	0.00	0.00
124	0.00	0.00	0.00	0.01	0.00
125	0.00	0.00	0.00	0.02	0.00
126	0.00	0.00	0.00	0.05	0.08
127	0.00	0.30	0.00	0.23	0.00
136	0.00	0.00	0.00	0.00	0.00
140	0.00	0.00	0.00	0.05	0.00
167	0.00	0.00	0.00	0.00	0.00
175	0.00	0.00	0.00	0.00	0.00
179	0.00	0.00	0.00	0.00	0.00
182	0.00	0.00	0.00	0.00	0.00
184	0.00	0.00	0.00	0.00	0.00
186	0.00	0.00	0.00	0.08	0.00
187	0.00	0.00	0.00	0.00	0.00
188	0.00	0.00	0.00	0.26	0.44
189	0.00	0.00	0.00	0.00	0.00
190	0.00	0.00	0.00	0.00	0.00
196	0.00	0.09	0.00	0.00	0.38
199	0.00	0.00	0.00	0.00	0.00
200	0.00	0.00	0.00	0.00	0.00
201	0.00	0.00	0.00	0.00	0.00
204	0.00	0.00	0.00	0.00	0.00
213	0.00	0.00	0.00	0.00	0.00
221	0.00	0.00	0.00	0.00	0.00
222	0.00	0.00	0.00	0.00	0.00
224	0.01	0.43	1.31	0.00	0.00
261	0.00	0.20	0.00	0.00	0.00
265	0.00	0.00	0.00	0.00	0.00
266	0.00	0.00	0.00	0.00	0.00
273	0.01	0.00	0.00	0.00	0.00
998	0.00	0.00	0.00	0.00	0.00
999	0.00	0.00	0.00	0.00	0.00
191	0.00	0.00	0.00	0.00	0.92
Total	100.00	100.00	100.00	100.00	100.00

Table A6-3 Continued: Confusion Matrix for Training Data, April 10, 2008 Run

Ground Truth (Percent)

Class	122	123	124	125	126
Unclass	0.00	0.00	0.00	0.00	0.00
102	1.11	2.05	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00
108	1.36	6.21	6.68	2.99	1.16
120	0.00	0.00	0.00	0.37	0.03
122	97.07	7.31	5.49	0.41	0.00
123	0.10	84.43	0.00	0.00	0.00
124	0.30	0.00	84.35	0.00	0.00
125	0.01	0.00	0.00	90.20	0.00
126	0.00	0.00	0.00	0.66	87.32
127	0.05	0.00	3.48	0.00	0.00
136	0.00	0.00	0.00	0.00	0.00
140	0.00	0.00	0.00	5.37	1.16
167	0.00	0.00	0.00	0.00	0.00
175	0.00	0.00	0.00	0.00	0.00
179	0.00	0.00	0.00	0.00	0.00
182	0.00	0.00	0.00	0.00	0.00
184	0.00	0.00	0.00	0.00	0.00
186	0.00	0.00	0.00	0.00	0.00
187	0.00	0.00	0.00	0.00	0.00
188	0.00	0.00	0.00	0.00	0.00
189	0.00	0.00	0.00	0.00	0.00
190	0.00	0.00	0.00	0.00	0.00
196	0.00	0.00	0.00	0.00	0.00
199	0.00	0.00	0.00	0.00	0.00
200	0.00	0.00	0.00	0.00	10.33
201	0.00	0.00	0.00	0.00	0.00
204	0.00	0.00	0.00	0.00	0.00
213	0.00	0.00	0.00	0.00	0.00
221	0.00	0.00	0.00	0.00	0.00
222	0.00	0.00	0.00	0.00	0.00
224	0.00	0.00	0.00	0.00	0.00
261	0.00	0.00	0.00	0.00	0.00
265	0.00	0.00	0.00	0.00	0.00
266	0.00	0.00	0.00	0.00	0.00
273	0.00	0.00	0.00	0.00	0.00
998	0.00	0.00	0.00	0.00	0.00
999	0.00	0.00	0.00	0.00	0.00
191	0.00	0.00	0.00	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Table A6-3 Continued: Confusion Matrix for Training Data, April 10, 2008 Run

Ground Truth (Percent)

Class	127	136	140	167	175
Unclass	0.00	0.00	0.00	0.00	0.00
102	0.62	0.00	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00
108	1.74	0.00	0.04	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00
122	0.17	0.00	0.00	0.00	0.00
123	0.10	0.00	0.00	0.00	0.00
124	0.31	0.00	0.00	0.00	0.00
125	0.00	0.00	0.16	0.00	0.00
126	0.00	0.00	0.00	0.00	0.00
127	96.96	0.00	0.01	0.00	0.00
136	0.00	94.02	0.27	8.71	0.00
140	0.06	1.79	98.06	0.00	0.50
167	0.00	0.69	0.00	89.04	0.00
175	0.00	0.00	0.23	0.00	90.90
179	0.00	0.00	0.09	0.00	0.00
182	0.00	0.00	0.00	0.00	0.00
184	0.00	0.26	0.82	0.00	7.66
186	0.00	0.00	0.13	0.00	0.00
187	0.00	0.00	0.00	0.00	0.00
188	0.00	0.00	0.00	0.00	0.00
189	0.00	0.00	0.00	0.00	0.00
190	0.00	0.00	0.00	0.00	0.00
196	0.00	0.00	0.00	0.00	0.00
199	0.00	1.79	0.00	2.25	0.00
200	0.00	0.91	0.00	0.00	0.00
201	0.00	0.00	0.00	0.00	0.00
204	0.00	0.00	0.05	0.00	0.00
213	0.02	0.51	0.00	0.00	0.00
221	0.00	0.00	0.00	0.00	0.00
222	0.00	0.00	0.00	0.00	0.00
224	0.00	0.00	0.00	0.00	0.00
261	0.00	0.00	0.00	0.00	0.00
265	0.00	0.00	0.00	0.00	0.00
266	0.00	0.00	0.09	0.00	0.95
273	0.00	0.00	0.00	0.00	0.00
998	0.00	0.00	0.00	0.00	0.00
999	0.01	0.04	0.07	0.00	0.00
191	0.00	0.00	0.00	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Table A6-3 Continued: Confusion Matrix for Training Data, April 10, 2008 Run

Ground Truth (Percent)

Class	179	182	184	186	187
Unclass	0.00	0.00	0.00	0.00	0.00
102	0.00	0.00	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00
108	0.00	0.00	0.00	1.57	0.76
120	0.00	0.00	0.00	0.00	0.00
122	0.00	0.00	0.00	0.00	0.00
123	0.00	0.00	0.00	0.00	0.00
124	0.00	0.00	0.00	0.00	0.00
125	0.22	0.00	0.00	0.00	0.00
126	0.00	0.00	0.00	0.00	0.00
127	0.00	0.00	0.00	0.00	0.00
136	0.00	0.00	0.02	0.00	0.00
140	6.65	0.00	0.24	1.80	0.00
167	0.00	0.00	0.00	0.00	0.00
175	0.00	0.00	0.02	0.00	0.00
179	93.13	0.00	0.00	0.00	0.00
182	0.00	86.89	0.54	0.00	0.00
184	0.00	8.24	98.90	0.00	0.00
186	0.00	0.00	0.00	88.82	1.44
187	0.00	0.00	0.00	2.97	96.50
188	0.00	0.00	0.00	0.02	0.00
189	0.00	0.00	0.00	0.00	0.00
190	0.00	0.00	0.00	0.00	0.00
196	0.00	0.00	0.00	3.84	0.27
199	0.00	4.28	0.18	0.00	0.00
200	0.00	0.00	0.00	0.00	0.00
201	0.00	0.00	0.00	0.00	0.00
204	0.00	0.00	0.00	0.00	0.00
213	0.00	0.00	0.00	0.00	0.00
221	0.00	0.00	0.00	0.00	0.00
222	0.00	0.00	0.00	0.27	0.85
224	0.00	0.00	0.00	0.06	0.18
261	0.00	0.00	0.00	0.00	0.00
265	0.00	0.32	0.07	0.00	0.00
266	0.00	0.00	0.00	0.00	0.00
273	0.00	0.00	0.00	0.66	0.00
998	0.00	0.28	0.00	0.00	0.00
999	0.00	0.00	0.01	0.00	0.00
191	0.00	0.00	0.00	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Table A6-3 Continued: Confusion Matrix for Training Data, April 10, 2008 Run

Ground Truth (Percent)

Class	188	189	190	196	199
Unclass	0.00	0.00	0.00	0.00	0.00
102	0.00	0.00	0.00	0.03	0.00
105	0.00	0.00	0.00	0.00	0.00
108	1.08	0.00	0.00	0.00	0.00
120	0.58	0.00	0.00	0.01	0.00
122	0.00	0.00	0.00	0.00	0.00
123	0.00	0.00	0.00	0.00	0.00
124	0.00	0.00	0.00	0.00	0.00
125	0.00	0.00	0.00	0.00	0.00
126	0.00	0.00	0.00	0.00	0.00
127	0.00	0.00	0.00	0.00	0.00
136	0.00	0.00	0.00	0.00	0.00
140	0.00	0.00	0.00	0.16	0.02
167	0.00	0.00	0.00	0.00	0.01
175	0.00	0.00	0.00	0.00	0.00
179	0.00	0.00	0.00	0.00	0.00
182	0.00	0.00	0.00	0.00	0.26
184	0.00	0.00	0.00	0.00	0.59
186	0.02	0.00	0.00	0.59	0.00
187	0.00	0.00	0.00	0.20	0.00
188	98.28	0.00	0.00	0.00	0.00
189	0.00	98.95	0.00	0.57	0.00
190	0.00	0.00	96.33	0.00	0.00
196	0.00	0.47	0.00	95.57	0.00
199	0.00	0.00	0.00	0.00	98.88
200	0.00	0.00	0.00	0.00	0.00
201	0.00	0.00	0.00	0.00	0.00
204	0.00	0.00	0.00	0.00	0.05
213	0.00	0.00	0.00	0.00	0.00
221	0.00	0.00	0.00	0.42	0.00
222	0.00	0.00	0.00	0.53	0.00
224	0.00	0.58	0.00	0.05	0.00
261	0.00	0.00	1.31	1.26	0.00
265	0.00	0.00	0.00	0.00	0.00
266	0.00	0.00	0.00	0.00	0.00
273	0.04	0.00	0.79	0.62	0.00
998	0.00	0.00	0.00	0.00	0.20
999	0.00	0.00	0.00	0.00	0.00
191	0.00	0.00	1.57	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Table A6-3 Continued: Confusion Matrix for Training Data, April 10, 2008 Run

Ground Truth (Percent)

Class	200	201	204	213	221
Unclass	0.00	0.00	0.00	0.00	0.00
102	0.00	0.00	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00
108	0.00	0.00	0.00	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00
122	0.00	0.00	0.00	0.00	0.00
123	0.00	0.00	0.00	0.00	0.00
124	0.00	0.00	0.00	0.00	0.00
125	0.00	0.00	0.00	0.00	0.00
126	0.81	0.00	0.00	0.00	0.00
127	0.00	0.00	0.00	27.06	0.00
136	0.55	0.00	1.01	9.02	0.00
140	0.18	0.00	1.41	0.00	0.00
167	0.00	0.00	0.00	0.00	0.00
175	0.00	0.00	0.00	0.00	0.00
179	0.00	0.00	0.00	0.00	0.00
182	0.00	0.00	0.00	0.00	0.00
184	0.00	0.00	0.47	0.00	0.00
186	0.00	0.00	0.00	0.00	0.00
187	0.00	0.00	0.00	0.00	3.53
188	0.00	0.00	0.00	0.00	0.00
189	0.00	0.00	0.00	0.00	0.00
190	0.00	0.00	0.00	0.00	0.00
196	0.00	0.00	0.00	0.00	6.99
199	0.00	0.00	5.31	0.00	0.00
200	96.41	4.68	0.00	0.00	0.00
201	0.67	95.18	0.00	0.00	0.00
204	0.00	0.00	89.46	0.00	0.00
213	0.00	0.00	0.00	63.92	0.00
221	0.00	0.00	0.00	0.00	67.78
222	0.00	0.00	0.00	0.00	5.24
224	0.00	0.00	0.00	0.00	11.76
261	0.00	0.00	0.00	0.00	0.40
265	0.00	0.00	0.00	0.00	0.00
266	0.00	0.00	0.00	0.00	0.00
273	0.00	0.00	0.00	0.00	4.32
998	1.38	0.14	0.00	0.00	0.00
999	0.00	0.00	2.34	0.00	0.00
191	0.00	0.00	0.00	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Table A6-3 Continued: Confusion Matrix for Training Data, April 10, 2008 Run

Ground Truth (Percent)

Class	222	224	261	265	266
Unclass	0.00	0.00	0.00	0.00	0.00
102	0.00	0.14	0.00	0.00	0.00
105	0.00	0.14	0.00	0.00	0.00
108	0.00	0.00	0.00	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00
122	0.00	0.00	0.00	0.00	0.00
123	0.00	0.00	0.00	0.00	0.00
124	0.00	0.00	0.00	0.00	0.00
125	0.00	0.00	0.00	0.00	0.00
126	0.00	0.00	0.00	0.00	0.00
127	0.00	0.00	0.00	0.00	0.00
136	0.00	0.00	0.00	0.00	0.00
140	0.00	0.00	0.00	0.00	6.48
167	0.00	0.00	0.00	0.00	0.00
175	0.00	0.00	0.00	0.00	0.51
179	0.00	0.00	0.00	0.00	0.00
182	0.00	0.00	0.00	19.16	0.00
184	0.00	0.00	0.00	39.85	0.64
186	0.03	0.00	0.00	0.00	0.00
187	0.15	0.06	0.00	0.00	0.00
188	0.00	0.00	0.00	0.00	0.00
189	0.12	0.24	0.00	0.00	0.00
190	0.00	0.00	0.91	0.00	0.00
196	1.98	1.00	0.65	0.00	0.00
199	0.00	0.00	0.00	0.00	0.00
200	0.00	0.00	0.00	0.00	0.00
201	0.00	0.00	0.00	0.00	0.00
204	0.00	0.00	0.00	0.00	0.00
213	0.00	0.00	0.00	0.00	0.00
221	2.37	2.13	0.51	0.00	0.00
222	77.88	8.16	0.94	0.00	0.00
224	6.68	83.73	0.38	0.00	0.00
261	2.75	0.64	93.88	0.00	0.00
265	0.00	0.00	0.00	41.00	0.00
266	0.00	0.00	0.00	0.00	92.20
273	7.78	3.15	2.34	0.00	0.00
998	0.00	0.00	0.00	0.00	0.00
999	0.00	0.00	0.00	0.00	0.18
191	0.27	0.62	0.40	0.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

Table A6-3 Continued: Confusion Matrix for Training Data, April 10, 2008 Run

Ground Truth (Percent)

Class	273	998	999	191
Unclass	0.00	0.00	0.00	0.00
102	0.00	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00
108	0.00	0.00	0.00	0.00
120	0.00	0.00	24.06	1.65
122	0.00	0.00	0.00	0.00
123	0.00	0.00	0.00	0.00
124	0.00	0.00	0.00	0.00
125	0.00	0.00	0.00	0.00
126	0.00	0.00	0.00	0.00
127	0.00	0.00	0.45	0.00
136	0.00	0.01	0.39	0.00
140	0.00	0.00	7.89	0.00
167	0.00	0.00	0.00	0.00
175	0.00	0.00	0.00	0.00
179	0.00	0.00	0.11	0.00
182	0.00	0.01	0.00	0.00
184	0.00	0.00	0.96	0.00
186	0.09	0.00	0.00	0.00
187	0.00	0.00	0.00	0.00
188	0.63	0.00	0.00	0.00
189	0.00	0.00	0.00	0.00
190	0.57	0.00	0.00	0.56
196	3.80	0.00	0.00	0.15
199	0.00	0.18	0.00	0.00
200	0.00	0.16	0.34	0.00
201	0.00	0.00	0.68	0.00
204	0.00	0.00	0.28	0.00
213	0.00	0.00	0.00	0.00
221	0.28	0.00	0.00	0.00
222	1.98	0.00	0.00	0.00
224	0.81	0.00	0.28	0.06
261	1.64	0.00	1.41	2.53
265	0.00	0.00	0.00	0.00
266	0.00	0.00	0.00	0.00
273	89.65	0.00	0.00	0.30
998	0.00	99.64	6.54	0.00
999	0.00	0.00	54.03	0.00
191	0.56	0.00	2.59	94.75
Total	100.00	100.00	100.00	100.00

Table A6-3 Continued: Confusion Matrix for Training Data, April 10, 2008 Run

```
Class Areal Percentages
```

Class	Total
Unclass	58.54
102	1.47
105	0.05
108	7.65
120	0.75
122	1.36
123	0.09
124	0.07
125	0.17
126	0.21
127	1.47
136	0.21
140	7.84
167	0.05
175	1.07
179	0.16
182	0.43
184	3.63
186	0.33
187	0.17
188	0.99
189	0.14
190	0.09
196	1.66
199	4.72
200	0.64
201	0.15
204	0.09
213	0.01
221	0.17
222	0.25
224	0.36
261	0.42
265	0.02
266	0.52
273	0.56
998	2.95
999	0.08
191	0.46
Total	100.00

Table A6-3 Continued: Confusion Matrix for Training Data, April 10, 2008 Run

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Unclass	0.00	0.02	0/831069	161/831230
102	4.48	2.41	933/20837	491/20395
105	3.65	14.81	27/740	124/837
108	1.34	1.44	1452/108576	1564/108688
120	7.96	2.67	852/10697	270/10115
122	2.22	2.93	429/19352	571/19494
123	4.62	15.57	56/1211	213/1368
124	12.94	15.65	137/1059	171/1093
125	8.41	9.80	202/2401	239/2438
126	5.22	12.68	154/2951	406/3203
127	2.11	3.04	440/20850	639/21049
136	15.36	5.98	468/3046	164/2742
140	1.30	1.94	1442/111222	2175/111955
167	3.94	10.96	26/660	78/712
175	1.98	9.10	300/15166	1489/16355
179	4.50	6.87	99/2201	155/2257
182	8.93	13.11	540/6047	831/6338
184	6.54	1.10	3369/51515	535/48681
186	8.74	11.18	412/4716	542/4846
187	12.61	3.50	310/2459	78/2227
188	2.66	1.72	374/14078	240/13944
189	7.45	1.05	152/2040	20/1908
190	10.62	3.67	131/1234	42/1145
196	3.94	4.43	931/23623	1051/23743
199	0.85	1.12	570/67001	753/67184
200	5.83	3.59	532/9132	320/8920
201	3.39	4.82	72/2127	104/2159
204	7.66	10.54	95/1241	135/1281
213	9.94	36.08	18/181	92/255
221	14.04	32.22	336/2393	978/3035
222	26.09	22.12	930/3564	748/3382
224	17.04	16.27	862/5060	816/5014
261	13.51	6.12	810/5996	338/5524
265	20.45	59.00	55/269	308/522
266	3.46	7.80	255/7368	602/7715
273	12.57	10.35	998/7942	802/7746
998	0.94	0.36	394/41894	152/41652
999	12.34	45.97	135/1094	816/1775
191	4.01	5.25	262/6527	347/6612

Table A6-3 Continued: Confusion Matrix for Training Data, April 10, 2008 Run

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Unclass	99.98	100.00	831069/831230	831069/831069
102	97.59	95.52	19904/20395	19904/20837
105	85.19	96.35	713/837	713/740
108	98.56	98.66	107124/108688	107124/108576
120	97.33	92.04	9845/10115	9845/10697
122	97.07	97.78	18923/19494	18923/19352
123	84.43	95.38	1155/1368	1155/1211
124	84.35	87.06	922/1093	922/1059
125	90.20	91.59	2199/2438	2199/2401
126	87.32	94.78	2797/3203	2797/2951
127	96.96	97.89	20410/21049	20410/20850
136	94.02	84.64	2578/2742	2578/3046
140	98.06	98.70	109780/111955	109780/111222
167	89.04	96.06	634/712	634/660
175	90.90	98.02	14866/16355	14866/15166
179	93.13	95.50	2102/2257	2102/2201
182	86.89	91.07	5507/6338	5507/6047
184	98.90	93.46	48146/48681	48146/51515
186	88.82	91.26	4304/4846	4304/4716
187	96.50	87.39	2149/2227	2149/2459
188	98.28	97.34	13704/13944	13704/14078
189	98.95	92.55	1888/1908	1888/2040
190	96.33	89.38	1103/1145	1103/1234
196	95.57	96.06	22692/23743	22692/23623
199	98.88	99.15	66431/67184	66431/67001
200	96.41	94.17	8600/8920	8600/9132
201	95.18	96.61	2055/2159	2055/2127
204	89.46	92.34	1146/1281	1146/1241
213	63.92	90.06	163/255	163/181
221	67.78	85.96	2057/3035	2057/2393
222	77.88	73.91	2634/3382	2634/3564
224	83.73	82.96	4198/5014	4198/5060
261	93.88	86.49	5186/5524	5186/5996
265	41.00	79.55	214/522	214/269
266	92.20	96.54	7113/7715	7113/7368
273	89.65	87.43	6944/7746	6944/7942
998	99.64	99.06	41500/41652	41500/41894
999	54.03	87.66	959/1775	959/1094
191	94.75	95.99	6265/6612	6265/6527

Table A6-4: Confusion Matrix for Threemile Hill Quad, April 10, 2008 Run

```
Confusion Matrix:
H:\ThreemileReRun\ErrorMatrix\PredictedArea\PredClassify

Overall Accuracy = (114295/153367)  74.5239%
Kappa Coefficient = 0.6785
```

Ground Truth (Percent)

Class	Unclass	108	123	127	140
Unclass	97.47	0.00	0.00	0.93	0.00
108	0.00	5.04	41.56	0.00	0.00
123	0.00	0.00	57.93	0.00	0.00
127	0.00	0.00	0.00	86.34	0.00
140	0.03	78.13	0.00	4.35	97.44
184	0.00	0.01	0.00	0.00	1.93
186	0.06	4.78	0.00	0.00	0.17
189	0.20	0.66	0.00	0.00	0.13
196	1.45	11.36	0.00	0.00	0.01
222	0.00	0.00	0.00	0.00	0.00
224	0.00	0.00	0.00	0.00	0.00
261	0.00	0.00	0.00	0.00	0.00
999	0.00	0.00	0.00	0.00	0.30
102	0.79	0.03	0.50	8.39	0.01
Total	100.00	100.00	100.00	100.00	100.00

Ground Truth (Percent)

Class	184	186	189	196	222
Unclass	0.00	0.00	0.00	0.00	0.00
108	0.00	3.94	1.04	0.00	0.00
123	0.00	0.00	0.00	0.00	0.00
127	0.00	0.00	0.00	0.00	0.00
140	1.49	17.73	0.58	0.00	0.00
184	83.52	1.40	0.00	0.00	0.00
186	0.19	50.79	0.12	2.33	0.00
189	14.56	0.35	97.11	6.00	1.98
196	0.12	24.11	0.16	91.66	20.46
222	0.00	0.00	0.00	0.00	20.13
224	0.00	0.00	0.00	0.00	9.57
261	0.00	0.00	0.00	0.00	40.26
999	0.01	0.00	0.00	0.00	0.00
102	0.11	1.69	0.99	0.00	7.59
Total	100.00	100.00	100.00	100.00	100.00

Table A6-4 Continued: Confusion Matrix for Threemile Hill Quad, April 10, 2008 Run

Ground Truth (Percent)

Class	224	261	999	102
Unclass	0.00	0.00	0.00	0.00
108	0.00	0.00	0.00	37.45
123	0.00	0.00	0.00	0.95
127	0.00	0.00	0.00	1.20
140	0.00	0.00	61.88	1.60
184	0.00	0.00	0.00	0.00
186	0.00	0.00	0.00	0.00
189	0.00	0.00	11.88	2.70
196	0.00	44.79	0.00	31.83
222	0.00	2.45	0.00	0.79
224	56.38	0.00	0.00	0.01
261	3.19	52.76	0.00	0.48
999	0.00	0.00	26.25	0.00
102	40.43	0.00	0.00	22.97
Total	100.00	100.00	100.00	100.00

Class Areal Percentages

Class	Total
Unclass	4.12
108	5.26
123	0.41
127	0.32
140	39.11
184	21.32
186	1.78
189	16.25
196	7.89
222	0.13
224	0.09
261	0.20
999	0.15
102	2.97
Total	100.00

Table A6-4 Continued: Confusion Matrix for Threemile Hill Quad, April 10, 2008 Run

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Unclass	0.05	2.53	3/6318	164/6479
108	90.68	94.96	7315/8067	14164/14916
123	26.87	42.07	169/629	334/794
127	43.50	13.66	214/492	44/322
140	22.40	2.56	13436/59977	1221/47762
184	2.97	16.48	971/32692	6261/37982
186	36.07	49.21	983/2725	1688/3430
189	25.99	2.89	6478/24929	550/19001
196	70.11	8.34	8486/12104	329/3947
222	70.39	79.87	145/206	242/303
224	22.06	43.62	30/136	82/188
261	71.33	47.24	214/300	77/163
999	63.95	73.75	149/233	236/320
102	10.51	77.03	479/4559	13680/17760

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Unclass	97.47	99.95	6315/6479	6315/6318
108	5.04	9.32	752/14916	752/8067
123	57.93	73.13	460/794	460/629
127	86.34	56.50	278/322	278/492
140	97.44	77.60	46541/47762	46541/59977
184	83.52	97.03	31721/37982	31721/32692
186	50.79	63.93	1742/3430	1742/2725
189	97.11	74.01	18451/19001	18451/24929
196	91.66	29.89	3618/3947	3618/12104
222	20.13	29.61	61/303	61/206
224	56.38	77.94	106/188	106/136
261	52.76	28.67	86/163	86/300
999	26.25	36.05	84/320	84/233
102	22.97	89.49	4080/17760	4080/4559