Winter 2007

Statistical and cartographic modeling of vernal pool locations: Incorporating the spatial component into ecological modeling

Tina A. Cormier
University of New Hampshire, Durham

STATISTICAL AND CARTOGRAPHIC MODELING OF VERNAL POOL

LOCATIONS: INCORPORATING THE SPATIAL COMPONENT INTO

ECOLOGICAL MODELING

BY

TINA A. CORMIER

H.B.A., Saint Anselm College, 2001

THESIS

Submitted to the University of New Hampshire

In Partial Fulfillment of

The Requirements for the Degree of

Master of Science

In

Natural Resources

December, 2007
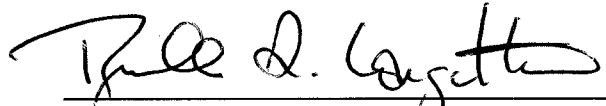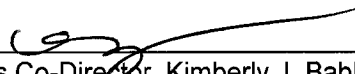
This thesis has been examined and approved.

Thesis Co-Director, Russell G. Congalton,
Professor of Remote Sensing & Geographic
Information Systems

Thesis Co-Director, Kimberly J. Babbitt,
Professor of Wildlife Ecology

Robert T. Brooks,
Research Wildlife Biologist, United States Forest
Service

Nov. 10, 2007
Date

ACKNOWLEDGEMENTS

I would like to thank a number of people who made the completion of this project possible. First, thank you to my committee, whose willingness to provide input and discussion throughout this study was invaluable. I especially want to thank Russ for providing me with necessary tools to complete this research. Perhaps the greatest achievement of this work was the autonomy I developed because of your ability to not only provide guidance when necessary, but also to let me struggle to find the answers for myself when they were within my reach.

I also want to extend a sincere thank-you to the Southern Nevada Water Authority; they were instrumental in the timely completion of this project. I genuinely appreciate the personal and professional support that I received from my co-workers as I concluded this research.
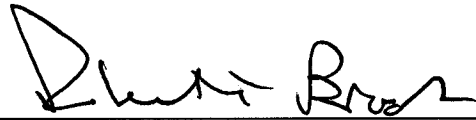
I would like to thank my family for their love and support, not only during my time at UNH, but always. Without your unending confidence in me from the time I was little, this accomplishment would certainly not have happened. Mum, thanks for believing in me and listening to my rants (both good and bad) as I waded through this project. Your faith and pride in me impels me to always do better! Dad, thanks for all of your encouragement, interest in my work, and, of course, for missing Patriots games to help me with field work. The sacrifices you (and Mum) have made to even allow me the opportunity to pursue an advanced degree often gave me the drive to continue even when I didn't want to anymore.

iii

Brat, thank you for rooting for me all the way through, for letting me bounce ideas off of you, and for never doubting me (you never did, right?). Sisters just have a special understanding of one another, and though I don't say it much, I genuinely appreciate how you were there when I needed you!

Finally, I want to thank my best friend, my partner, my love, Jesse. Your review of and contributions to this document vastly improved its quality. Also, without you there, standing by me day after insane day, keeping me focused, and urging me forward, I am positive that this project would have taken much longer or remained incomplete. You helped me *wu wei* and *bok choi*, and for that, I am truly grateful. You are my favorite.

TABLE OF CONTENTS

vi

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

STATISTICAL AND CARTOGRAPHIC MODELING OF VERNAL POOL LOCATIONS: INCORPORATING THE SPATIAL COMPONENT INTO ECOLOGICAL MODELING

By

Tina A. Cormier

University of New Hampshire, December, 2007

Vernal pools are small, isolated, depressions that experience cyclical periods of inundation and drying. Many species have evolved strategies to utilize the unique characteristics of vernal pools; however, their small size, seasonal nature, and isolation from other, larger water bodies, suggest increased risk of damage/loss by development. The goals of this research were to statistically determine physical predictors of vernal pool presence and, subsequently, to represent the output cartographically for use as a conservation tool. Logistic regression and Classification and Regression Tree (CART) routines were used to define important variables (slope, aspect, land use, soils, and reflectance) of 405 known vernal pools across northeastern Massachusetts. The CART models performed most favorably, achieving cartographic accuracies as high as 97% and providing a set of rules for vernal pool prediction. This combined statistical and spatial approach represents an efficient and accurate method of identifying vernal pools in Massachusetts and other, similar landscapes.

INTRODUCTION

In the Northeast, seasonal forest pools, often referred to as "vernal pools," are ephemeral wetlands that are biologically active (mainly) during the spring and summer months. They provide essential breeding habitat for amphibians and invertebrate species that are adapted to ephemeral and fish-free environments. For this reason, vernal pools are generally defined by the wildlife found within them, rather than by their physical features, as is characteristic of other habitat definitions. Most pools, however, have some basic physical attributes in common: they are small depressional basins, they are geographically isolated from other wetlands (no permanent inlet and/or outlet of surface water), and they exhibit cyclical/seasonal periods of inundation and drying. As a result of this particular set of characteristics, vernal pools are often left unprotected under wetland legislation and are therefore easily overlooked by developers.

In response to the vulnerability of seasonal forest ponds to filling and fragmentation of adjacent uplands, Massachusetts has developed legislation to help protect them. Massachusetts has been a pioneer in accepting the difficult issues surrounding vernal pool protection; it was one of the first states in the nation to pass regulations that specifically protect vernal pool habitat (Burne and Griffin 2005). Many other states have used Massachusetts regulations as a model for developing their own vernal pool protection regulations.

While legislation is a necessary step in the process of safeguarding vernal pools, a complete inventory of vernal pool locations across the landscape is necessary to begin effective enforcement of these regulations. Until recently, vernal pool identification in Massachusetts relied almost exclusively upon vernal pool certification though citizen participation, resulting in patchy distributions of known pools. These distributions were merely a reflection of areas where groups of interested individuals worked to identify pools rather than their actual distribution throughout the landscape. Until 2001, this "certification" method was the primary technique for inventorying vernal pools. In fact, prior to 2001, there had never been an attempt to comprehensively map vernal pools in the state of Massachusetts (Burne 2001). In the spring of 2001, an intensive effort was made to more completely identify potential vernal pools on a statewide scale by photo interpreting aerial photographs (Burne 2001). While this method was considered to be relatively fast and effective for pool detection across the landscape, there are other, newly evolving methods that may prove to be more time and cost effective than aerial surveys.

Ecological modeling may provide a less labor-intensive solution for identifying vernal pool locations over large geographic areas. Predictive ecological models endeavor to correlate the presence of a feature in the landscape (in this case, a seasonal forest pool) with other significant "predictor" variables at the same location (Guisan and Zimmermann 2000). From the model, rules can be generated for predicting the feature of interest in other, similar

2

areas. Inherently, this particular problem is a spatial one, which lends itself to the use of Geographic Information Systems (GIS) and remote sensing.

Objectives

The overall goals of this study were to statistically determine physical predictors of vernal pool presence in central and northeastern Massachusetts and, subsequently, to represent the output cartographically (as a map) for use as a conservation tool. Specifically, the goals were to:

- Explore the use of logistic regression as a modeling technique for the prediction of vernal pool locations.

- Explore the use of Classification And Regression Tree (CART) analysis as a modeling technique for the prediction of vernal pool locations.

- Implement and assess each model using Geographic Information Systems.

- Choose the model that most comprehensively identifies vernal pools (the model with the fewest omission errors).

- Facilitate and focus the efforts of those individuals and/or groups who are interested in identifying vernal pools over a large geographic area.

Assumptions

- There is a correlation between where vernal pools occur in the landscape and the physical features at those locations.

- This correlation can be determined with GIS and remotely sensed data, and predictive (statistical) modeling.

3

- The physical characteristics of vernal pool locations do not vary significantly over the geographic range of the study area (Central – Northeastern Massachusetts).

4

CHAPTER I

LITERATURE REVIEW

Seasonal Forest Pools

Definitions

Vernal pools, found throughout the United States, have been described in various ways. Generally, they are defined as "seasonal wetlands that form in shallow basins and alternate on an annual basis between a stage of standing water and . . . drying conditions" (Keeley and Zedler 1998). Those found in the northeast were generally formed by retreating glaciers at the end of the last ice age (~10,000 years ago) (Colburn 2004). As the large mountains of ice melted, they left depressions in the landscape; many of these depressions remain evident today as vernal pools and other wetlands (Colburn 2004; Preisser et al. 2000). Other vernal pools formed where suitable geology, slope, and land use allowed for proper water retention and drainage.

The term "vernal pool" has become very popular in the literature to describe many types of ephemeral wetlands; however, pools in the northeastern United States are often not vernal *per se*. Though they are typically most full during the early spring, the hydrological cycle of most vernal pools is characteristically autumnal in origin; therefore, they are more appropriately termed "seasonal forest ponds," (Brooks et al. 1998; Brooks 2004). Both "vernal pool," "seasonal forest pond," "seasonal forest pool," and "seasonal woodland

5

pond" will be used interchangeably in this study, since the National Heritage and Endangered Species Program is still officially using the term "vernal pool."

In the northeastern U.S. specifically, seasonal forest ponds are generally described/defined/valued, at least in part, by the species which use them (i.e. obligate or facultative species), which, for most purposes, is an acceptable and appropriate way of discussing them. For example, the state of Massachusetts, through the Wetland Protection Act (310 CMR 10.04), defines vernal pools as: "confined basin depressions which, at least in most years, hold water for a minimum of two continuous months in the spring and/or summer, and which are free of adult fish populations . . . [and] are essential breeding habitat . . . for a variety of amphibian species and other wildlife" (as cited in Burne and Griffin 2005). Colburn (2004) describes vernal pools similarly:

> a shallow, isolated, non-flowing woodland water body that attains its maximum depth and volume in spring, remains flooded for a minimum of two months, and periodically loses all or most of its water volume and surface area, and in which the biological community lacks fish and includes species requiring the absence of fish predation and adapted to seasonal drying. (p.292)

For modeling purposes, however, a species-centric definition is not appropriate; instead a definition based upon physical characteristics is more acceptable. Seasonal forest ponds are technically classified as "seasonally to semi-permanently flooded, scrub-shrub or forested palustrine wetlands (Cowardin et al. 1979) and are characterized as occurring in isolated, confined basins with no permanent hydrological connection to a stream or other permanent water body" (Brooks et al. 1998).

6

Physical Characteristics

Hydroperiod. There are a number of important physical characteristics that vernal pools, to some degree, tend to have in common. Hydroperiod, the duration of inundation, is a critical element to the survival of vernal pool species; in fact, it is one of the most important factors in determining the habitat suitability for specific amphibian species (Babbitt 2005; Babbitt et al. 2003; Brooks 2004; Skidds and Golet 2005). Hydroperiod is largely determined by site, morphology, and weather-related factors (Brooks 2004). Climate plays a substantial role in vernal pool hydrology; since there is no permanent inflow or outflow of surface water, the water balance of these systems is generally controlled by precipitation (snow melt and rain), evapotranspiration, and groundwater exchange (Brooks and Hayashi 2002). Vernal pool water sources may include: rainfall, surface run-off, intermittent stream flow, groundwater, and/or flooding from adjacent water bodies (Colburn 2004). Seasonal forest pond water levels have a strong positive correlation to precipitation and a negative relationship with Potential Evapotranspiration (PET) (Brooks 2004). Simply stated, the periodic drying most vernal pools experience is a result of pool morphology and the fact that pools tend to have negative water balances between June and August (i.e. evapotranspiration is greater than precipitation) (Brooks 2004).

Little is understood about the surface water-groundwater connection in vernal pools and how it may affect their hydrology; however, many agree that the connection exists. A study of prairie pothole wetlands in North Dakota revealed that, at intermediate elevations, the wetlands were receiving groundwater

7

discharge for much of the year (Winter and Rosenberry 1995). Further, pools at higher elevations were found to recharge groundwater during precipitation events. In the Northeast, many vernal pool depressions intersect and fluctuate with the groundwater table (Colburn 2004). During summer drawdown, however, the water table of most vernal pools remains above that of the underlying groundwater table because they are hydrologically isolated by an extensive layer of organic material (Colburn 2004). Similarly, Brooks and Hayashi (2002) assert that almost all pools have some degree of interaction with groundwater; pools that have no groundwater connection are more ephemeral than those that do, because their water balance is strictly determined by the difference between precipitation and evapotranspiration.

Pool Morphology. While vernal pools exhibit variable size and depth (Colburn 2004), they are generally characterized as small, shallow depressions throughout the landscape. Most pools described in the literature are less than 0.1 ha in surface area, though they can be larger (Brooks et al. 1998; Colburn 2004). In 34 Massachusetts vernal pools, Brooks and Hayashi (2002) found that the maximum depth ranged from 0.11 m to 0.94 m (measurements acquired at maximum storage in early spring). They found maximum surface area to range from 68 m^2 to 2,941 m^2, and maximum volume ranged from six to 506 m^3. Pool perimeter ranged from 30 m to 388 m. Pool morphology has also been weakly correlated to hydroperiod: Brooks and Hayashi (2002) found that pools with a surface area greater than 1,000 m^2 or a volume greater than 100 m^3 and a depth greater than 0.5 m were inundated more than 80% of the times they were visited

8

(visits occurred between March and August). Other, smaller pools had much more variable hydroperiods, which indicated that pool morphology was not the only factor controlling hydroperiod in these pools.

Soils. There has not been much published work regarding soil types in vernal pools in the Northeast. Generally, many pools are found on poorly drained, moderately drained, and somewhat to excessively well drained soils (Colburn 2004). Surprisingly, few pools are considered to be truly perched, as there is evidence of groundwater-vernal pool interactions. Perched pools depend solely upon precipitation and run-off for their water supply; therefore, depressions on bedrock and very poorly drained soils typically support very few seasonal woodland ponds (Colburn 2004).

As part of a larger study in Rhode Island, Skidds and Golet (2005) observed the soil characteristics at 65 vernal pools. They recorded the properties of the O (organic) horizon, the A horizon, and parent material textures. They found that the mean thickness of the O horizon was variable and ranged from 0 cm – 255 cm. The mean depth of the organic layer was 33.92 cm, and 75% of pools had less than 40 cm of organic material. In the A horizon, they observed that the most common texture was silt-silt loam, followed by sandy loam-fine sand. Finally, parent material textures were largely loamy sand-sand and sandy loam-fine sands. They analyzed the relationship between soil texture and hydroperiod, and found that A horizon coarseness was positively correlated with mean hydroperiod, and parent material texture had no relationship with hydroperiod.

9

Land Cover/Land Use. Despite the widespread global distribution of vernal pools, California vernal pools appear to be the only ones that have evolved extensive endemic floral species (Keeley and Zedler 1998). In the glaciated northeast, vernal pool flora consists of typical wetland species found locally in other habitats (Colburn 2004), illustrating how landscape setting is a primary determinant of wetland structure (Godwin et al. 2002). Vernal pools can occur in "isolation" (i.e. surrounded by uplands), or within larger wetland systems. Those that occur in uplands tend to have typical local wetland species on the outer edges of the basin. Within the basin, ferns, mosses, herbaceous annuals and perennials, shrubs, and trees are common (Colburn 2004). Pools that are within larger wetland systems are generally found in red maple swamps, spruce fir swamps, Atlantic and northern white cedar swamps, shrub swamps, fens and bogs (Colburn 2004).

Wildlife

Obligate vs. Facultative Species. Vernal pools provide essential habitat for many species of wildlife. Some species, referred to as "obligate species," have developed life history strategies that take advantage of and require fishless habitat and relatively short hydroperiods. Massachusetts has compiled a list of these species to aid in their certification program, and many are state listed as threatened, endangered, or of special concern (Table 1). Several other faunal species use vernal pool habitat for a portion of their life cycle; however, they are also able to survive in other types of wetlands: these are called "facultative species" (Table 2).

10

Table 1: Obligate vernal pool species. Table adapted from Commonwealth of Massachusetts Division of Fisheries and Wildlife (2001).

MA Breeding Obligate Species
Wood frog (*Rana sylvatica*)
Spotted Salamander (*Ambystoma maculatum*)
Blue-spotted salamander (*Ambystoma laterale*)**
Jefferson salamander (*Ambystoma jeffersonianum*)**
Marbled salamander (*Ambystoma opacum*)**
Eastern spadefoot toad (*Scaphiopus holbrooki*)**
Fairy shrimp (*Eubranchipus spp.*)

**State Listed Species

Table 2: Facultative vernal pool species. Table adapted from Commonwealth of Massachusetts Division of Fisheries and Wildlife (2001).

MA Facultative Species
Amphibians
Breeding Spring peeper (*Pseudacris crucifer*)
Breeding Gray tree frog (*Hyla versicolor*)
Breeding American toad (*Bufo americanus*)
Breeding Fowler's toad (*Bufo woodhousii*)
Breeding Green frog (*Rana clamitans melanota*)
Breeding Pickerel frog (*Rana palustris*)
Breeding Leopard frog (*Rana pipiens*)
Breeding Four-toed salamander (*Hemidactylium scutatum*)**
Adult or Breeding Red-spotted newt (*Notophthalmus v. viridescens*)
Reptiles
Spotted turtle (*Clemmys guttata*)**
Blanding's turtle (*Emydoidea blandingii*)**
Wood turtle (*Clemmys insculpta*)**
Painted turtle (*Chrysemys p. pictata*)
Snapping turtle (*Chelydra serpentina*)
Invertebrates
Predaceous diving beetle larvae (*Dytiscidae*)
Water scorpion (*Nepidae*)
Dragonfly larvae (*Odonata: Anisoptera*)
Damselfly larvae (*Odonata: Zygoptera*)
Dobsonfly larvae (*Corydalidae*)
Whirligig beetle larvae (*Gyrinidae*)
Caddisfly larvae (*Trichoptera*)
Leeches (*Hirundinea*)
Freshwater (fingernail) clams (*Pisidiidae*)
Amphibious, air-breathing snails (*Basommatophora*)

**State Listed Species

Evolutionary Strategies.   The temporary, seasonal hydrology that is characteristic of vernal pools precludes species that require permanent inundation (Leibowitz 2003), while favoring those that have evolved an ability to

11

respond rapidly to flooding conditions, quickly reach reproductive size, and survive in or near the pools during drought conditions (Colburn 2004; Zedler 2003). As stated above, hydroperiod is one of the most important factors in determining species composition. In wetlands where fish are excluded because of short hydroperiods, wildlife species (specifically amphibians and some arthropods) have not evolved strong "antipredator defenses," such as unpalatability, large body size, behavioral changes, etc. (as cited in Babbitt et al. 2003). While some species can breed in permanent wetlands that contain fish, their offspring (eggs and larvae) are extremely vulnerable to predation, as they have only weak defenses for this type of threat (Burne and Griffin 2005). Instead, vernal pool species have:

> life history strategies that provide for successful completion of an aquatic developmental phase when water is present, for survival during the dry period, and for . . . [persisting even when] successful reproduction may be impossible in some years when weather results in unfavorable hydrologic conditions in pools (Colburn 2004, 71).

Even within vernal pools themselves, hydroperiod can vary based on a number of physical factors (i.e. basin morphology, weather, groundwater interaction etc.). Variable hydroperiods result in different assemblages of amphibians (and likely other wildlife). For example, Degraaf and Yamasaki (2001) reported that wood frogs (*Rana sylvatica*) require between 52 and 135 days of inundation for hatching and metamorphosis; Spotted salamanders (*Ambystoma maculatum*) need between 92 and 164 days (as cited in Brooks 2004). Babbitt et al. (2003) determined that intermediate hydroperiods (more

12

than four months, but not permanent) were important for the survival and breeding success of spotted salamanders, wood frogs, and blue spotted salamanders (*Ambystoma laterale*). In anuran species specifically, and presumably for other amphibian species, differences in life history traits (i.e. ability to coexist/breed successfully in habitats with fish predators and the length of larval development) restrict the range of wetlands in which a species can successfully breed (Babbitt et al. 2003; Babbitt and Tanner 2000).

Upland Importance. Many species (both vertebrates and invertebrates) that use vernal pools for breeding spend the majority of their time in the surrounding uplands feeding, hibernating, nesting, and estivating (Gibbons 2003; Semlitsch 1998; Semlitsch and Bodie 2003). For example, many amphibian species have stage-specific habitat requirements: they require aquatic habitat for breeding and larval development and terrestrial habitat for foraging and hibernation (Leibowitz 2003). Herrmann et al. (2005) found that, in order to maintain amphibian species richness, ponds should be surrounded by greater than 60% forest cover within 1,000 m buffer. Ponds with less than 40% forest cover within a 1,000 m buffer experienced diminished larval assemblages. Similarly, Gibbs (1998) found that wood frogs and spotted salamanders were absent from areas with less than 30% forest cover. Semlitsch and Bodie (2003) gathered information from the literature regarding buffer widths for amphibians and reptiles, and reported that the necessary range of core habitat surrounding a wetland is 159 m - 290 m for amphibians, and 127 m - 289 m for reptiles.

13

Not only is the area immediately surrounding the pools important, but, for some species, viable corridors *between* pools are also important in maintaining populations and genetic diversity. Marsh and Trenham (2001) assert that amphibians act as metapopulations and that ponds are patch habitats where local extinctions and recolonizations can be common. Smith and Green (2005) are more wary of assuming amphibians act as "metapopulations," especially when the dispersal of amphibians is often (though not always) too little or too frequent to support metapopulation structure. In this case, whether a specific population is part of a metapopulation mainly depends upon whether or not the population is truly isolated. If the dispersal distance is such that a high rate of dispersal occurs between ponds, "disjunct" populations are essentially united into a single unit, which excludes it from being a metapopulation (Smith and Green 2005).

Regardless of whether a particular amphibian population qualifies as a metapopulation, upland connectivity between pools within dispersal distance (up to 10 km for some species (Smith and Green 2005)) is invaluable. Even though many vernal pool amphibians have shown high site fidelity to their natal and/or breeding ponds (Vasconcelos and Calhoun 2004), members of new, *successful* generations (there are many failure years) disperse to other breeding habitats. Unless there is reproductive failure in a certain year, the dispersal of juveniles may help to ensure survival if the original pond is lost, to ensure gene flow between populations or ponds, and to colonize new breeding sites (Colburn 2004). Maintaining the integrity and connectedness of wetland/vernal pool

14

mosaics is important because of inter-pool dispersal of individuals, which may result in larger, patchy populations (Smith and Green 2005) or in metapopulations (Gibbons 2003; Gibbs 2000; Lichko and Calhoun 2003; Semlitsch 1998). So, while vernal pools are isolated in the landscape, they are connected on many levels, including biologically (Zedler 2003).

Ecosystem Services

Seasonal woodland ponds serve important ecological, biological, and hydrologic functions in the landscapes in which they occur (Lichko and Calhoun 2003). First, they are important for energy exchange between aquatic and terrestrial ecosystems. Energy, in the form of biomass, is exchanged when amphibians and invertebrates complete their aquatic stages and disperse to the surrounding uplands, thus "extending the trophic interactions of the pool into the surrounding habitat" (Burne and Griffin 2005). The high perimeter-to-area ratio characteristic of small pools may magnify this effect (Palik et al. 2001).

In addition to energy exchange, vernal pools contribute disproportionately to the biodiversity of landscape. While they are generally small in size, their significance in maintaining the diversity of the landscape is large (Leibowitz 2003; Semlitsch and Bodie 1998). Vernal pools often have even higher biodiversity than other, larger and more permanent wetlands. Their small, shallow morphology and seasonal hydrology means that they typically have gentle slopes and varying moisture conditions that encourage specialization in the species that inhabit them (Leibowitz 2003). "Loss of these wetlands may have a

15

disproportionate effect on regional biodiversity relative to other wetlands" (Leibowitz 2003).

Vernal pools are also habitats for non-breeding wetland-dependent species. Many species of amphibians, reptiles, birds, and mammals use vernal pools as stepping stone habitat between wetlands. Typically, they are used as refugia, feeding/foraging areas, and watering stops. Many species of turtles, such as spotted turtles (*Clemmys guttata*), Blandings turtles (*Emydoidea blandingii*), painted turtles (*Chrysemys picta*), and snapping turtles (*Chelydra serpentina*), use vernal pools as food sources, feeding on amphibian eggs and young (Colburn 2004). These pools may be especially important feeding areas for female turtles that are developing their eggs (Colburn 2004). Garter snakes (*Thamnophis sirtalis*), ribbon snakes (*Thamnophis sauritus*), and water snakes (*Nerodia sipedon*) feed on tadpoles, metamorphs, and adult frogs and salamanders (Colburn 2004). A number of other taxa, including avian and mammalian species, also utilize vernal pools for non-breeding activity, such as feeding and watering (Colburn 2004).

In addition to important *functions* within the landscape, there are also *values*, from a human standpoint, that are fulfilled by vernal pools. For instance, vernal pools can promote flood control by reducing flood peaks associated with run-off (Leibowitz 2003). Flooding waters entering the depressions through run-off and precipitation can likely be dampened in two ways: 1. The basin itself can store water, 2. Groundwater exchange - during flood events, the groundwater can be recharged through vernal pools (Leibowitz 2003). Also, they improve

water quality by intercepting run-off, trapping sediments and nutrients, and stabilizing soils (Wolfson et al. 2002).

Threats to Vernal Pool Systems

Despite their ecological functions and values, wetlands, in general, were lost at an alarming rate over the past two centuries; Dahl (1990) reported that since the 1980s, 44 million hectares (109 million acres) of wetlands have been destroyed in the United States, which is a 50% reduction from the original 87 million ha (215 million acres) (as cited in Wolfson et al. 2002). Woodland vernal pools are especially susceptible to loss because of their small size and seasonal hydroperiod. Often times, vernal pools are either regarded as unimportant because of their size or are completely overlooked due to seasonal drying. Even in federal legislation, small wetlands are excluded from protection. Semlitsch and Bodie (1998) caution that if the goal of current legislation is to maintain/protect biodiversity, small, isolated wetlands are not expendable. The bias against small wetlands is unfounded in current literature. Wolfson et al. (2002) conducted a study analyzing wetland size and its ability to perform a given function and found that there was no significant difference between a large and a small wetland's functional capability. Further, they found that no specific wetland type (i.e. forested, scrub-shrub, emergent, etc.) had a greater probability of performing any of the functions they tested than another wetland type.

As small wetlands, vernal pools are capable of performing important ecological functions; however there are a number of significant threats to vernal pools that hinder or terminate their ability to carry out those functions. Most of

them are related to human alteration of the land: physical destruction of vernal pools; disturbance/fragmentation of adjacent uplands; changes to vernal pool hydrology, including changes in water source, depth, volume, and timing of filling/drying; watershed alterations, including changes in water quality and energy flow; pollution; invasive species, etc. (Colburn 2004).

Outright destruction occurs when pools are filled and built upon. Permanent dwellers in the pools are immediately lost, while individuals that inhabit the surrounding terrestrial areas may either experience direct morality and/or local extinctions due to loss of breeding areas (figure 1). Adjacent upland habitat must also be a consideration in vernal pool loss. When changes are made to the landscape that introduce gaps into an organism's core habitat (fragmentation), often times the organism cannot cope. For example, all amphibians that use vernal pools spend the majority of the year in the surrounding uplands, which, if destroyed, eliminates crucial core habitat. Also, disturbed upland habitat may mean that individuals can no longer reach their breeding pools or that juvenile dispersers cannot migrate to other, nearby pools (figure 1). Many, though not all, vernal pools occur in the landscape in clusters (Brooks 1998), and source-sink dynamics often occurs between pools within dispersal distance (figure 1) (Semlitsch and Bodie 1998). Source-sink dynamics means that local extinctions are common in small communities, such as vernal pools (Marsh and Trenham 2001); however, recolonization by individuals from surrounding populations is also common and aids in assuring the continued existence of the metapopulation (rescue effect). Loss of "stepping stone" pools or

18

corridors between them reduces the connectivity among remaining populations and dampens the possibility that isolated subpopulations can be rescued from neighboring pools, resulting in more local extinctions and an overall decline in amphibian populations (due to less available breeding area and greater distances to travel between wetlands) (Semlitsch and Bodie 1998). Many studies have shown the adverse effects of fragmentation on amphibian species (a few studies include: Rittenhouse and Semlitsch 2006; Rothermel and Semlitsch 2002; Rothermel and Semlitsch 2006; Semlitsch et al. 2007).
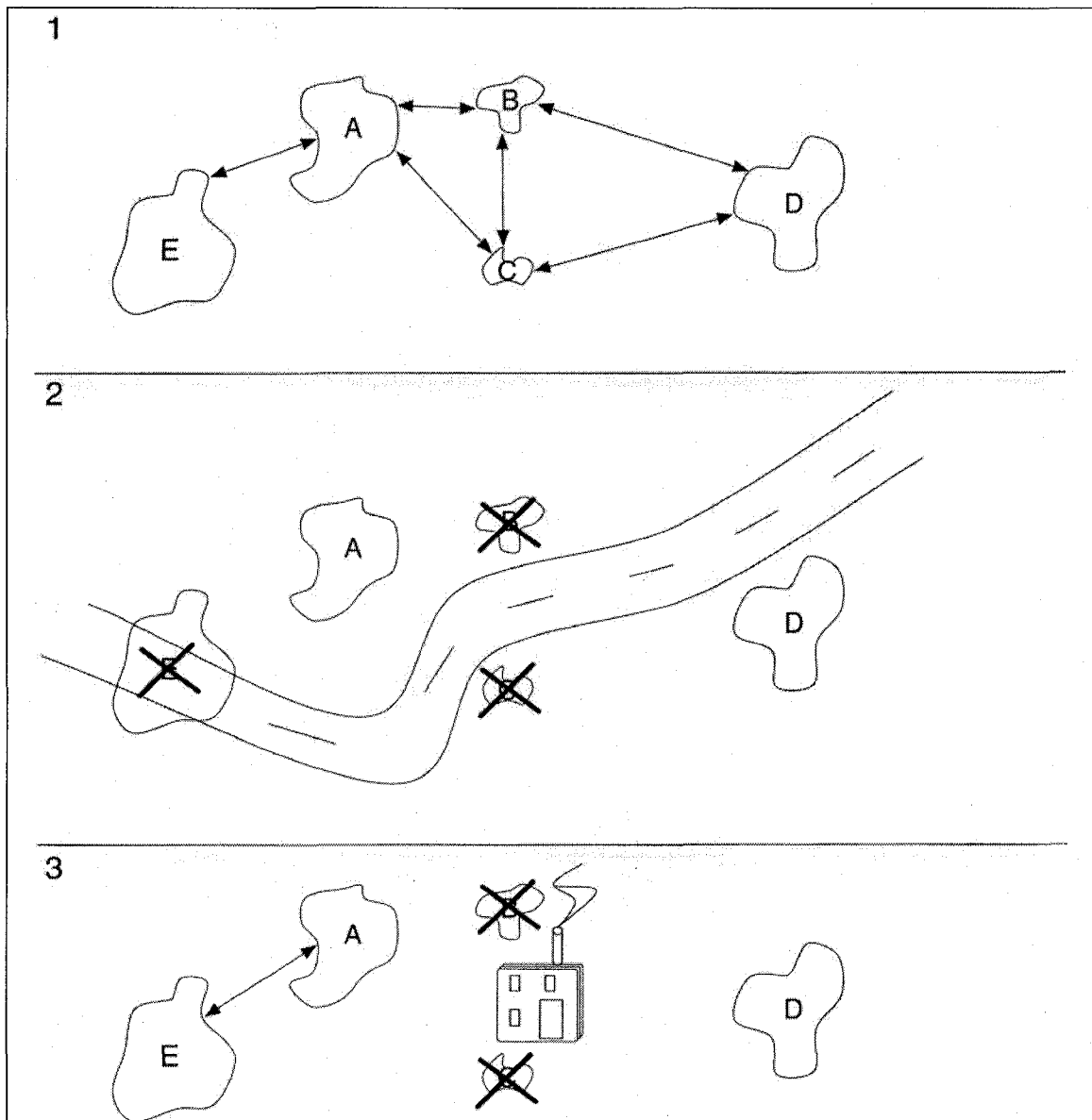
Figure 1: Scenario 1 shows an undisturbed cluster of vernal pools. In this example, all migratory populations can theoretically exchange genetic information through dispersal (arrows). For example, pool E can share material with pool D through migrations between pools A, B, and/or C. In scenario 2, a road has been built through the middle of the patch, fragmenting the uplands surrounding the pools and directly destroying pool E (direct mortality). Disruption of the adjacent uplands near pools B and C has indirectly eliminated those as well. Consequently, pools A and D have been isolated from one another and individuals can no longer migrate between them. The loss of pools E, B, and C increases the risk of local extinctions at the remaining pools, and there is no (or extremely little) chance of rescue/recolonization from a nearby pool. In scenario 3, a factory has been built, destroying pools B and C. Again, there is upland fragmentation that acts as a barrier to genetic exchange with pool D. While dispersal between A and E is still possible, the overall genetic variability of the original cluster (i.e. metapopulation) is diminished. (Figure and explanation adapted from Colburn 2004).

Policy

<u>Federal Legislation.</u>   Vernal pools, more than other wetlands, are vulnerable to loss due to their small size and ephemeral hydrology. Federal laws regarding wetland protection perpetuate this problem. The Clean Water Act (CWA) (1972) regards only navigable waters under federal jurisdiction. Responsibility for interpretation and enforcement of the CWA lies with the Army Corps of Engineers and the Environmental Protection Agency (EPA) (Downing, et al. 2003; United States Army Corps of Engineers 1987). While the CWA itself does not protect small, isolated wetlands, the EPA and Army Corps used the "Migratory Bird Rule" for protecting isolated waters, which themselves were not navigable (Downing et al. 2003). The "Migratory Bird Rule" was not applicable to birds only, however. It included waters that were or would be used "(1) as habitat by birds protected by Migratory Bird Treaties or that cross state lines, (2) as habitat for endangered species, or (3) to irrigate crops sold in commerce" (Downing et al. 2003). The conglomeration of these three cases collectively became the Migratory Bird Rule and provided the necessary nexus between important (ecological or agricultural) isolated waters and navigable ones (waters of the United States).

The Supreme Court decision in the *Solid Waste Agency of Northern Cook County v. United States Army Corps of Engineers* case in 2001, hereon referred to as "SWANCC," represented a significant weakening of Corps jurisdiction over isolated wetlands. The Court found that the use of the "Migratory Bird Rule" exceeded the authority of the Corps under the Clean Water Act. They asserted

21

that "the presence of migratory birds is by itself not a sufficient basis for asserting jurisdiction over 'isolated,' intrastate, non-navigable water bodies" (as cited in Downing et al. 2003). Presently, then, the status of non-navigable, "isolated" waters calls for a case-by-case investigation of whether there is a "significant nexus" with navigable waters, or if they are truly isolated (Downing et al. 2003). Isolation is defined here by whether degradation or destruction of such water bodies would significantly affect navigable waters (Downing et al. 2003). The SWANCC decision has caused much concern for the future of U.S. wetlands, specifically small, isolated ones. "The SWANCC decision, based more on commercial interests than on ecological resources and functions *per se*, has severely jeopardized the number, area, integrity, and value of national wetlands" (Gibbons 2003).

Massachusetts State Legislation. Massachusetts was among the first states in the nation to generate legislation that specifically protects vernal pools by adding amendments to its Wetlands Protection Act (WPA) in 1987 (Burne and Griffin 2005). Many local governments and conservation commissions have created even more stringent regulations under local wetland laws (Burne and Griffin 2005). The state has implemented a vernal pool certification program through the National Heritage and Endangered Species Program (NHESP). To be certified, vernal pools must have certain characteristics: (1) Evidence of a confined basin depression with no permanently flowing outlet and (a) a breeding obligate amphibian (Table 1), or (b) an adult obligate invertebrate (i.e. fairy shrimp), or (2) Evidence of a confined basin depression with no permanently

flowing outlet and photographs of two or more facultative species (Table 2), or (3) Evidence of a confined basin depression containing no standing water (during dry phase) and evidence of specific invertebrate presence (Commonwealth of Massachusetts Division of Fisheries and Wildlife 2001). Certification does not necessarily guarantee state protection; vernal pools are protected by the WPA only *if* they fall within a jurisdictional wetland. The upland areas surrounding CVPs are also protected, up to 30.5 m, but only if the buffer area *also* falls within the jurisdictional wetland (Burne and Griffin 2005; Commonwealth of Massachusetts Division of Fisheries and Wildlife 2001). The Wetland Protection Act itself protects eight wetland functions: "protection of public and private water supply, protection of groundwater supply, flood control, storm damage prevention, prevention of pollution, protection of land containing shellfish, protection of fisheries, and protection of wildlife habitat" (Burne and Griffin 2005). The WPA defines vernal pools as confined depressions that are inundated for at least two continuous months in the spring/summer, are essential breeding habitat for certain indicator species, and are free of adult fish populations (Burne 2001; Burne and Griffin 2005). Within the act, the wildlife habitat value of certified vernal pools (within jurisdiction) is addressed:

> Any project that would alter a certified vernal pool must demonstrate that there would be no substantial reduction in the pool's capacity to provide food, shelter, migratory and breeding areas, and overwintering areas for amphibians, or food for other wildlife. No changes to the topography, soil structure, plant community composition and structure, or hydrologic regime are permissible if, after 2 growing seasons, the habitat functions listed above would be substantially reduced (Burne 2001; Burne and Griffin 2005).

23

The WPA does not specifically provide protection for uncertified vernal pools, which is a limitation of the WPA's protection of vernal pools (Burne 2001; Burne and Griffin 2005).

Massachusetts has other regulations that offer vernal pools legal protection under specific circumstances. First, some pools that are not under jurisdiction by falling within another wetland may be protected as "Isolated Land Subject to Flooding" (ILSF) resource areas (under the WPA) (Commonwealth of Massachusetts Division of Fisheries and Wildlife 2001). ILSFs are inland wetlands that have no connections to other wetlands (Burne 2001; Burne and Griffin 2005). These habitats are not presumed to be significant to wildlife, unless, on a case by case basis, they are proven to be so (Burne 2001; Burne and Griffin 2005). The establishment of a vernal pool as an ILSF with important wildlife functions is accomplished though vernal pool certification (Burne 2001; Burne and Griffin 2005). The limitation with this legislation is that ILSF protection has no provision for the surrounding upland habitat; therefore, it does not effectively protect the wildlife functions (Burne and Griffin 2005).

The Rivers Protection Act, an amendment to the WPA, provides protection for vernal pools (both certified and uncertified) that are within 61 m (200 ft) of the banks of a perennial stream (Burne 2001; Burne and Griffin 2005). Jurisdiction under this act includes both wetland and upland areas within the resource area (Burne 2001; Burne and Griffin 2005). It is the only legislation that considers uncertified vernal pools. The act protects all vernal pools from any project that

24

would have "adverse effects" on the wildlife habitat value of vernal pools or their adjacent terrestrial, non-breeding habitat (Burne 2001; Burne and Griffin 2005).

There are other, non-WPA regulations that provide additional protection to vernal pools. The Surface Water Quality Standards, for which the Massachusetts Department of Environmental Protection is responsible, certifies that wetland filling projects comply with the federal Clean Water Act (Burne 2001; Burne and Griffin 2005; Commonwealth of Massachusetts Division of Fisheries and Wildlife 2001). Certified vernal pools that meet federal criteria for "Waters of the United States," which means that they must be navigable waters or adjacent to navigable waters (Burne 2001; Burne and Griffin 2005). Under this act, vernal pools are designated as "Class B Outstanding Resource Waters," which means that any new or increased discharge of pollutants or fill material is prohibited (Burne 2001; Buren and Griffin 2005). It also prohibits discharges of solid or liquid fill into Certified Vernal Pools. Run-off from roads or roof-tops is also not permissible (Commonwealth of Massachusetts Division of Fisheries and Wildlife 2001). For this legislation to be activated, the wetland must warrant federal jurisdiction. This legislation does not provide protection to surrounding upland habitats, rendering it less effective in protecting vernal pool species than legislation that does protect the adjacent uplands.

There are two other notable laws protecting vernal pools in Massachusetts. The first, "subsurface sewage disposal regulations," more commonly referred to as "Title 5," establishes minimum setbacks from certified vernal pool boundaries for septic systems and leach fields. In most cases, septic

25

tanks must be at least 15 m (50 ft) from vernal pool boundaries, while leach fields (and their reserves) must be a minimum of 30 m (100 ft) from pool boundaries (Burne 2001; Burne and Griffin 2005; Commonwealth of Massachusetts Division of Fisheries and Wildlife 2001).

Finally, the "Forest Cutting Practices Act" is designed to protect vernal pools from harvesting impacts. It provides both certified and uncertified vernal pools protection within 15 m (50 ft) of the pool boundary (Burne 2001; Burne and Griffin 2005). It limits harvesting, within the designated 15 m (50 ft) radius, to 50% of the basal area of the surrounding trees (Commonwealth of Massachusetts Division of Fisheries and Wildlife 2001). It also prohibits vernal pools from being used as staging areas or skidder trails and trees or tree tops from being felled into vernal pools (Burne 2001; Burne and Griffin 2005; Commonwealth of Massachusetts Division of Fisheries and Wildlife 2001).

Further, in 2007, the National Heritage and Endangered Species Program released a document of forestry Conservation Management Practices (CMP) for Massachusetts state-listed mole salamander species (National Heritage and Endangered Species Program 2007b). This document requires that additional precautions are taken during forestry activities that occur within delineated mole salamander habitat (cool, shaded, and moist forested conditions surrounding vernal pools/breeding sites) (National Heritage and Endangered Species Program 2007b). Based upon mole salamander life history requirements, these CMPs attempt to reduce direct mortality of individuals from motorized vehicles and soil compaction during harvests and to avoid habitat alteration that would

26

make forested land inhospitable for mole salamanders (National Heritage and Endangered Species Program 2007b). Some of these regulations include: a 50 foot buffer must be maintained around specified vernal pools/breeding sites, 75% canopy cover must be maintained within 70% of a 450 foot buffer from breeding sites, and no motorized equipment can be used within 450 feet of Blue-spotted and Jefferson's salamander breeding sites between March 1st and May 15th (the time of year when these species are most mobile) (National Heritage and Endangered Species Program 2007b). Similarly, no motorized equipment can be used within 450 feet of a Marbled salamander breeding site between August 15th and October 15th (the time of year when these species are most active) (National Heritage and Endangered Species Program 2007b). To minimize forest floor disturbance, soil compaction, and direct mortality, NHESP recommends forest harvesting happen during the winter months.

Mitigation.   Vernal pool mitigation has not been well-studied in the northeastern United States. One specific study, though, has attempted to evaluate the success of mitigation projects in New England. Lichko and Calhoun (2003) studied documentation of 15 vernal pool creation projects in New England to determine whether they replaced key vernal pool functions. They found that most vernal pool creation projects likely failed to reproduce the functions lost when the original pool was damaged because of poor planning; however, poor record keeping and inconsistent monitoring made success difficult to determine. They reported poor pool design as a major flaw; in fact, the pool design criteria were not well documented, and those projects that did document their plans had

27

no rationale for the specific design choices they made. Creation attempts were rarely based on successfully functioning reference wetlands, but rather on speculation. The majority of the projects considered vegetation, depth, soil, and adjacent upland habitat in their project design; however, many did not consider the water regime, egg mass attachment sites, woody debris surrounding the pool, or the transfer of amphibian eggs or adults (Lichko and Calhoun 2003). Further, none of the projects proposed to monitor water regime or pool surface area (Lichko and Calhoun 2003), even though it is well documented that pool hydrology is often the cause for the success or failure of a vernal pool (Brooks 2004; Skidds and Golet 2005). While some of the projects claimed to monitor amphibians at the pools, most were not targeting specific species (i.e. wood frogs, spotted salamanders, etc.) (Lichko and Calhoun 2005). Most projects did not even have *the goal* of replacing lost vernal pool functions; therefore, they generally failed to do so. This study illustrates the importance of understanding seasonal woodland ponds and their functions, especially for mitigation purposes. Conservation strategies should reflect the current knowledge of the life history requirements of vernal pool dependent species and also the landscape functions of small wetlands (Lichko and Calhoun 2003).

The Role of GIS and Remote Sensing in Identifying Vernal Pools

Within the last decade, there have been dramatic improvements in the spatial technology available to environmental scientists. With these improvements, there has been an increase in the number of ecological studies attempting to better incorporate a spatial component. These studies have ranged

in extent from global to local. For example, Tiner (2003) completed a nationwide study on the extent of isolated wetlands. He used National Wetland Inventory (NWI) layers, hydrology layers, and Digital Raster Graphics (DRGs) to estimate isolated wetlands in the U.S. In the Northeast, he found that isolated pools occupy about 5 - 28% of the landscape.

Many more studies have been done at the state level. For instance, in California, Smith and Verrill (1998) used GIS to create a hierarchical framework for identifying present and extant vernal pools. Their hierarchy, derived from GIS data layers, included landform, geologic formation, soil great groups, soil series, and phase of soil series. Because of the availability of statewide spatial information, they were able to identify California vernal pools, not only in the present, but also historical pools, which serve as possible mitigation sites for disturbed or destroyed pools.

Northeast vernal pools have been identified using GIS and photo interpretation in many studies. Lathrop et al. (2005) used on-screen visual interpretation of 1 meter resolution color infrared Digital Ortho Quarter Quadrangles (DOQQs) to map vernal pool occurrence in New Jersey. They identified more than 13,000 pools with 88% accuracy. They reported 12% commission error and 15% omission error using this method. They observed that the ability to discern vernal pools on aerial photography is related to pool size, pool shape, and surrounding land cover. Additionally, they did not find a consistent minimum detectable pool size, though their ability to identify pools decreased at an area of 120 m^2.

29

In Maine, Calhoun et al. (2003) experimented with different types of aerial photography to see how scale affected their ability to identify vernal pools. Specifically, 1:12,000 and 1:4800 scales were evaluated to decipher the efficacy with which vernal pools could be identified. On the 1:4,800 scale imagery, 516 pools were identified, approximately 93% of which were correct. Only 170 pools were identified on the 1:12,000 scale imagery; an estimated 90% of those pools were correctly identified. Eight percent of the pools mapped on the 1:4,800 imagery were also identified by the 1:12,000 imagery; whereas 83% of pools were delineated on the 1:12,000 photos were also mapped by the 1:4,800 scale images. The importance of scale when trying to identify isolated wetlands by photo interpretation was demonstrated.

There have been other, similar studies done specifically in Massachusetts. For example, Brooks et al. (1998) used 1:12,000 spring, leaf-off, color infrared imagery to identify vernal pools in the Quabbin Reservoir watershed. With the quality of the imagery, pools greater than 0.025 ha in size could be consistently identified. They observed that vernal pools were generally clustered in the watershed, and that overall, they occur at a density of about 1.1 ponds/km^2, with inter-pool distances ranging from 19 m to 2.4 km. Errors of omission were not computed. In a similar, but much larger project, Burne (2001) used 1:12,000 color infrared imagery to identify potential vernal pools on a statewide level, resulting in the National Heritage and Endangered Species Program Potential Vernal Pool (NHESP PVP) layer (National Heritage and Endangered Species Program 2000). He reported that pools under 15 m – 18 m (50 ft - 60 ft) in

diameter could not be accurately detected. He also observed that pools occurring beneath coniferous canopies are obscured, except where they are large enough to cause a gap in the canopy, illustrating both the strengths and limitations to photo interpretation.

Finally, in a recent study, Grant (2005) combined GIS and statistical modeling to predict vernal pools in Massachusetts. Logistic regression was used to identify specific physical predictors of vernal pool presence. The independent variables, which began as a large suite of possible predictors, were derived from GIS data layers. The best model used slope, surficial geology, percentage of cropland, urban/commercial development, and residential development as predictors of vernal pool presence. Sand/gravel, fine grained, and floodplain alluvium surficial geology types were positive correlates of vernal pool presence. Slope, percentage of cropland, urban/commercial development, and high density residential development were negatively associated with vernal pool presence. Statistically, 64% of his validation set of pools were correctly predicted; however, the results were not displayed or analyzed cartographically.

Ecological Modeling

Many ecologists are using ecological modeling to acquire important information about environmental processes, species distributions, habitat distributions, etc. Models are simplifications of reality used to explain, in this case, ecological processes (Vogiatzakis 2003). Ecological data sets are generally multivariate (contain more than one variable and often times many variables) and location specific in nature (Vogiatzakis 2003). Ecological

31

problems, therefore, lend themselves well to the use of GIS; however, most GIS lack the predictive and analytical capabilities necessary to examine complex modeling problems, while statistics-oriented problems lack important spatial components (Vogiatzakis 2003). With this problem in mind, there are two common solutions presently available. First, ecologists can use a single interface to integrate spatial and statistical models (Vogiatzakis 2003). Currently, there are few viable software options that are capable of this integration. New editions of the Idrisi software (designed by Clark University) are capable of implementing complex machine learning statistical procedures, like Artificial Neural Networks (ANNs). It can also perform simple linear regression, multiple linear regression, and logistic regression between images or attribute files (Clark Labs 2007). Also, Insightful's S+ software has the "SpatialStats" module which is capable of parametric and nonparametric trend surface analysis, Kriging, spatial regression models, nearest neighbor searches, spatial randomness tests, etc. ("S+ SpatialStats Product Features" 2007). There are relatively few other software packages that are appropriate for both statistical and spatial modeling. When such an option is not available, modelers are forced to run their models in statistical software outside of the GIS, and then interpret the model spatially in the GIS (Vogiatzakis 2003). This task is often difficult because GIS and statistics lack common data structures and have different interfaces (Vogiatzakis 2003).

While difficult, many studies have managed to integrate GIS and ecological modeling. The modeling process starts with a conceptual model, derived either from field knowledge of the subject, laboratory experiments, or

32

gathered from the literature (Guisan and Zimmermann 2000). At this step, it is important to define the goals of the study. If the purpose is to identify locations, with certainty, where an organism or habitat definitely exists, then choosing a technique and variables that minimize errors of commission (false positives) is of utmost importance. If, however, the goal is to conserve an organism or habitat, then errors of omission (false negatives) must be minimized (Muñoz and Felicísimo 2004). The next step is to choose a statistical technique; often statistical literature and/or other models are the basis for this choice. The model is then formulated and calibrated on a test set of data, which is often an iterative process. In another iterative process, the model is then tested on an independent (ideally) set of data and evaluated.

Ecological Modeling Techniques

Generalized Linear Models. Regression has long been used in ecology to determine relationships between the biological and the physical environment (Vogiatzakis 2003). In general, regression attempts to correlate a response variable to one or more environmental predictors (Guisan and Zimmermann 2000). Generalized Linear Models (GLM) are mathematical extensions of simple linear models that allow for non-linearity and non-constant variances in the data (Guisan et al. 2002). They are based on an assumed relationship, called a link function, between the mean of the dependent variable and a linear combination of predictor variables. In GLM, the independent variables are combined to produce a "linear predictor" (LP), which is related to the expected value of the response variable through a link function (Guisan et al 2002). The link function

33

used depends upon the GLM technique chosen. GLM are more flexible than simple linear models because they are appropriate for data from any of the exponential family distributions: Gaussian, Poisson, binomial, negative binomial, or Gamma, some of which may be better suited for analyzing ecological relationships than methods assuming a classical Gaussian distribution (Guisan et al. 2002; Guisan and Zimmermann 2000; Lehmann 1998). Further, they allow the use of continuous and/or categorical data (Lehmann 1998).

Logistic regression is a specific type of GLM. With this routine, the dependent variable (response variable) must be binomial (yes/no, present/absent, etc.) (Guisan et al. 2002; Lehmann 1998). It uses a logistic link (logit/logit transformation) that can fit polynomial equations to a higher degree than linear (supports non-linear data) (Hirzel et al. 2001). It allows the user to predict a discrete outcome (i.e. presence/absence) from a set of categorical or continuous predictors, though it has a difficult time modeling *complex* interactions between variables and general rule exceptions.

Logistic regression outputs a number of statistical results for determining overall model fit and the contribution of each independent variable in predicting the response variable. There are several statistics that indicate model fit. The most commonly recognized statistic is the pseudo R^2 value, which summarizes the overall strength of the model. Akaike's Information Criterion is another model fit statistic often utilized to identify the most efficient and simple model: a lower value means better model fit (Akaike 1979). Additionally, a non-significant

Hosmer and Lemeshow "Goodness of Fit Test" means that the model has adequately fit the data ("Logistic Regression" 2007).

In addition to the model fit, logistic regression is capable of determining the strength of each predictor variable. For instance, the Wald Statistic tests the significance of the logistic regression coefficients for each independent variable ("Logistic Regression" 2007). The logistic regression coefficients, often used to generate probability of prediction equations, explain the strength and sign of each variable's contribution in predicting the response variable. Significant negative values indicate avoidance or an inverse correlation to the presence of the response variable, where significant positive values indicate a positive relationship between the predictor and the presence of the response (Mace et al. 1999). Finally, the most common way of interpreting a logistic regression is by the "odds ratio." An odds ratio above one indicates positive odds that the response variable is "present" ("Logistic Regression" 2007) while odds ratios below one indicate negative odds or an inverse relationship between the predictor and the response ("Logistic Regression" 2007). Odds ratios close to one mean that the independent variable does not explain the presence of the dependent variable.

Of course, with multiple predictor variables included in the model, there is the opportunity to create multiple models. Caution should be used when choosing independent variables for the logistic regression; many variables should not be carelessly added into the model because it is well-documented that as the number of parameters increase, the accuracy with which they can be estimated

35

decreases (Bonney 1987). Specifically, more predictor variables mean increased multicollinearity (Muñoz and Felicísimo 2004). "Multicollinearity occurs when one or more variables are exact or near exact linear functions of other variables in the data set" (Muñoz and Felicísimo 2004). When this happens, it becomes very difficult to determine the effects of any one variable. More variables also equates to more noise specific to the training data set. There are methods for choosing the best model fit that penalize for complexity. One of the most common methods of determining the appropriate model from a large number of models is Akaike's Information Criteria (AIC) (Akaike 1979). AIC penalizes the model fit measure for unnecessarily increasing model complexity (i.e. number of variables). The minimum AIC denotes the best model.

The ability to model presence and absence of particular phenomenon inherently involves relating spatial data to ecologic data. To do so, landscape variables must be correlated to species/habitat presence. For this reason, most studies in this field utilize GIS in some way. Typically GIS data layers are utilized as independent (predictor) variables (i.e. elevation, slope, land use, soil type, geology, precipitation, etc). Information about the physical attributes related to species/habitat presence is collected on a site-specific basis. Once the model is created, calibrated, and evaluated, it can be transferred back into the GIS to produce a probability map depicting the likelihood that the phenomenon of interest is present in a given area. To create this layer, the inverse logistic transformation can be used, which yields a raster with each cell having a value

36

between 0 and 1 (0 meaning no chance of presence, 1 meaning 100% chance of presence).

A number of studies have used logistic regression to predict species and/or habitat presence and absence. For instance, Mladenoff et al. (1995) used a stepwise logistic regression to correlate landscape variables derived from GIS data layers to essential wolf (*Canis lupus lycaon*) habitat to assess the feasibility of recolonizing the Great Lakes area. Mace et al. (1999) used GIS and logistic regression models to describe grizzly bear (*Ursus arctos*) habitat in Montana. Gibson et al. (2004) modeled rufous bristlebird (*Dasyornis broadbenti*) habitat by coupling GIS with logistic regression. Compton et al. (2002) used a paired logistic regression to determine habitat preferences for the wood turtle (*Clemmys insculpta*). Carroll et al. (1999) used a multiple logistic regression to model fisher (*Martes pennanti*) distribution. Finally, Bian and West (1997) used logistic regression and GIS to predict elk (*Cervus Canadensis*) calving habitat preferences in Kansas. These are just a few of the examples of how logistic regression can be applied in ecology.

Generalized Additive Models. While not utilized in this study, and not to be discussed in full detail, Generalized Additive Models (GAM) represent an alternative to GLM. GAM is described as non-parametric or semi-parametric extensions of Generalized Linear Models (Guisan et al. 2002; Lehmann 1998). They build models by using smoothed functions taken from the predictor variables instead of pre-establishing a parametric model (Lehmann 1998). When predictors do not fit the traditional linear model, polynomials and transformations

37

are often used; however, they are tedious and often imprecise (Guisan et al. 2002). Generalized Additive Models facilitate this process. Like GLM, they use a link function to establish a relationship between the mean of the response variable and the "smoothed" function of the independent variable(s) (Guisan et al. 2002). GAM assesses each variable separately and can automatically identify the appropriate transform or polynomial (smoother) for each one and additively calculates the response (Guisan and Zimmermann 2000); some variables can be modeled normally while others must be modeled as transforms or polynomials (Guisan et al. 2002). This type of technique is advantageous because it can handle highly non-linear relationships, often enabling it to better represent the underlying data (Guisan et al. 2002). Since it is a nonparametric approach, however, there is one main disadvantage: when performing ecological modeling with a spatial component, interpretation of the results into a GIS is difficult because GAM do not produce a conventional mathematical function or equation (Lehmann 1998).

Classification and Regression Tree Analysis (CART). Classification and Regression Tree analysis (CART) is a technique that has recently been receiving increased attention in ecological studies. It is a routine that recursively splits predictor variables into a hierarchical sequence of groups based upon the independent variables' ability to predict the response (Andersen et al. 2000). The undivided data resides at the top of the tree and is called the "root node" (De'ath and Fabricius 2000). The routine initially splits the data into two groups, based upon the variable that most minimizes the deviance in the dependent variable

38

(Iverson and Prasad 1998; Lawrence and Wright 2001). At each subsequent split, the data are again divided into two (branches), mutually exclusive groups which are as pure/homogenous as possible (De'ath and Fabricius 2000). Each split is based on only a single variable, and variables may be used once, multiple times, or not at all (Muñoz and Felicísimo 2004). For categorical variables, splits divide the categories into two groups. For continuous variables, splits are defined by less than or greater than some chosen value (De'ath and Fabricius 2000). The result of the analysis is a dichotomous decision tree. Each path through the tree defines "the conditions that lead to the most probable class" (Lawrence and Wright 2001). The final decision points are called "leaves" or "terminal nodes." Variables that work on regional scales tend to be captured early in the model near the top of the tree (i.e. climate), while variables working on a more local scale are captured toward the terminal nodes (i.e. soil, elevation, etc.) (Iverson and Prasad 1998).

Trees will grow until completely homogenous groups are obtained or until some stopping criterion is met. For instance, in the S+ statistical package, the stopping criterion is when a node explains less than 1% of the total tree deviance (Lawrence and Wright 2001). Most of the time, CART analyses over fit the model, meaning that they begin to explain idiosyncrasies inherent in the training data only; they begin to explain noise. In these cases, the trees often become exceedingly large and difficult to interpret, so pruning methods have been developed with the goal of explaining the same, or similar, amount of variance, but with fewer terminal nodes.

There are several advantages to using CART. First, both categorical and continuous independent variables can be used together (Iverson and Prasad 1998). Further, the response variable can either be categorical (classification tree) or numeric (regression tree) (De'ath and Fabricius 2000). Prediction rules can be directly induced (Guisan and Zimmerann 2000) and hierarchical relationships between independent variables are explicitly illustrated from the tree structure (North et al. 1999). Implementation of these rules in decision-making is generally very easy (Andersen et al. 2000). For this reason, realization of tree-based models into a GIS to create predictive maps is facilitated. Statistically, CART makes no assumptions about the distribution of the response or predictor variables (Andersen et al. 2000): CART can handle complex data, non-normal data, missing values, and non-linear and high order interactions between variables (Andersen et al. 2000; De'ath and Fabricius 2000). Finally, the biggest advantage to using a CART analysis is its ability to capture non additive behavior. In other words, sometimes relationships between the response variable and some of its predictors are conditional, based upon the values of other predictors; CART can detect exceptions to general rules (Iversen and Prasad 1998). The main disadvantage to CART analyses is that, when more than a few predictor variables or cases are used to classify a data set, trees can become extremely complex and almost impossible to interpret.

There is less ecological application-centered research on CART than there is for GLM, including logistic regression. North et al. (1999) used CART to model spotted owl habitat (*Strix occidentalis*). Skidmore et al. (1996) compared

CART, BIOCLIM, and supervised classification to see how each could classify multiple species of kangaroo's habitat. They found that CART performed the best out of the three models tested. Andersen et al. (2000) compared multiple regression to a CART analysis to model desert tortoise (*Gopherus agassizii*) density. They determined that the CART results were much more revealing than the multiple regression results were. Interestingly, they began the analysis with 73 independent (predictor) variables, but the model only utilized eight predictors. Finally, Iverson and Prasad (1998) used a regression tree analysis (RTA) to map the current distribution of tree species in the eastern U.S. They were also able to map future distributions based upon climate change models. Overall, CART is beginning to receive more attention due to its applicability to ecological and spatial problems.

Other Advanced Modeling Techniques. There are myriad other techniques to choose from when creating an ecological model. There are a few relatively new, progressive routines that have recently entered into the ecological modeling literature. Multivariate Adaptive Regression Splines (MARS) is one of those techniques. It is a combination of classical linear regression, the mathematical construction of splines, and the binary recursive partitioning of CART, to model linear or non-linear response-predictor relationships (Muñoz and Felicísimo 2004). It creates a regression line; however, at points on the regression line where the trend (i.e. the slope) changes, it is allowed to bend at a point termed the "knot," which denotes the beginning of a new region of data with different behavior (Muñoz and Felicísimo 2004). These models always over fit

41

the data at first, but in the subsequent steps, the "knots" that contribute the least to the effectiveness of the model are removed by backwards pruning (Muñoz and Felicísimo 2004). This method has not yet been used extensively in ecological modeling; however, it shows tremendous promise to be an effective method in future studies.

Another interesting choice for ecological modeling is Artificial Neural Network (ANN) analysis. Often overlooked because of their obscure statistical routines, ANNs have only rarely been used in ecology (Lek and Guégan 1999). ANNs are non-linear structures that are designed to emulate the human brain. They rapidly learn from experiences to solve computational problems (Lek and Guégan 1999). Though there are multiple algorithms, back propagation (also known as multi-layer feed-forward neural network) is used most often (Lek and Guégan 1999). It is a supervised routine (user provides training data) in which information flows from the input layers, through a hidden layer that assigns weightings to the input layers, and finally to the output layer/response (Figure 2). Guisan and Zimmermann (2000) described them as more powerful than multiple regression models for describing non-linear relationships. They are advantageous because they accommodate non-parametric variables (Zhou 1999), learn adaptively from existing examples (Thurston 2002), handle noisy and missing data, adapt to patterns not observed in the training data and find the best fit (Thurston 2002), and continually learn and adjust weights with more training data (Thurston 2002). Their main pitfall is that they are still "black box" in terms of what happens within the hidden layers. Researchers are therefore

42

hesitant to use them for some applications in which it is important to gain insight into the characteristics of the data set. With more research, however, they may have exceptional potential in ecological and GIS modeling.
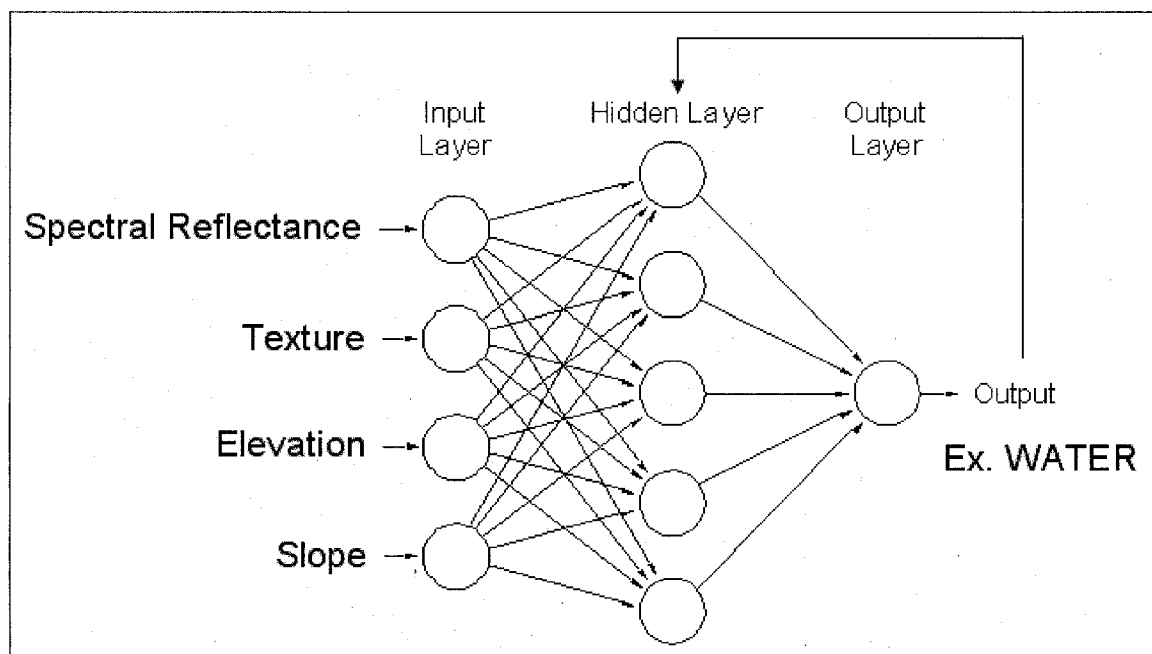
Figure 2: Simplification of Artificial Neural Network processes.

43

CHAPTER II

METHODS

Study Area Description

The area examined for vernal pools was in Massachusetts, USA (north and west coordinates: 42° 44' 45.0", 73° 15' 52.9"). Massachusetts occupies about 20,958 km (8,092 mi^2), or 1/8 of New England's total land area (MassGIS 2002a). It is also the most populous state in New England, with 6.4 million residents and an overall population density of 312 people/km^2 (810 people/mi^2) (United States Census Bureau 2006). Of the 6.4 million people residing in the state, about 3 million are within an 80 km (50 mi) radius of Boston (United States Census Bureau 2006).

The climate in Massachusetts is temperate with mild, humid summers and cold, snowy winters. Weather can change very quickly, and there are large ranges in temperature on a daily and annual basis (NOAA National Climatic Data Center 2005). Average summer temperatures range from 70°-75°F in the central part of the state, but can be greater than 90° (NOAA National Climatic Data Center 2005). Average winter temperatures are generally 23° to 27° in central Massachusetts. The growing season usually lasts between 140 - 160 days (NOAA National Climatic Data Center 2005). There are no defined wet and dry seasons; the state receives precipitation uniformly throughout the year. Total

44

precipitation averages 102 cm – 127 cm (40 in – 50 in) per year (NOAA National Climatic Data Center 2005).

Presently, about 12,002 km^2 (57.3 %) of Massachusetts is forested; approximately 466 km^2 (2.2%) of the land surface is characterized as wetlands; approximately 2,147 km^2 (10.2%) is developed (includes urban, industrial, and residential areas); and about 1,269 km^2 (6.1%) is farmland (MassGIS 2002b). The soil in the state is generally rocky. The vegetation is characterized by temperate species of trees, shrubs, and herbaceous plants. Forests in Massachusetts are described as "Deciduous Forest Land" and/or "Mixed Forest land" (Anderson et al. 1976). Deciduous areas in Massachusetts are often composed of the following tree species: red maple (*Acer rubrum*), oak (*Quercus spp.*), birch (*Betula, spp.*), and American beech (*Fagus grandifolia*). Mixed areas contain both deciduous (listed above) and evergreen species. The most common evergreens in Massachusetts are eastern hemlock (*Tsuga canadensis*) and white pine (*Pinus strobus*). Prevalent shrub species in Massachusetts include: dogwood (*Cornus amomum*), high and lowbush blueberry (*Vaccinium spp.*), buckthorn (*Rhamnus spp.*), speckled alder (*Alnus incana*), staghorn sumac (*Rhus typhina*), witch-hazel (*Hamamelis virginiana*), and many others. Additionally, Massachusetts has many herbaceous species, some of which include: meadowsweet (*Spirea latifolia*), steeplebush (*Spirea tomentosa*), Canada mayflower (*Maianthemum canadense*), indian cucumber (*Medeola virginiana*), sensitive fern (*Onoclea sensibilis*), royal fern (*Osmunda regalis*), cinnamon fern (*Osmunda cinnamomea*), and many others.

45

Training And Validation Study Sites

Training and validation sites were chosen by analyzing the Certified Vernal Pool (CVP) layer across Massachusetts (National Heritage and Endangered Species Program 2002) (Figure 3). The statewide layer was searched for assemblages of pools with similar geography and vernal pool density to represent training and validation sites. Once desirable clusters of Certified Vernal Pool points were identified, a convex hull (the smallest polygon containing all of the points) was generated around each cluster using the custom convex hull extension for ArcView 3.3 (Jenness 2004). The resultant polygons were then buffered by 100 m to account for the error associated with the CVP layer. Four training sites and four validation sites were used in the models (Figure 4; Table 3). The training sites totaled 9,145 ha, and the validation sites totaled 8,911 ha.

Training and validation study sites were used in model generation, calibration, and evaluation. Training sites were used to gather information about the predictor variables used in the various statistical models examined in this study; validation sites were used to test the success and robustness of the models (Guisan and Zimmermann 2000).

46
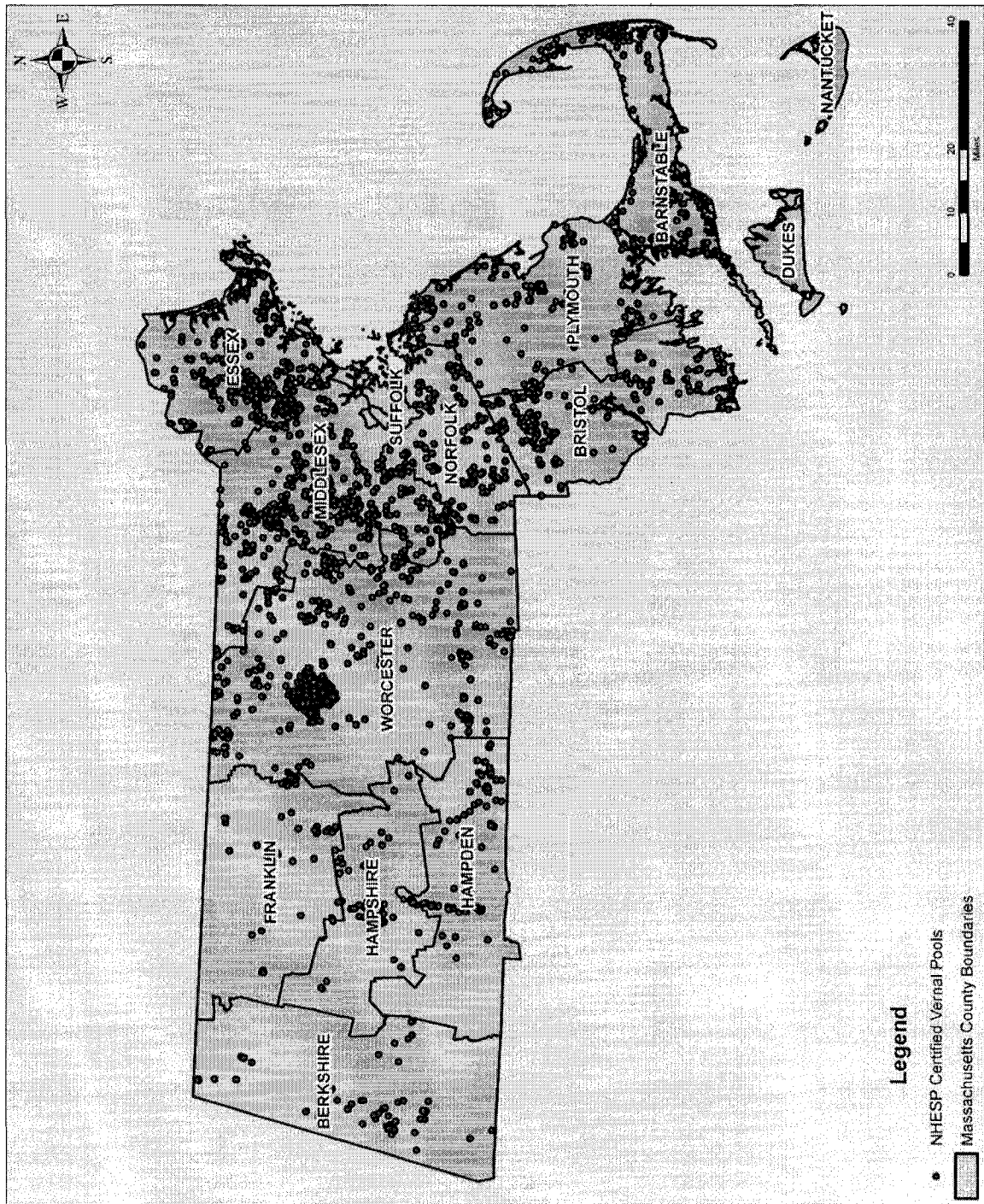
Figure 3: The Massachusetts Certified Vernal Pools layer, which was examined for pairs of complementary assemblages of pools for model training and validation.

47
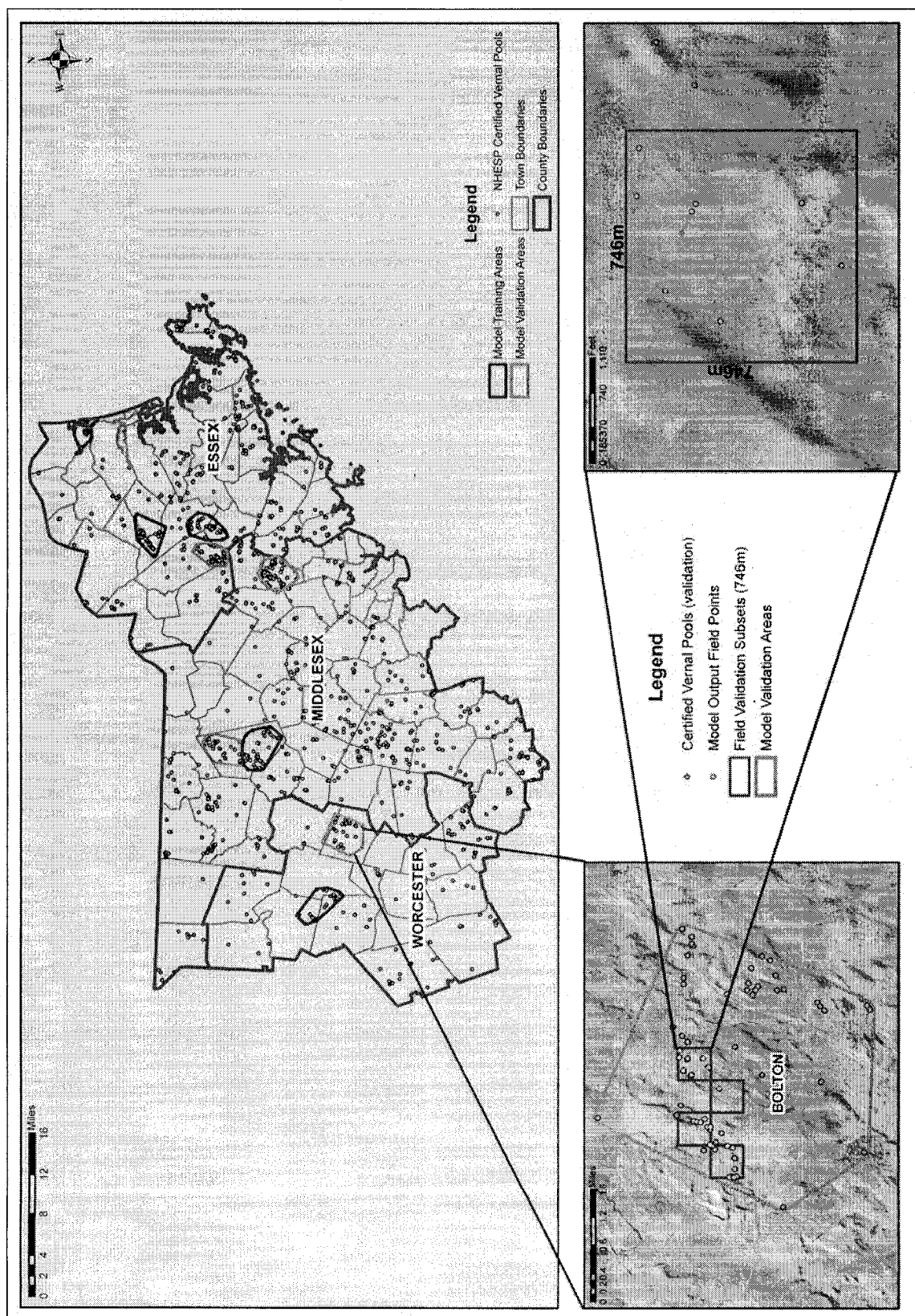
Figure 4: Model training and validation areas used for model creation, calibration, and evaluation. Inset maps depict field validation areas.

48

Table 3: Training and validation areas: location, total area, and density of pools.

ID	Town	# CVPs	Area (Acres)	Density (pools/acre)	Description
Train 1	Boxford	59	4662.9182	0.0127	Northeast Massachusetts,
Validation 1	N. Andover	63	3563.9139	0.0177	Essex County.
Train 2	Georgetown	71	4878.9553	0.0146	Northeast Massachusetts,
Validation 2	Reading	69	5526.9445	0.0125	Middlesex & Essex
Train 3	S. Westford	44	7578.2786	0.0058	Northern Massachusetts,
Validation 3	N. Westford	44	5951.2665	0.0074	Middlesex County.
Train 4	Sterling	40	5477.9293	0.0073	Central Massachusetts,
Validation 4	Bolton	39	6977.8449	0.0056	Worcester County

Field validation of model outputs was difficult due to the vastness and discontinuity of the total validation area. To make fieldwork more manageable, the validation polygons were subset, resulting in four field validation subsets per polygon (totaling 16 subsets). The subsets totaled 10% of the total validation area (891.12 ha) (Figure 4) and made field checking more achievable.

Modeling Framework

In this project, statistics, GIS, and remote sensing were combined to create predictive models of vernal pool locations. Guisan and Zimmermann's (2000) modeling framework was chosen for this study, whereby a conceptual model is formulated based upon potential model inputs. These inputs are chosen from information gathered during extensive literature review and from field experience (Figure 5). Appropriate statistical models are then identified, tested, and eventually calibrated on a set of training data. This process is iterative, and once an acceptable model is formulated with the training data, it is then evaluated on a separate set of validation data. The processes of training and testing the models are inherently coupled, and are, again, iterative. Model

production continues until some stopping criterion is met or until an acceptable model is produced.

Data

With a conceptual model in mind (Figure 5), a number of GIS data layers were gathered to reflect the necessary inputs into the model and to generate information about known vernal pool points. All layers acquired were projected to Massachusetts State Plane, NAD83 meters. Of utmost importance was the National Heritage and Endangered Species Program (NHESP) Certified Vernal Pool (CVP) point layer. It was chosen to represent known vernal pools that have been documented and certified by the NHESP as of December 2002 (National Heritage and Endangered Species Program 2002). Certification of Massachusetts vernal pools involves documentation of obligate or facultative vernal pool species and of pool location (Commonwealth of Massachusetts Division of Fisheries and Wildlife 2001). These data were converted into a GIS data layer by mapping the points on 1:24,000 or 1:25,000 USGS topographic quadrangle maps and using the coordinates from the topographic maps to create an Arc/Info coverage. The accuracy of this layer is 100 meters (Szczebak, personal communication, June 5, 2006). These vernal pool points were used as the response variable in all models in this study. Since they are known points on the ground, collection of both spectral and ancillary data was facilitated. Since information about CVPs within the study area (soils, slope, aspect, land cover, spectral data, etc.) was used to train the various models created in this study, it was important that vernal pool points be accurately represented. With permission

50

from the National Heritage and Endangered Species Program, inaccurately mapped certified vernal pool points that were within study area boundaries (training and validation study sites) were corrected. The determination of which CVP points were to be edited was done by photo interpretation of 1:5,000 color Digital Orthophoto Quadrangles (DOQs) provided by MassGIS, the Massachusetts Geographic Information System repository. Using ArcGIS 9.1, points that were within 100 m of a photo-interpreted vernal pool on the imagery were manually corrected using on-screen digitizing. By overlaying the CVP points onto the imagery, it was possible to determine if each point was spatially accurate. When a vernal pool was not apparent (on the imagery) in the location of the CVP point, a 100 m radius around the point was analyzed to determine if there was an obvious pool within the boundary of error (Figure 6). If there was more than one potential pool within a 100 m radius of the point, the closest one was chosen. If the presence or absence of a vernal pool could not be determined within 100 m of a given CVP point, that point was removed from the analysis. Removal of points was prevalent in areas of dense coniferous canopy cover. The National Wetlands Inventory (NWI) and slope layers were used in conjunction with the imagery to determine whether an area on the imagery could be a vernal pool. Overall, there were 198 CVP points used as training data and 205 CVP points used as validation data.

In addition to the CVP layer, the NHESP Potential Vernal Pool (PVP) layer was also utilized. This layer represents unverified vernal pools identified by photo interpretation of 1:12,000 color-infrared, spring, leaf-off aerial photography.
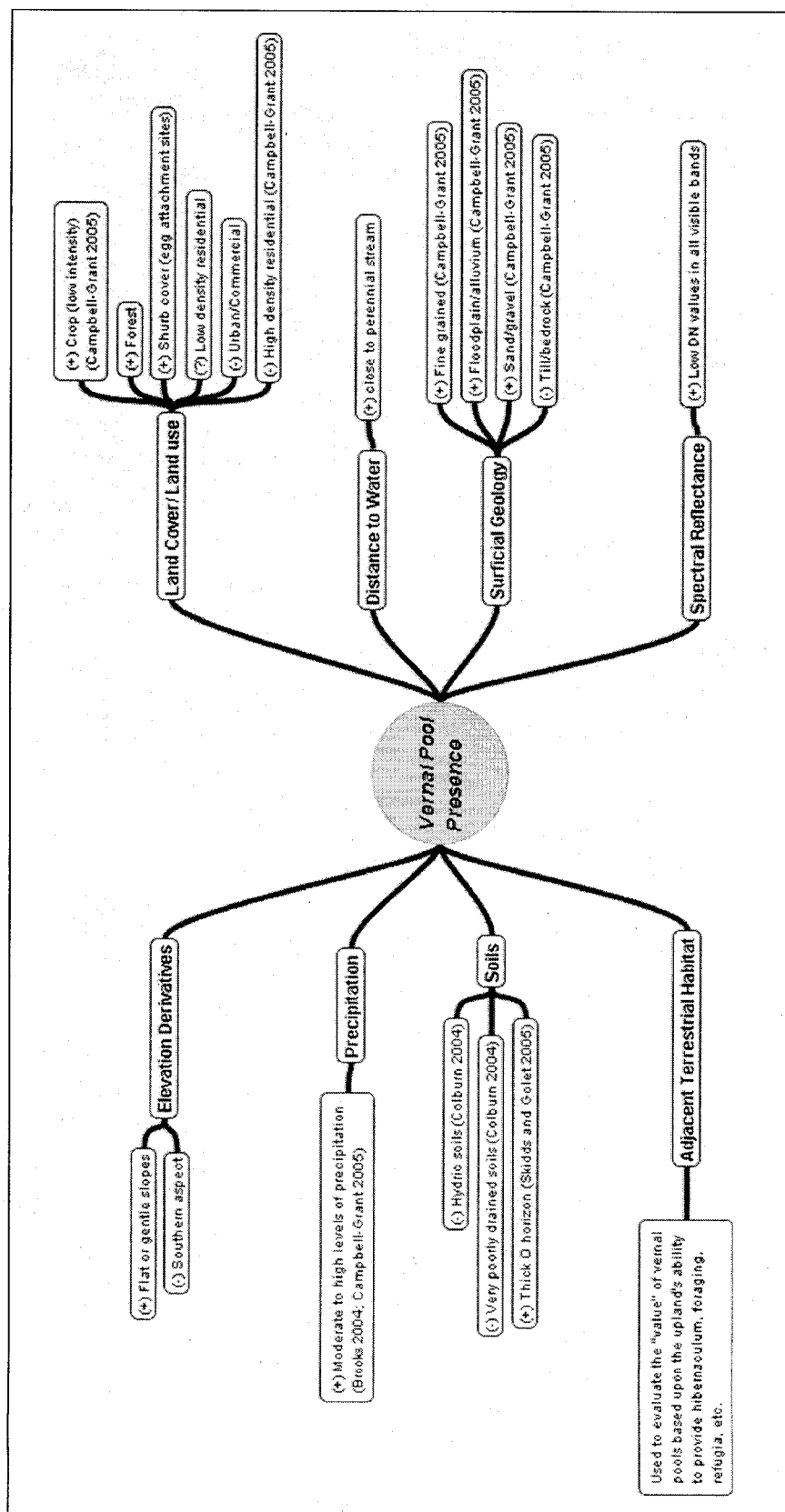
Figure 5: Initial brainstorming regarding predictors of vernal pools in the landscape.

52

These data have not been field verified, and should not be confused with the Certified Vernal Pools. The layer is comprised of more than 29,000 potential pools (National Heritage and Endangered Species Program 2000), and was used to help validate model output results.

For model building and validation, vernal pool *absence* was as important as vernal pool *presence*. The specific modeling techniques chosen for this project required knowledge of the conditions under which vernal pools do and *do not* exist. For this reason, a point layer containing "absent" points was created. The new layer was created using ArcCatalog and edited using ArcMap. It was created with the Massachusetts State Plane NAD83 projection. Fifty points were chosen in each of the 4 training and validation sites totaling 400 validation points to match the 198 training pool points and 205 validation pool points. Points were selected based upon photo interpretation of MassGIS 1:5000 (0.5 m) color ortho photos. Slope, NWI, and land use were used as supplemental layers in decision-making for choosing absent points. Points were designated as "absent" if there was certainty that a vernal pool was not present: such places included areas of significant development, like buildings, roads, and parking lots; areas of extreme slope where water would not pool; and obviously dry areas in forested and open areas (typified by high red, green, and blue DN values on the imagery). Places that had similar physical characteristics as vernal pools (mainly other wetlands) were avoided so as not to confuse the models.

53
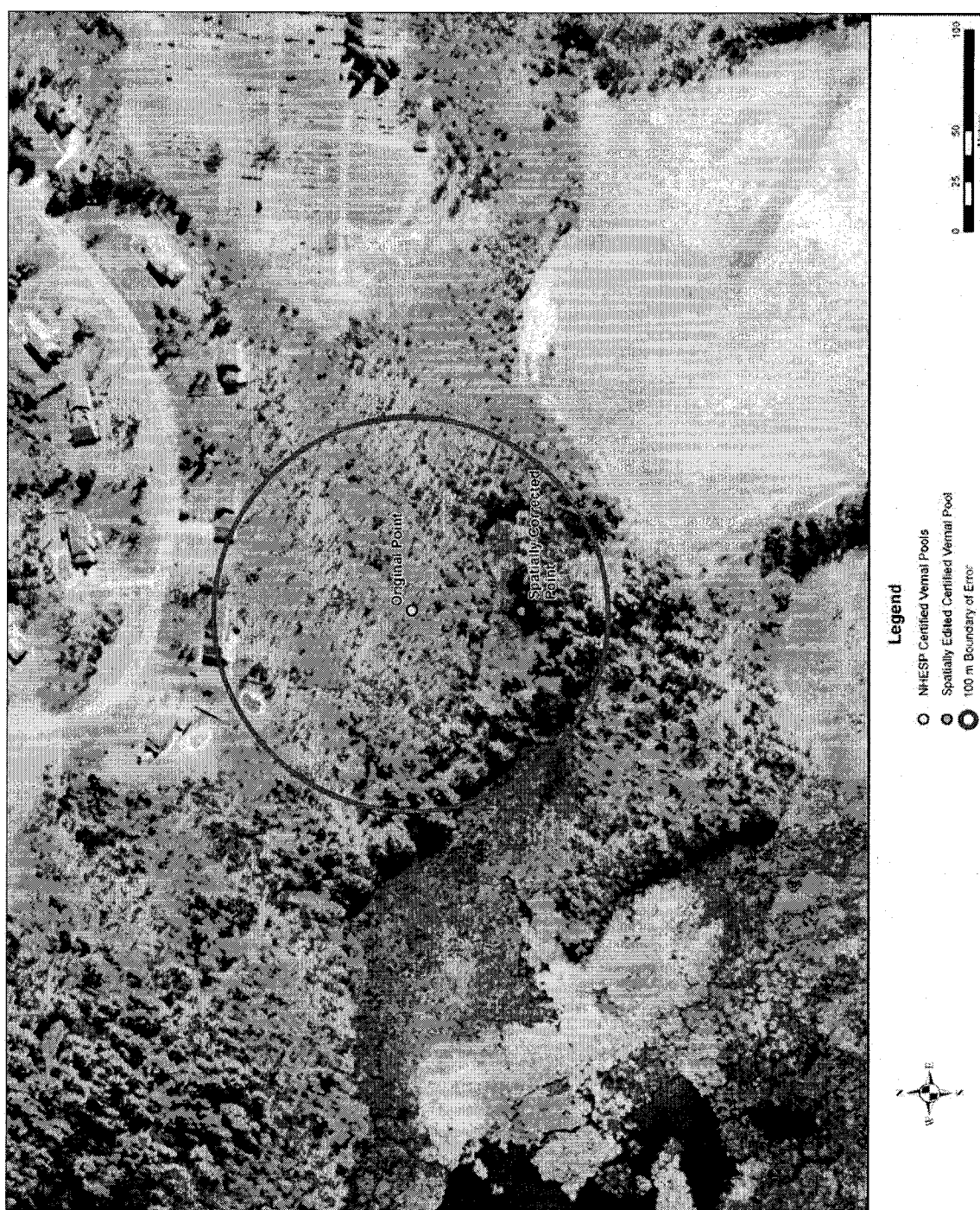
Figure 6: Illustration of CVP spatial correction method. In this example, the original point clearly does not overlay a vernal pool, as evidenced by high reflectance in all three bands (visually bright). The corrected point lies within the boundary of error and has the characteristic tone, texture, shape, and land association (site) as other, known vernal pools.

54

Besides the vernal pool layers, there were a number of other layers used in the analyses. The 1:5,000 color ortho photo imagery was integral to this project. Not only was it used for visual assessment of vernal pool locations and model output, but the RGB (Red, Green, Blue) values were also used as model inputs. The imagery was flown for the entire state in April of 2001, when the deciduous trees were still bare and there was little or no snow left on the ground. The spatial resolution of the imagery is 0.5 m (MassGIS 2001), and the radiometric resolution is 8-bits. Each tile covers 16 km^2 on the ground. There were multiple images covering the study area, so in order to facilitate the analysis, the individual tiles were mosaicked using Leica Image Analysis Extension for ArcGIS (Leica Geosystems 2006).

One of the other important data layers used in the project was the Digital Elevation Model (DEM). It is a raster layer with a scale of 1:5,000. The cell size is 5 m (MassGIS 2005a). It was created from Digital Terrain Models (DTMs). DTM points were collected at a density sufficient to support 3 m contours while conforming to National Map Accuracy Standards (+/- 1.5 m). Variable density (dependent on topography and ground features) mass points were collected along parallel lines 75 m apart, with spot elevations collected at significant features, summits, and depressions (MassGIS 2003). From these points, a Triangulated Irregular Network (TIN) was created and then converted to a lattice. The final product was an integer raster (rounded from floating point) (MassGIS 2003).

Slope and aspect surface layers were derived from the DEM layer using the Spatial Analyst extension in ArcGIS. Both had a resolution of 1:5,000 and a cell size of 5 m (identical to the DEM). These layers were used as inputs (independent variables/predictor variables) into the models for predicting vernal pool presence. Slope was calculated in degrees and ranged in value between 0 and 78. Aspect, or the direction of the slope, was also calculated, and ranged from -1 (flat) to 360. For modeling and querying purposes, the aspect surface layer was recoded (Table 4).

Table 4: Aspect reclassification rules. Reclassification of the original, continuous values into categorical values, which aided in the description of the conditions under which vernal pools were generally present.

Original Aspect Values	Re-coded Aspect Definitions
-1	Flat
0 - 23; 339 - 360	North
24 - 68	Northeast
63 - 113	East
114 - 158	Southeast
159 - 203	South
204 - 248	Southwest
249 - 293	West
294 - 338	Northwest

The land use layer was also used as an input in the analysis. The layer was created by photo interpretation and automated techniques by the Resource Mapping Project at the University of Massachusetts, Amherst (MassGIS 2002b). It contains land use information stored in polygon format for 1971, 1985, and 1999. The most detailed (37 categories) and the most recent (1999) land use data available were utilized. The scale of the layer is 1:25,000 (MassGIS 2002b), and the minimum mapping unit was 1 acre (large enough that it would not

56

sufficiently map most vernal pools). As a model input, there were too few points in each of the 37 categories for it to be statistically meaningful. For model building, it was important that there was a representative number of CVPs in each land use category (Ducey, M. J., personal communication, June 7, 2006). In other words, the ability of the models to predict vernal pool locations depended upon their ability to determine patterns within the data. With 37 categories and few vernal pools in each one, the models were not statistically sound and were unable to determine meaningful trends. To rectify this problem, the land use layer was reclassified into two different schemes. One contained four categories (forest, wetland, field/open, developed), and the other contained five categories (forest, wetland, field/open, urban development [high density], and residential development [low density]). The five-category reclassification identified an important difference between high and low density development. Commercial and industrial lands were split from residential areas, with the idea that a vernal pool would more likely occur in a low density residential area than a commercial/industrial one. These reclassifications ensured that there were enough pools in each category to perform the necessary statistical analyses.

Additionally, the United States Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS) soils layer was used as a predictor during the vernal pool analysis. This layer was obtained from MassGIS, but is maintained by the Massachusetts Department of Agricultural Resources (DAR) (MassGIS 2005b). This polygon layer was digitized from 1:25,000 published soil surveys, and it had a minimum mapping unit of 1.21 ha (three acres). This layer

57

was very complex, with many different categories and methods of classification. Similar to the land use layer, it was imperative that a representative number of CVPs be in each soil category to enable the models to statistically determine patterns in the data (Ducey, M. J., personal communication, June 7, 2006). Since the soils layer is very complex, it was reclassified to a number of simpler classification schemes. One re-classification was based upon soil type and contained the following classes: fine sandy loam, loamy, loamy sand, sandy, muck, urban land, and rock outcrop. The other method of reclassifying the soils layer was by drainage capability, and the classes included: excessively drained, well drained, poorly drained, and very poorly drained. For both schemes, reclassification decisions were based on information gathered from soil surveys.

Finally, the National Wetlands Inventory (NWI) layer was used for visual assessment of vernal pool locations; it was not an input into either of the statistical models. Like many of the other layers in this analysis, in order to gain meaningful information from this layer, it was necessary to reclassify it into simpler categories. In this case, the specific wetland categories were scaled up to a more general level on the Cowardin Wetlands and Deepwater Habitats Classification hierarchy (Cowardin et al 1979). For example, the category "PFO1E," which translates to "Palustrine Forested, Broad-leaved Deciduous, Seasonally Flooded/Saturated," would be scaled by two levels to "Palustrine Forested Wetland." The System and Class hierarchy levels were used to reclassify the layer.

58

Once all of the layers were acquired and preprocessed, the data had to be properly arranged for analysis. To input information about the present and absent vernal pool points into the models, ancillary information about those points had to be compiled into a single table. To create this table, the training present and absent layers were first merged using ArcMap, resulting in a shapefile/table with 398 total points (198 present, 200 absent). Next, this combined table was overlayed with the layers selected for modeling, effectively "drilling down" under each point and extracting information about each of the ancillary layers. The overlay was performed using the "point intersect tool" in Hawths Analysis Tools for ArcGIS (an external extension for ArcGIS). The selected layers included: land use; soils; slope; aspect; and bands 1, 2, and 3 (BGR) from the color aerial photography. The resulting shapefile contained a table with the information from each of the abovementioned layers appended to each present and absent point. The same process was completed for the validation set of present and absent points.

Descriptive statistics for the CVPs used as training data were calculated in an attempt to preliminarily describe the conditions of vernal pool presence. For continuous variables (slope, and the three bands of imagery), the minimum, maximum, mean, median, mode, and standard deviation were reported. For categorical variables (land use, soils, and aspect), a count of the number of points in each category was presented, as well as the percent of the total represented in each category.

59

Analysis

This study was performed on two levels. First, strictly statistical models were created to explain the conditions in which vernal pools are typically found in the landscape. Those models were each statistically evaluated on an independent data set. Once these models performed acceptably, they were translated into cartographic models. The cartographic models were spatial representations of the output from the statistical models. While they should perform similarly in both the statistical and spatial realms, both analyses were completed to test that assumption.

Statistical Modeling

Logistic Regression. The logistic regression modeling technique was utilized in an attempt to reproduce and/or build upon the work of Grant (2005). In order to determine which combination of the 7 independent variables would provide the best model, Akaike's Information Criterion (AIC) was employed (smaller values indicated worse fit) (Akaike 1979; Bonney, 1987). The choice not to automatically include all variables into the model was made to ensure that both the simplest and the most effective model was selected from the numerous possible models. A stepwise logistic regression was performed on the training data set (of present and absent points) using SAS 9.1. Parameters for entry into the model (SLENTRY) were very relaxed (significance level of .99). This liberal value was to ensure that all independent variables could enter the model and the AIC could be assessed at each level. The significance levels for variables remaining in the model (SLSTAY) were varied to determine if model differences

60

were observed (SLSTAY .05 - .99). These liberal SLENTRY and SLSTAY values were used for data exploration only; more stringent values would be used if it had been the final model (see Hosmer and Lemeshow 2000 for examples). In all cases, the smallest AIC value indicated that the best fit model would include band 2 from the aerial imagery (green band), land use (four categories), and slope.

Based upon these preliminary findings from the stepwise logistic regression, the three independent variables indicated to produce the best model were entered into the SAS PROC LOGISTIC routine. From the test statistics provided in the logistic regression routine, SAS also provided a prediction table, which was generated for the independent validation set of CVPs. The table contained a column that calculated the probability that each point was a vernal pool; essentially, each validation point was statistically classified as a present or an absent point using the results of the logistic regression. This table was analyzed in two ways: first, a liberal cut-off value for success was applied. Second, a more conservative cut-off value was used. The liberal cut-off value was 50%, meaning that if the probability of a validation point being correctly predicted was greater than or equal to 50%, then the model was considered successful for that point (Grant [2005] used 53% as the threshold). The second approach used a 75% cut-off value for success. Using these two approaches, classification errors were calculated as percentages.

Classification and Regression Tree. CART was chosen as a modeling technique because of its ability to handle nonparametric data and both

61

continuous and categorical variables. Also, it provides a list of rules for predicting the dependent variable that can be more easily incorporated into a GIS than many other statistical modeling techniques. Further, since CART examines all explanatory variables at each step, it was unnecessary to predetermine which independent variables should be considered in the model; CART will not use predictor variables that do not enhance the accuracy of the model. For this reason, all variables were added to the models: slope, aspect, raw imagery (three bands, natural color), land use, soil type and soil drainage.

In this study, the CART analyses were completed in S+. Two versions of the CART analysis were performed. The single difference between them was that one analysis used the reclassified land use layer with four categories (forest, field/open, developed, water), while the other analysis used the five-category land use layer which split the "developed" class into low density residential land ("residential") and high density urbanized areas ("urban") (forest, field/open, water, residential, urban). The models were named "CART4" and "CART5," respectively. The defaults of the S+ CART modeling routine were maintained, meaning that splits occurred only if there were more than five observations in a node before a split and terminal nodes were achieved when either the total number of observations for a particular node was less than ten, or when the deviance of the node was less than 1% of the total tree deviance (as cited in Lawrence and Wright 2001). Since an unrestricted CART analysis will generally over-fit the model to even the slightest variations specific to the training data set (noise), cost complexity and cross validation pruning methods were tested in an

attempt to make the models more robust. Both yielded similar results; however, with these data, the two classification and regression trees were relatively simple (in some cases, they can be very complex and therefore, difficult to interpret). Further, there was no scenario in which the deviance remained similar to the original tree, yet the number of end nodes significantly decreased, meaning that pruning some of the end nodes would have reduced the explanatory power of the analysis. After exploring multiple pruning scenarios and testing each output on the independent validation set, the trees with the fewest misclassification errors on the validation set were chosen; in both cases, the trees remained unpruned and contained 20 terminal nodes.

Like the logistic regression, a prediction table was output based upon the statistical results of the analysis. This table was generally composed of ones and zeros, indicating the failure or the success of the model in classifying the independent validation set of points. In a few cases, where the characteristics of a particular validation point did not perfectly fit into the decision tree, the output was presented as a decimal/probability. These types of predictions were expected, since models are generalizations of reality and do not account for every anomalous occurrence. Probabilities that were less than one indicated doubt or confusion in the model. In such cases, model success was determined by a 50% or higher probability of a correct prediction.

Cartographic Modeling

In landscape scale environmental modeling, reliance on only statistical analyses is not sufficient. For management purposes especially, the process of

63

integrating the statistical analyses into a *spatial* context is very important. Understanding if/how certain statistics translate onto the landscape is the crux of environmental modeling, and it is often overlooked. In this study, both the logistic regression models and the CART models were used to create predictive maps of vernal pools across the validation study sites. The cartographic outputs were compared to the statistical ones, and differences in accuracies were recorded if present.

Logistic Regression. Using the results from the PROC LOGISTIC model, three cartographic representations of the model were created, two equal-weighted scenarios and one weighted scenario (Sperduto and Congalton 1996), based on the strength of the independent variables in predicting vernal pools. The first representation was a conservative, equal weighted interpretation of the model. In this version, the statistics from the continuous variables (slope and band 2 reflectance) were queried within a range of one standard deviation of the mean using the Raster Calculator in Spatial Analyst. The categorical variables (land use categories) considered important in predicting vernal pool locations were queried based upon both the results of the initial descriptions of pool count and percent in each category, and the logistic regression odds ratios. The classes with the most vernal pools were considered to be important positive predictors. These same classes were also identified as strong positive predictors by the odds ratios in the logistic regression. Conversely, the classes with the least vernal pools were considered strong negative predictors; similarly, odds ratios that were less than one indicated a negative association with the

64

dependent variable and aided in determining which classes were inversely related to vernal pool presence.

The second interpretation of the model, which was more liberal but still equal weighted, allowed a range of two standard deviations from the mean for the continuous variables. The categorical variables did not change in this interpretation. The queries were executed using conditional statements written in the Raster Calculator (Spatial Analyst). A separate statement was written for each of the variables: slope, band 2 reflectance, and land use. The results of each query were then overlayed (intersection), yielding an output that illustrated where all of the criteria for vernal pool presence converged. The results of these model interpretations/queries were two predictive maps that strictly portrayed vernal pool presence or absence.

The third interpretation of the model was one that weighted the independent variables based on their maximum likelihood coefficients in the logistic regression output. The inverse logistic transformation was used to create an equation that resulted in each raster (cell) being assigned a value between 0 - 1. The equation is as follows:

$$PVPO = \frac{\exp(LP)}{(1 + \exp(LP))}$$

where *PVPO* is the Probability of Vernal Pool Occurrence, and *LP* is the linear predictor fitted by the logistic regression (Guisan and Zimmermann 2000). This transformation is necessary to generate values between 0 and 1. The equation was computed in the Raster Calculator in Spatial Analyst. The resulting map represented the probability of vernal pool occurrence across the validation study

65

areas, rather than only presence or absence. Probabilities were classified as: (1) Low (0 - 25%), (2) Moderately Low (25 - 50%), (3) Moderately High (50 - 75%), and (4) High (75 - 100%).

The resolution of all three of the predictive maps was 5 m. There is little literature regarding the appropriate raster resolution when converting vectors to rasters for modeling. The most conservative method would be to consider the Minimum Mapping Unit (MMU) of the coarsest layer as the output resolution. In this case, the soils layer had a MMU of 3 acres (1.21 ha), which is approximately 110 m by 110 m. For the purposes of mapping vernal pools, 110 m was unacceptable, as it would miss many of the smaller pools. Instead, it was decided that an intermediate cell size between the highest (0.5 m - three bands of imagery) and the lowest (110 m - soils) input resolutions would be acceptable. Of the layers used in this study, the DEM, slope, and aspect layers were all 5 m, so this value was chosen as the output resolution for the predictive maps. Other resolutions were tested; however, the 5 m resolution seemed to preserve a satisfactory amount of model detail without absorbing the smaller pools into the rest of the landscape.

CART. Similar to the logistic regression outputs, predictive maps were created for both of the CART analyses. The CART maps were much more complex, as they required a query for each node on the tree that lead to a "present" prediction. Each node effectively represented a rule for determining the presence or absence of a vernal pool. Queries were written from the initial split (root node), through a series of non-terminal nodes, to each terminal node

66

(Figure 7). Each series of queries leading to a "present" end node were merged.

Once there was a collection of queries for all "present" end node paths, those

products were also merged together using the raster calculator to generate one

map predicting vernal pool presence. The resolution for the output models, like

the logistic regression, was 5 m. This resolution was chosen to be consistent with

the logistic regression model for comparison purposes.
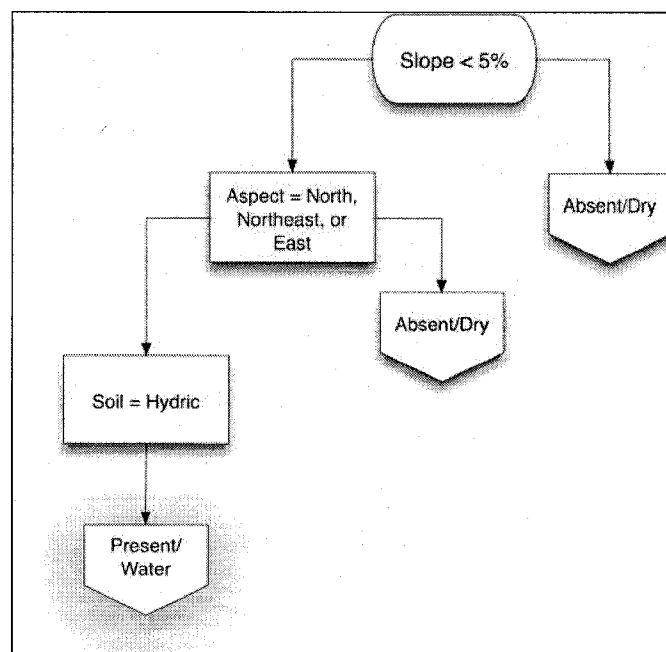
Figure 7: Example of CART model output. In this scenario, the dependent variable is "presence of water." At the top of the tree is the root node, which represents the predictor variable that most minimizes the deviance in the response variable. To create a map output of the CART analysis, the tree must be interpreted and converted into queries. If the condition at each node is true, then the statement proceeds to the left; if it is false, the statement continues to the right. In this case, the query for water presence would be: Slope is less than 5%; Aspect is north, northeast, or east, and soils are hydric.

Cartographic Model Validation.   Error analysis of spatial data means calculating overall map accuracy as well as individual class accuracies. In this study, however, the vernal pool presence "class" was most important; therefore overall accuracy was not used as a measure of success. The two measures of individual class accuracy used in this research were producer's and user's accuracy (Congalton and Green 1999). Producer's accuracy, the complement to omission error, describes how well an individual class on the map matches the reference data for that class (Figure 8). User's accuracy, the complement to commission error, describes how well a mapped class represents what is actually on the ground (Figure 8). In other words, if someone wanted to use the map for navigation, user's accuracy illustrates how well he/she would be able to find a specific class.

Model validation for both the logistic regression and the CART models was done identically. It was performed in two main steps. The first step was to determine how many of the validation pools (original 205 validation points) were correctly predicted by the models (Producer's accuracy). The second step was to evaluate the locations in which the models predicted vernal pools, but a validation pool was not present (User's accuracy). In other words, it was the goal of the modeling activities to predict not only the validation set (where vernal pools are known to exist), but also to predict new pools in the landscape; both accuracies needed to be calculated. This second step was especially important because the National Heritage and Endangered Species Program CVP layer, which is larger than and encompasses the validation set, is not a total

68

enumeration of vernal pools on the ground. For this reason, predicting additional pools beyond the CVP layer was expected and, in fact, desirable.

The first step of the validation process began with converting the model outputs from raster to vector format, which facilitated record-keeping. The conversion was done using ArcToolbox Conversion Tools, and, in order to preserve the original integrity of the data, the edges were not smoothed/splined. Next, each point in the validation layer (both present and absent points) was examined (Figure 9). Records were kept regarding whether or not each model correctly predicted each point. From those data, percent accuracy and percent error were calculated, indicating how well the models performed in trying to predict the validation set only.

The second step of model validation was much more complex. Since model output was not restricted to predicting only occurrences of Certified Vernal Pools, it was important to have some way of evaluating the models as a whole, rather than solely where they predicted the validation set of pools. Overall, the models predicted much more area than just the validation set, which meant there also had to be some method of measuring of how well they did at predicting other pools. Each output polygon (representing where vernal pools were predicted to be present) was examined and overlayed with other layers, such as CVP, PVP, NWI and high resolution imagery, to determine success or failure of the model. Each polygon was assigned a code: 1. a National Heritage and Endangered Species Program CVP, 2. a National Heritage and Endangered Species Program PVP, 3. an NWI-determined wet area, 4. an otherwise wet area, termed here as

69

70

Evaluating Class Accuracy using User's and Producer's Accuracy

A **B** **C**

FIELD REFERENCE

Example A	Bare Soil	Agriculture	Open Water	Total
Bare Soil	3	4	5	12
Agriculture	0	0	0	0
Open Water	0	0	0	0
Total	3	4	5	12

Bare Soil Producer's Accuracy: 3/3 = 100% Bare Soil User's Accuracy: 3/12 = 25%

FIELD REFERENCE

Example B	Bare Soil	Agriculture	Open Water	Total
Bare Soil	0	0	0	0
Agriculture	2	4	0	6
Open Water	1	0	5	6
Total	3	4	5	12

Bare Soil Producer's Accuracy: 0/3 = 0% Bare Soil User's Accuracy: 0/0* = 100%

*While 0/0 is technically not a number, it still represents 100% user's accuracy for this class

FIELD REFERENCE

Example C	Bare Soil	Agriculture	Open Water	Total
Bare Soil	3	0	0	3
Agriculture	0	4	0	4
Open Water	0	0	5	5
Total	3	4	5	12

Bare Soil Producer's Accuracy: 3/3 = 100% Bare Soil User's Accuracy: 3/3 = 100%

In example 'A,' the analyst mapped the entire area as "Bare Soil." Everywhere that there is bare soil on the ground (reference data), the analyst mapped bare soil, yielding 100% producer's accuracy for that class, or 0% omission error. On the ground, however, there are three classes represented: open water, agriculture, and bare soil. Since the analyst classified everything as bare soil, he/she "committed" a lot of area to the wrong classification, or had high commission error. This type of error means that the user's accuracy is very low; a person attempting to use the map would go to the ground and expect to see nothing but bare soil, which is not representative of what is actually there.

In example 'B,' the analyst did not map any of the area as "Bare Soil," though bare soil clearly exists on the ground (reference data). The analyst "omitted" bare soil from the correct class by calling it something else - water or agriculture, resulting in a low producer's accuracy or high omission error. Conversely, since no areas were mapped as bare soil, the commission error is 0%. The user's accuracy for this class is 100%, because nowhere on the map would one expect to find bare soil and find some other category on the ground.

In this example, both producer's and user's accuracies were 100%. Every place where the reference data was bare soil, the analyst classified bare soil (high producer's accuracy, low omission error). Also, there were no areas that were mapped as bare soil that were actually some other class on the ground (high user's accuracy, low commission error). In other words, there were no ground reference points in bare soil that were over-looked on the map.

Legend

Bare Soil (MAP)   Agriculture (MAP)   Open Water (MAP)   Bare Soil (Ground)   Agriculture (Ground)   Open Water (Ground)

Figure 8: Example of individual class accuracy calculations. In this illustration, the Bare Soil class is analyzed to demonstrate the importance of evaluating both producer's and user's accuracy.
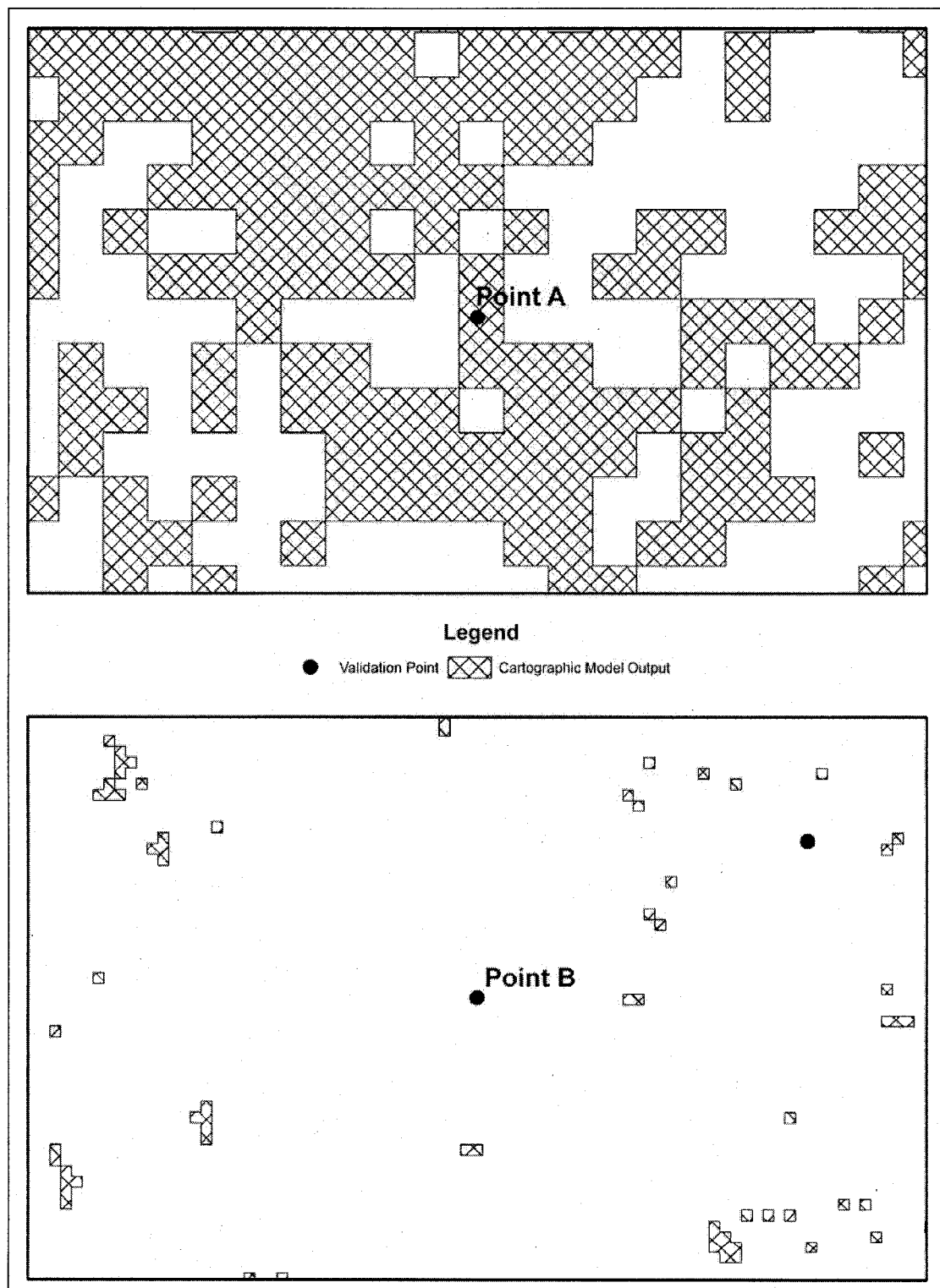
Figure 9: Picture representation of how CART models were evaluated using validation points. Point A (top) was successfully predicted by the cartographic model output, while point B (bottom) was erroneously excluded by the cartographic model.

71

"Possible Vernal Pool" (PoVP). These were defined as areas of interest for field investigation as determined by photo interpretation of 1:5,000 color DOQs (not a PVP, CVP, or NWI area), 5. Other (definitely NOT a vernal pool). Polygons that overlayed CPVs or PVPs were considered successful predictions; those that overlayed "other," non-pool areas were considered errors. Those polygons in NWI areas not consistent with vernal pool presence represented a "gray area" in the classification of errors. While they did not technically predict only vernal pools, they were successful at predicting water in the landscape. For this class, two representations of error were reported: one which considered these polygons erroneous, and another, "fuzzy" report, which considered these polygons to be partially successful predictions. In other words, misclassification of vernal pools as an NWI wetland was considered less severe than misclassifying them as dry upland. Since some of the larger vernal pools are classified as wetlands by the NWI, considering all of them correct or incorrect was not appropriate. Finally, those polygons coded as "otherwise wet" required field investigation to determine success or failure of the models.

Field investigation was, initially, an unrealistic task for this project. The entire validation study area covered 8,911.20 hectares and consisted of four, discontinuous polygons (four separate geographic areas throughout northern and eastern Massachusetts) (Figure 4). With this large tract of land to validate across such a wide geographic range, it was necessary to subset the study area and field sample a representative area within the original boundaries. A 10% sub-sample (891.12 ha) of the total area was extracted for field verification. Using

72

Hawths Tools, four subsets were randomly generated per validation study location (Essex county, eastern Middlesex county, northern Middlesex county, and eastern Worcester county); a total of 16 field validation subsets were created. Each subset was 746 m by 746 m, an area totaling 556,516 m^2 (55.65 ha). Within these subsets, model output polygons designated for field verification were visited on the ground for confirmation of vernal pool presence.

Field visits were completed in August of 2006. Generally, most pools have dried by this time of the year, making it undesirable for vernal pool field work; however, the summer of 2006 was a very wet season and most of the pools were still full. At that time, however, it was impossible to identify pools based upon certification criteria, as obligate and facultative species had already emigrated from the pools. Field verified Possible Vernal Pools were identified based upon common physical features observed at most vernal pools (discussed in Literature Review). There were some verification sites that were questionable as to whether they had been wet earlier in the season. At these places, comprehensive observations were made of the potential basin's morphology, soil moisture, litter cover etc. and a judgment was made as to whether or not it was likely a vernal pool. This scenario was not frequently encountered and does not represent a large percentage of the field validation results; in most cases, vernal pool presence or absence was still very obvious, even at that late time during the season. There was an equally small percentage of points that were inaccessible for various reasons. To definitively decide whether these pools function as vernal

pools, field visits would have to be conducted at the identified sites when they are

biologically active.

74

CHAPTER III

RESULTS

Vernal Pool Descriptive Statistics

Preliminary analyses of vernal pool locations yielded descriptive statistics describing the physical attributes that are characteristic of the vernal pool locations in this study.   First, the land use characteristics of vernal pools were explored. Of the 198 training pools examined, 157 were found in forested environments. Only about 30% of the study area was found to be forested, though almost 80% of vernal pools were found in forested landscapes (Figure 10). Low density development, which occupied about 40% of the study area, accounted for less than 10% of vernal pools.

The soil characteristics of the training set were very variable. They were categorized in two ways: by soil type and by drainage class (Figure 11, Figure 12, respectively). Categorization of pools by soils type revealed inconsistent information as well. Most pools were described as occurring atop fine sandy loam (88 pools), rock outcrops (54 pools), or on mucky soils (29 pools), which were also the most abundant types in the study area (Figure 11). Further, the majority of the vernal pools were reported in well-drained soils (122 pools), which was also the most abundant class over the entire study area. Very poorly drained soils (36 pools) and excessively drained soils (29 pools) comprised 33% of the pools and of the study area.
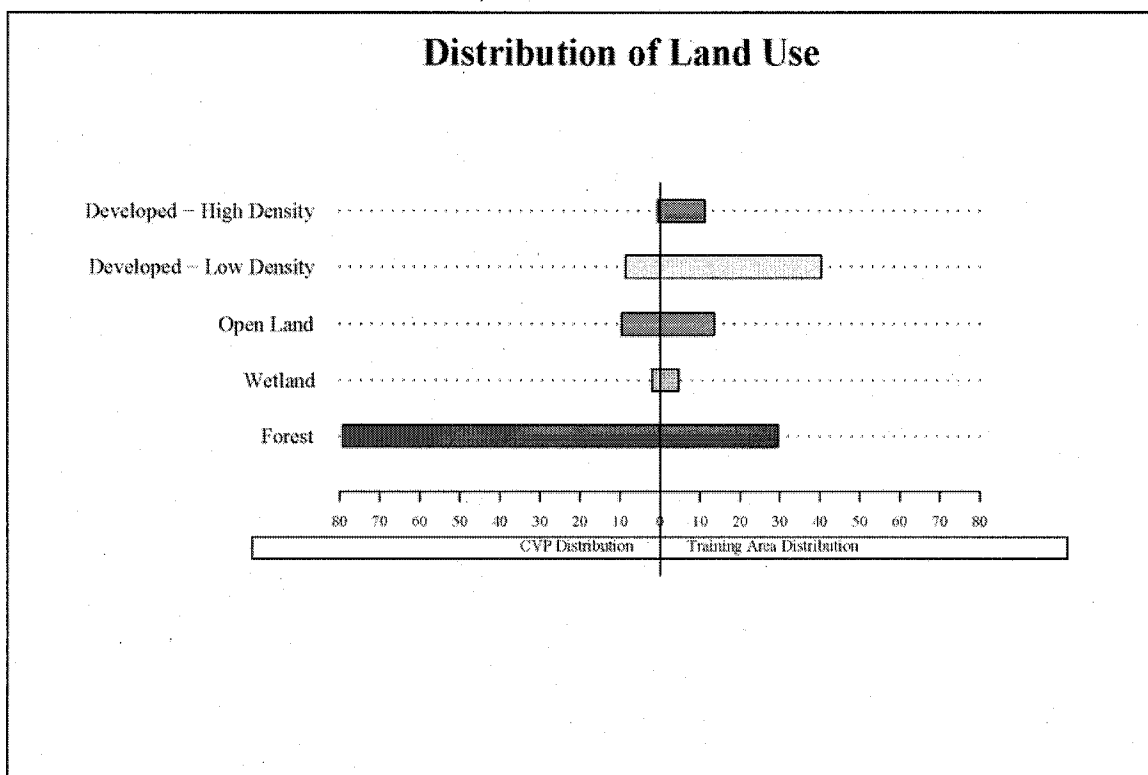
Figure 10: Distribution of Certified Vernal Pools by land use vs. Distribution of land use over the training study area.
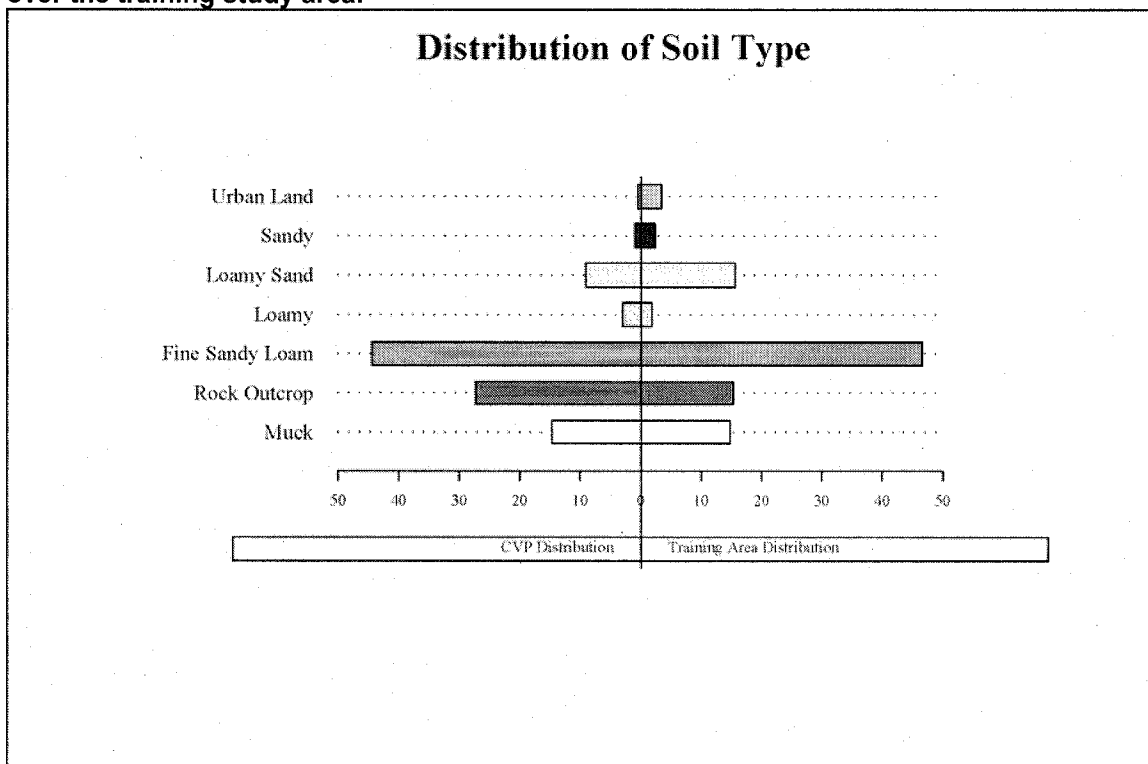
Figure 11: Distribution of Certified Vernal Pools by soil type vs. Distribution of soil type over the training study area.

76

Distribution of Soil Drainage Type
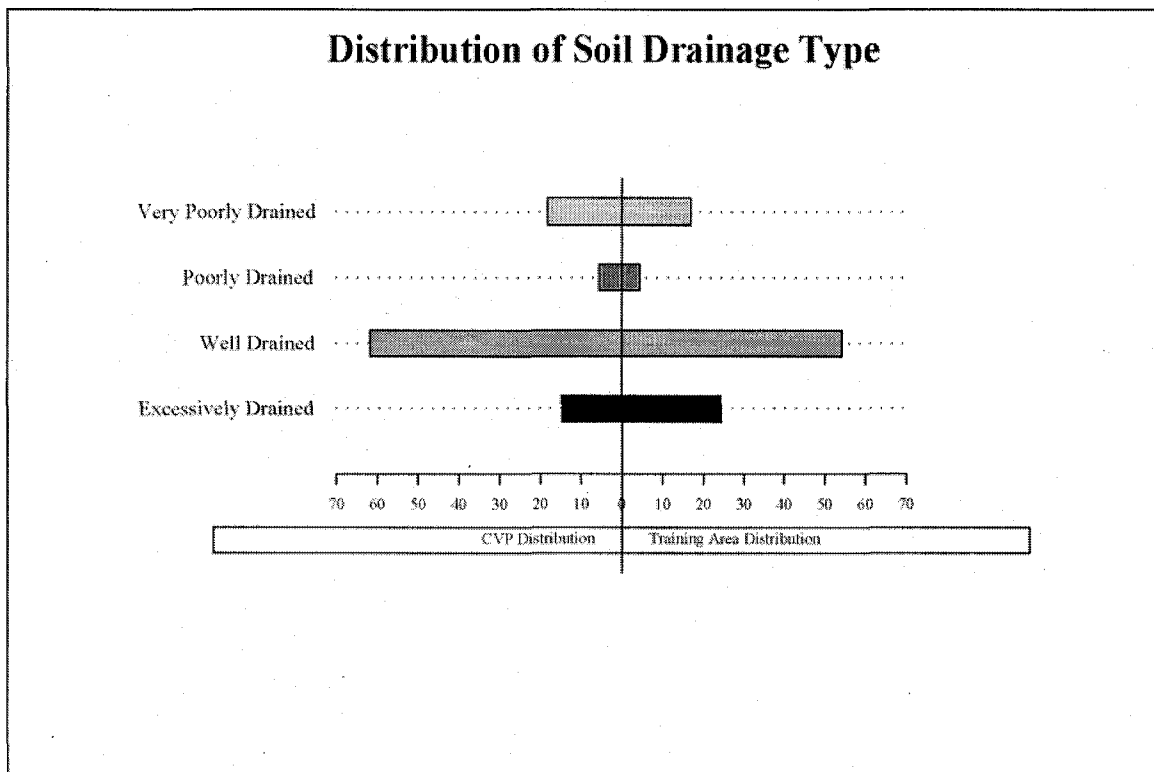
Figure 12: Distribution of Certified Vernal Pools by soil drainage class vs. Distribution of soil drainage class over the training study area.

While NWI was not used in the modeling portion of this study, it was used in multiple steps of the pre-processing methods, and it was used for observation of the types of locations in which vernal pools exist. The majority of the training CVPs were located in either an upland area (29%) or in a Forested Wetland (27%) (Figure 13). Interestingly, forested wetlands comprised less than 10% of the area, though contained many vernal pools.

The final categorical variable examined was aspect. Not surprisingly, the majority of vernal pools were found in flat areas, meaning that there was no aspect (104 of 198 pools). Flat areas also represented the majority of the training study area. The fewest pools were found on south facing slopes (7 pools). All other categories were fairly evenly distributed (Figure 14).

77

Distribution of NWI

Upland
Riverine
Lacustrine
Palustrine Scrub Shrub
Forested Wetlands
Palustrine Emergent Vegetation
Palustrine Unconsolidated Bottom
No Data

70 60 50 40 30 20 10 0 10 20 30 40 50 60 70

CVP Distribution | Training Area Distribution
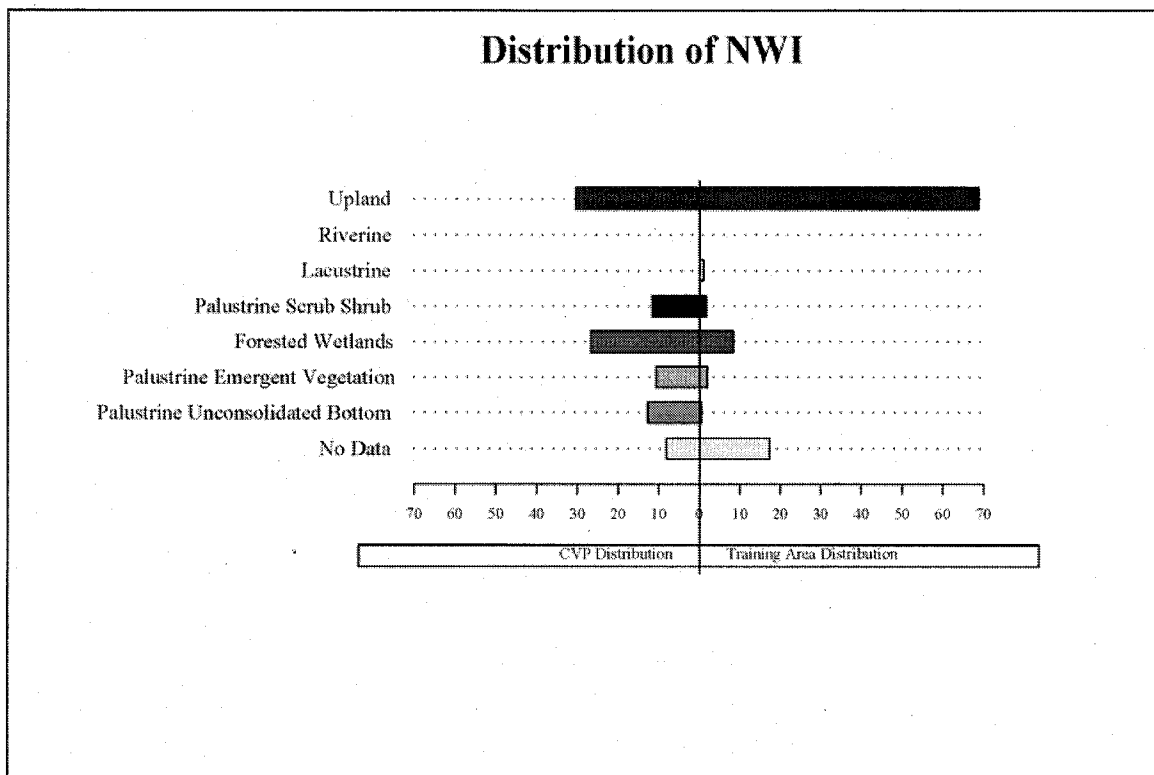
Figure 13: Distribution of Certified Vernal Pools by National Wetland Inventory class vs. Distribution of National Wetland Inventory class over the training study area.

Distribution of Aspect

Northwest
West
Southwest
South
Southeast
East
Northeast
North
Flat

60 50 40 30 20 10 0 10 20 30 40 50 60

CVP Distribution | Training Area Distribution

Figure 14: Distribution of Certified Vernal Pools by Aspect vs. Distribution of Aspect over the training study area.
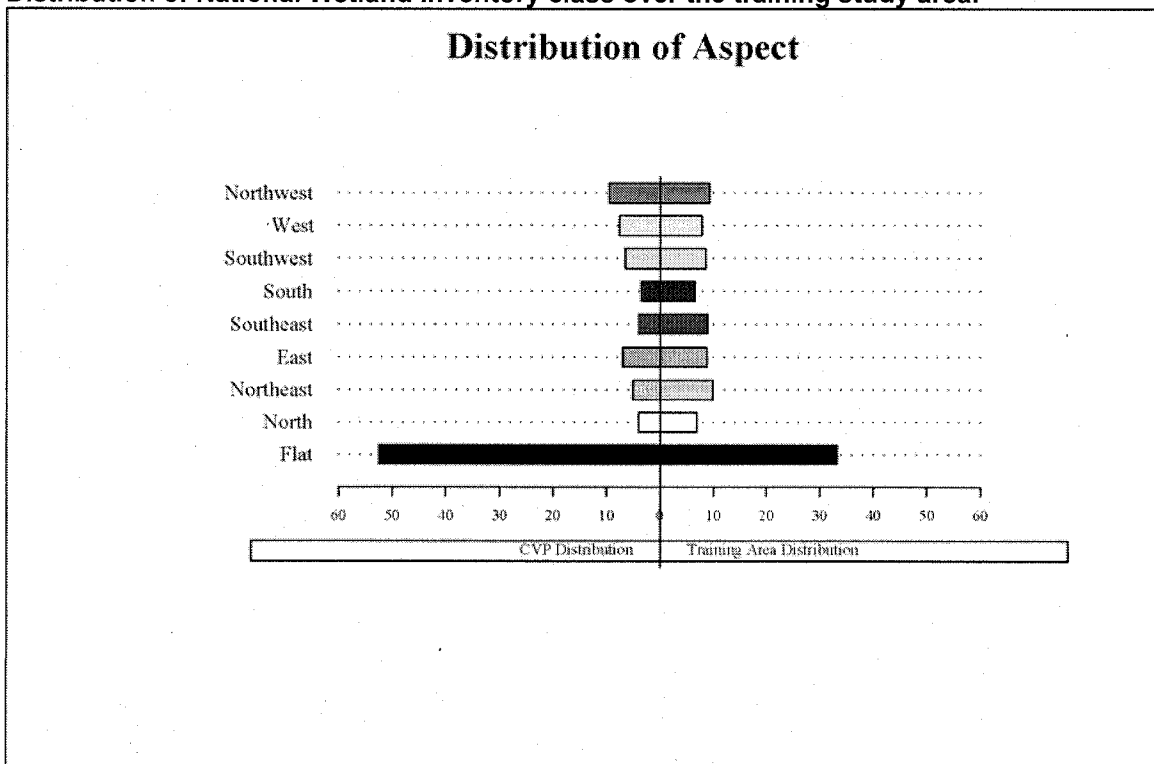
The continuous variables were summarized in a different way than the categorical variables were, since only a frequency distribution was possible for the latter. With slope and the blue, green, and red bands of the imagery, basic statistics were generated (Table 5). The minimum, maximum, mean, median, mode and standard deviation of each were calculated. The slope upon which the certified vernal pools were located ranged between 0 - 12.6 degrees. The mean was 2.4 degrees, with a standard deviation of 3 degrees. Training pools showed similar ranges, means and standard deviations for reflectance values in all three bands of imagery. CVPs showed reflectance values between 0 - 154 in the blue wavelength of light. The mean blue light reflectance was 16.0, and the standard deviation was 28.4. In the green wavelength, the range of light reflected at vernal pool locations was 0 - 143. The mean green light reflectance was 14.8, and the standard deviation was 27.7. Finally, red light reflectance ranged between 0 - 118. The mean was 18.8, and the standard deviation was 24.9.

Table 5: Certified vernal pools continuous variables: basic statistics.

Certified Vernal Pools: Statistics for Continuous Variables				
	Slope	Band 1 (blue)	Band 2 (Green)	Band 3 (red)
Minimum	0.00	0.00	0.00	0.00
Maximum	12.60	154.00	143.00	118.00
Mean	2.42	15.98	14.78	18.81
Median	2.02	4.00	3.00	9.50
Mode	0.00	0.00	0.00	0.00
Standard Deviation	3.01	28.43	27.73	24.94

79

Statistical Modeling

Logistic Regression Model - Model Fit and Predictor Strength Statistics

The logistic regression modeling technique, which used the input variables slope, land use, and the green band of imagery, produced a maximum rescaled R^2 value of .8535. The Akaike's Information Criterion and the -2 Log Likelihood both decreased significantly with the addition of the three variables to the model intercept, indicating that the predictor variables improved the model (Table 6). The difference in the -2 Log Likelihood with the intercept only and with the covariates added was 405.698, and was calculated by the "Likelihood Ratio Test" (a Chi Square statistic). The Chi Square value was significant ($p < 0.0001$), again indicating that the addition of the chosen covariates to the model significantly improved the overall model fit. Finally, the Hosmer-Lemeshow Goodness of Fit Test was not significant (.7980), indicating that the model adequately fit the data.

Table 6: Logistic regression model fit statistics.

Model Fit Statistics - Best Model		
Criterion	Intercept Only	Intercept and Covariates
AIC	552.356	156.658
-2 Log Likelihood	550.356	144.658

Further, the "Analysis of Effects" indicated that all three variables were significant in predicting the presence of vernal pools ($p < .05$) (Table 7). The odds ratios provided a method of describing the strength of the relationship between each predictor variable and the presence of vernal pools. The odds ratios for slope and band 2 of the imagery were less than one (0.789 and 0.959, respectively), indicating that there was an inverse relationship between these two individual variables and vernal pool presence (Table 8). In other words, if slope increased one degree, the odds of finding a vernal pool would increase by 0.789

80

times. This negative relationship between the dependent variable (vernal pool presence) and slope and/or band 2 of the imagery was further solidified by the sign of the maximum likelihood estimate: both are negative (Table 9).

Table 7: Logistic regression analysis of effects.

Type 3 Analysis of Effects			
Effect	Degrees of Freedom	Wald Chi-Square Value	Significance
Slope	1	13.3	0.0003
Green Light Reflectance	1	86.0	<0.0001
Land Use	3	23.0	<0.0001

Table 8: Logistic regression odds ratio estimates.

Odds Ratio Estimates		
Effect	Point Estimate	95% Wald Confidence Limits (Lower - Upper)
Slope	0.789	0.694 0.896
Band 2	0.959	0.950 0.967
Developed vs. Wetland	0.505	0.049 5.185
Forested vs. Wetland	8.078	0.819 79.638
Open Land vs. Wetland	2.163	0.187 24.991

Table 9: Logistic regression maximum likelihood coefficients.

Maximum Likelihood Estimates			
Variable Strength	Coefficient (Max Likelihood Estmiate)	Odds Ratio Estimate	Significance
Intercept	3.1211	N/A	<0.0001
Slope	-0.2372	0.789	0.0003
Green Light Reflectance	-0.0423	0.959	<0.0001
Land use - Development	-1.2268	0.505	0.0087
Land use - Forest	1.5446	8.078	0.0003
Land use - Field/Open	0.2268	2.163	0.6650
Land use - Wetland	1*	1*	N/A

Reference variable to which all other categorical variables are compared.

The odds ratios of the land use categories were interpreted a bit differently. In logistic regression, categorical variables are divided into dummy variables, with

81

one class acting as the reference class to which all other classes are compared. In this regression, the "wetland" class was the reference category. The most dramatic instance of vernal pool presence occurred between forested land and wetlands (reference). A vernal pool was 8.1 times more likely to occur in a forested area than in a wetland area (Table 8). Further, a vernal pool was 2.2 times more likely to occur in open lands than in wetlands, and 0.505 times more likely to occur in developed areas than in wetland areas. The odds ratio of 0.505, since it was less than 1, indicated that vernal pools were negatively associated with developed areas (as compared to wetland areas).

From the logistic regression statistics, SAS generated a prediction table for the points in the validation set (both present and absent). Classification errors were calculated using both a 50% and a 75% threshold for success. The 50% threshold yielded an overall accuracy (including present and absent points) of 90.6%, meaning that 367 of the 405 validation points were correctly predicted by the statistical model. Of the certified vernal pools, 85.9%, or 176 of 205 pools, were correctly predicted. Absent points analyzed at the 50% threshold were correctly predicted in 95.5% of cases (191 out of 200 cases).

When using a 75% threshold for success, an overall accuracy of 77.8% (315 of 405) was achieved. Validation pools were correctly predicted in 64.9%, or 133 of 205 cases. Absent points were correctly predicted in 91% of the validation points (182 of 200).

82

Classification and Regression Tree Model

Two CART models were produced (Appendix A). Both had 20 terminal nodes. Similarly, both models had nine terminal nodes that defined situations where vernal pools would be present. The rules for arriving at those final "present" designations were, however, different. Their differences did not manifest until the fourth level of the trees; the root nodes and the subsequent 2 levels of the hierarchy were identical. Finally, both analyses utilized all inputs except for soil drainage class.

Like the logistic regression routine, the CART analysis was also able to evaluate the models' performances on an independent validation set. CART4 correctly classified (statistically) 93.3% of all points (present and absent), or 378 of 405 points. Of the vernal pools, it identified 94.6% of the pools, or 194 of 205 points. Absent points were correctly predicted in 92.0%, or in 184 of 200 of the cases. CART5 had an overall statistical accuracy of 91.9% (372 of 405 points). It correctly predicted 92.7% (190 of 205), validation pools, and 91.0 % (182 of 200) of absent pools.

Cartographic Modeling

Logistic Regression Mapping

Three cartographic interpretations of the logistic regression model were created. The first one, termed the "conservative model," included values within one standard deviation of the mean for continuous variables and specific classes for categorical variables, as determined from the logistic regression odds ratios.

The rule set for identifying vernal pools that was associated with this model was as follows:

1.   Slope must be between 0 and 5.4 degrees, and;

2.   Green band reflectance must be between 0 and 43, and;

3.   Land use must be forest or open/field.

For mapping purposes, these rules were translated into conditional statements using Spatial Analyst.

The second model, called the "liberal model," included values within two standard deviations of the mean for continuous variables; the queries for categorical variables were the same in this model as in the conservative one. The rule set for determining vernal pool locations in this model was as follows:

1.   Slope must be between 0 and 8.4 degrees, and;

2.   Green band reflectance must be between 0 and 70, and;

3.   Land use must be forest or open/field.

Conditional statements were written to achieve the model output of vernal pool locations.

The third model, the probability model, the following equation was computed:

$$PVPO = \frac{\exp(LP)}{(1+\exp(LP))}$$

where LP =

3.211 - 0.2372*(Slope) - 0.0423*(Band 2) - 1.2268*(Developed Land) + 1.5446*(Forested Land) + 0.2266*(Field/Open Land) + 1*(Wetland).

84

The logistic regression coefficients were used to weight each variable (Table 9). Weighting was based upon the sign and intensity of the coefficients. Output maps were generated for each of four validation study areas (Figures 15 - 18).

Legend

Town Boundaries

Model Validation Areas

NWI - Identified Wetland

removed from model output

Logstic Regression Model Output

Probability of Vernal Pool Occurrence

0 - 25%

25 - 50%

50 - 75%

75 100%

0.5   0.25   0   0.5 Miles
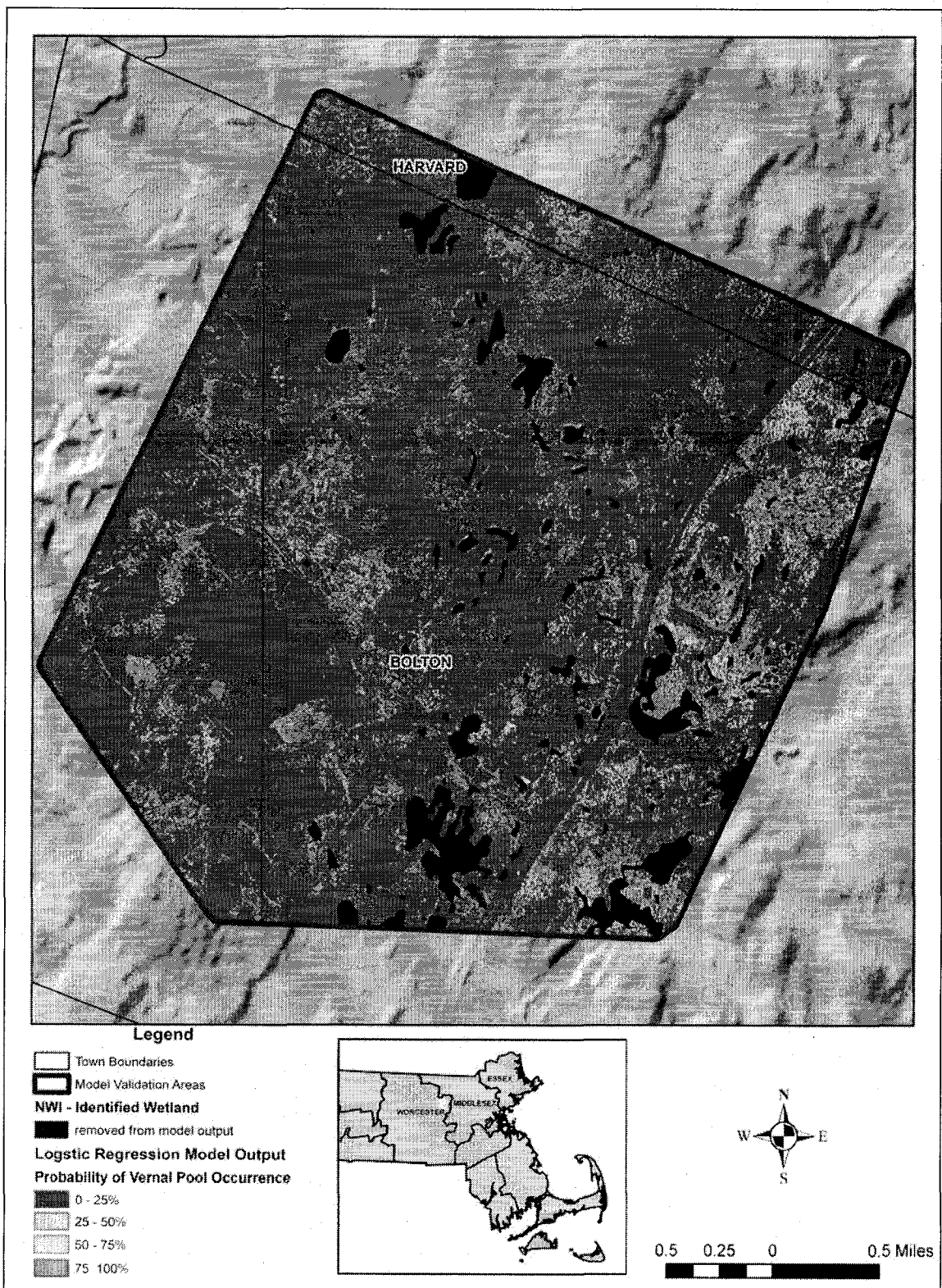
Figure 15: Logistic regression weighted cartographic model: Bolton study site.

86
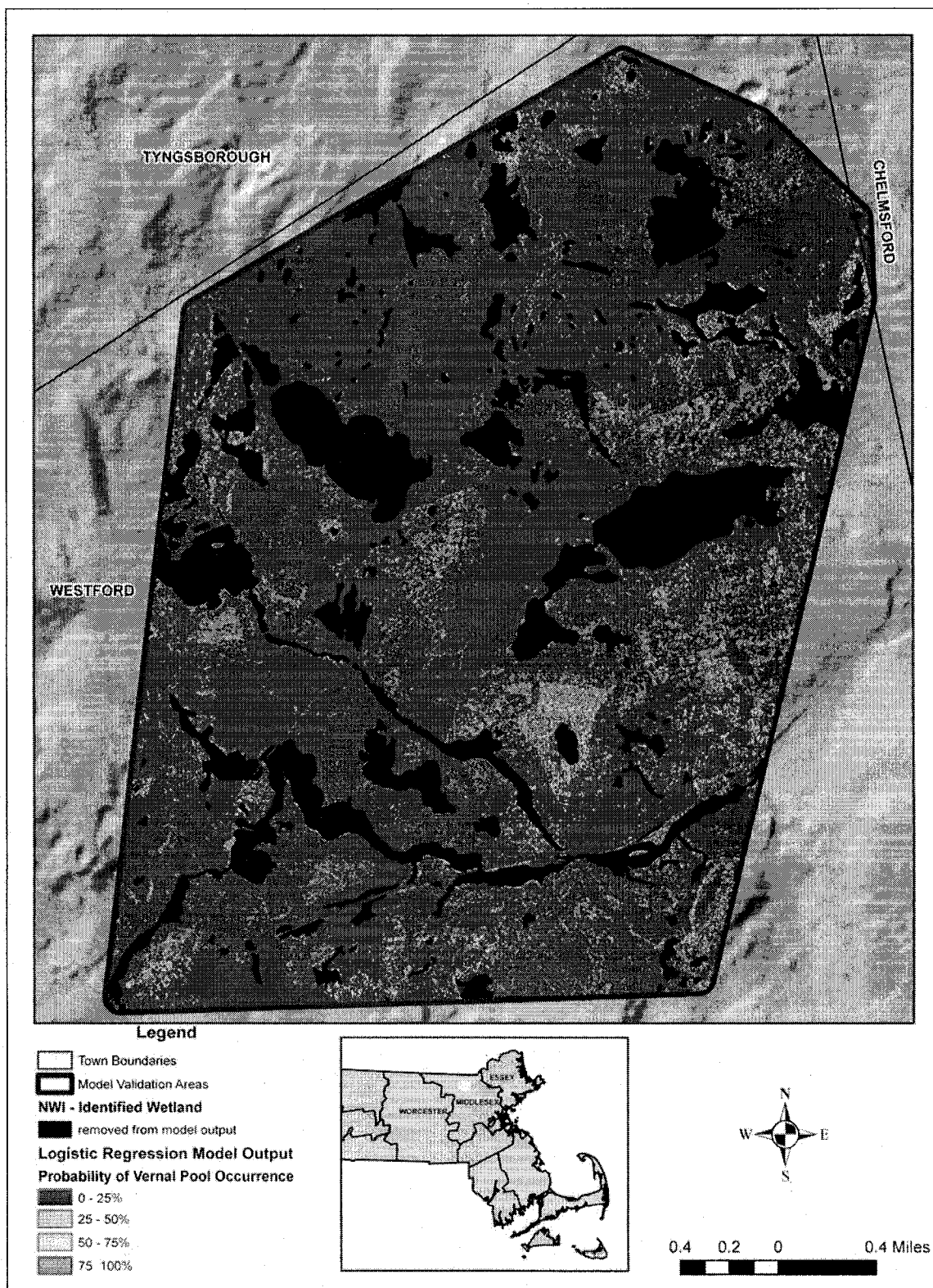
Figure 16: Logistic regression weighted cartographic model: South Westford study site.

Legend

- Town Boundaries
- Model Validation Areas

NWI - Identified Wetland
- removed from model output

Logistic Regression Model Output
Probability of Vernal Pool Occurrence
- 0 - 25%
- 25 - 50%
- 50 - 75%
- 75 - 100%

0.4   0.2   0   0.4 Miles

87
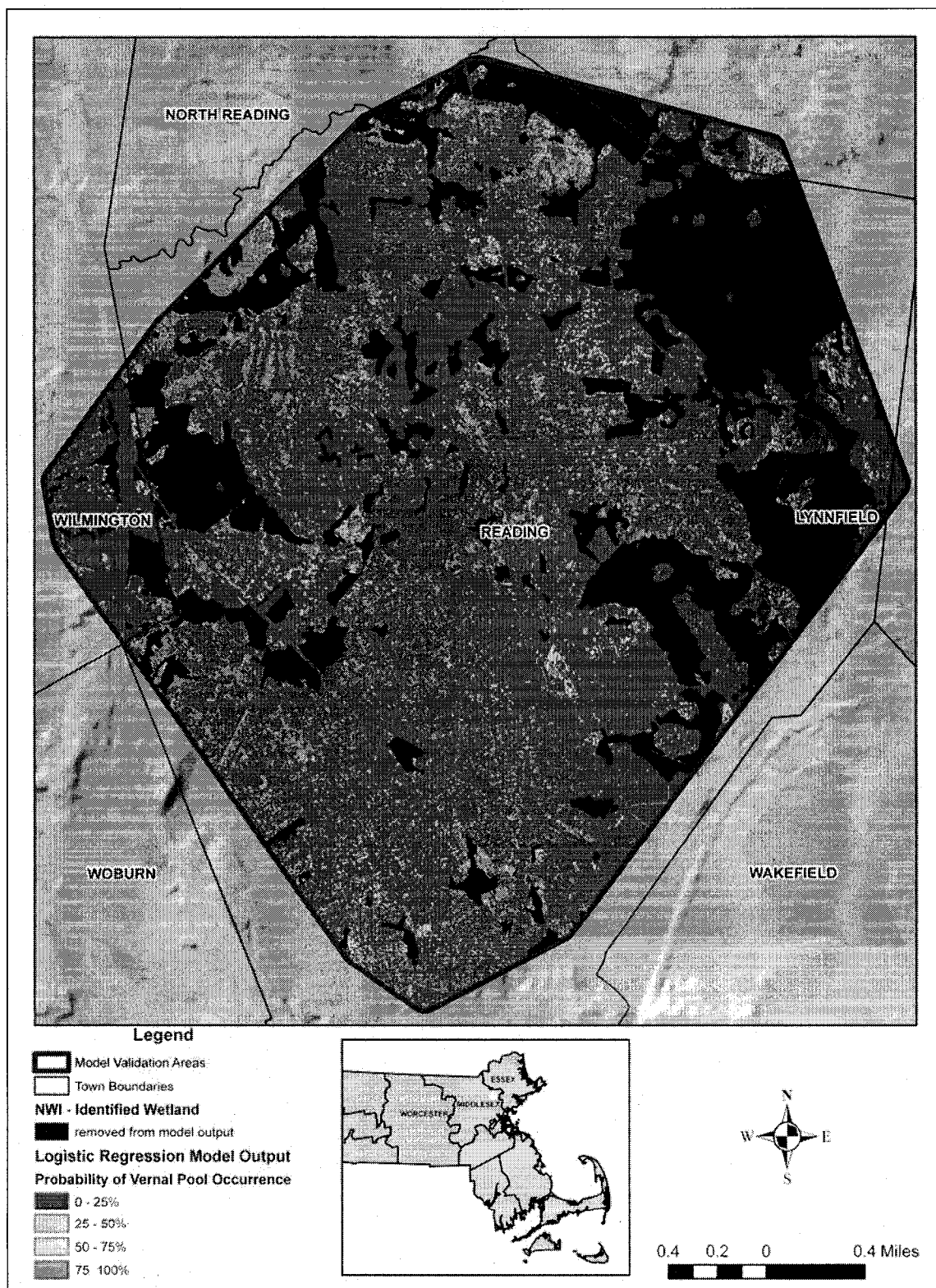
Figure 17: Logistic regression weighted cartographic model: Reading study site.

Legend

Model Validation Areas
Town Boundaries
NWI - Identified Wetland
removed from model output
Logistic Regression Model Output
Probability of Vernal Pool Occurrence
0 - 25%
25 - 50%
50 - 75%
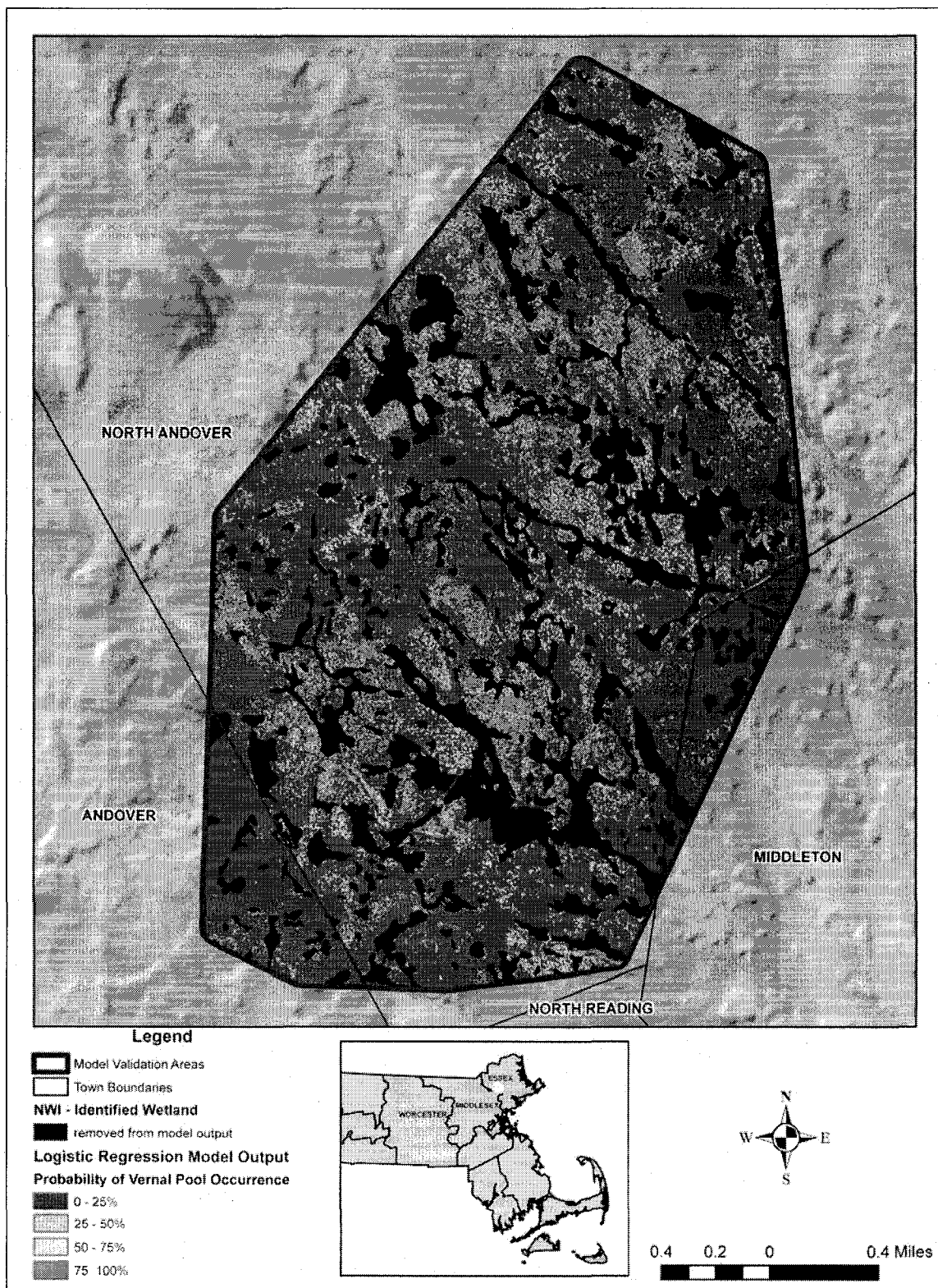75 - 100%

0.4   0.2   0   0.4 Miles

Figure 18: Logistic regression weighted cartographic model: North Andover study site.

89

Classification and Regression Tree Mapping

The full CART models output predictions for both vernal pool presence *and* absence (Appendix A). Mapping of the CART results required writing queries/conditional statements *only* for those nodes on the tree that lead to a "present" classification. Each set of rules leading to a present classification predicted a subset of the model; the model, as a whole, was a combination of each of the subsets. The rule sets for CART4 were as follows (the number of points statistically predicted by each rule set and the percentage of those points that represented correct predictions are found in parentheses):
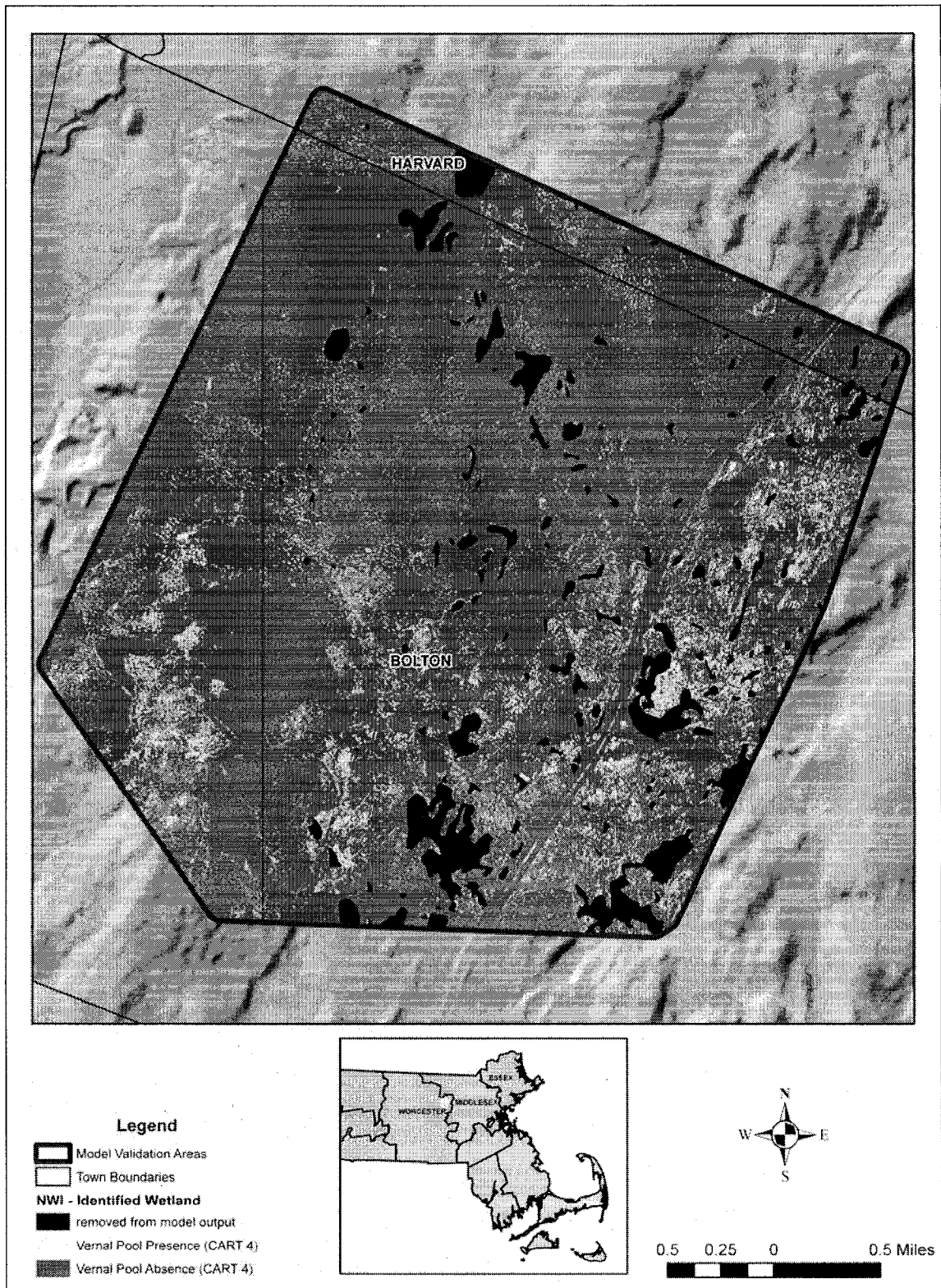
1. (6, 100%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is north, northeast, or south; and slope is less than 3.7 degrees, or;

2. (16, 81.25%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is north, northeast, or south; Slope is greater than or equal to 3.7 degrees; Red band reflectance is greater than one, or;

3. (10, 100%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is flat, east, southeast, southwest, west, or northwest; Land use is developed or field/open; Soil is loamy, muck, or rock outcrop, or;

4. (6, 83%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is flat, east, southeast, southwest,

west, or northwest; Land use is developed or field/open; Slope is less than 5.1 degrees; Blue band reflectance is less than 4, or;

5. (7, 100%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is flat, east, southeast, southwest, west, or northwest; Land use is developed or field/open; Soil type is fine sandy loam, loamy sand, or urban land, Slope is greater than or equal to 5.1 degrees, or;

6. (119, 99.16%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is flat, east, southeast, southwest, west, or northwest; Land use is forest or wetland, or;

7. (14, 100%) Green band reflectance is less than 88.5; Red band reflectance is greater than or equal to 37.5; Aspect is flat, south, southwest, or northwest; Green band reflectance is less than 43, or;

8. (6, 100%) Green band reflectance is less than 88.5; Red band reflectance is greater than or equal to 37.5; Aspect is flat, south, southwest, or northwest; Green band reflectance is greater than or equal to 43; Soil is loamy sand or rock outcrop, or;

9. (6, 83.33%) Green band reflectance is greater than or equal to 88.5; Red band reflectance is less than 118.5; Land use is forest; Slope is less than 7.3 degrees; Aspect is east, southeast, or northwest.

When compiled as a Boolean algebra OR statement, these nine statements predict vernal pool presence within each validation study area (Figures 19 - 22).

91

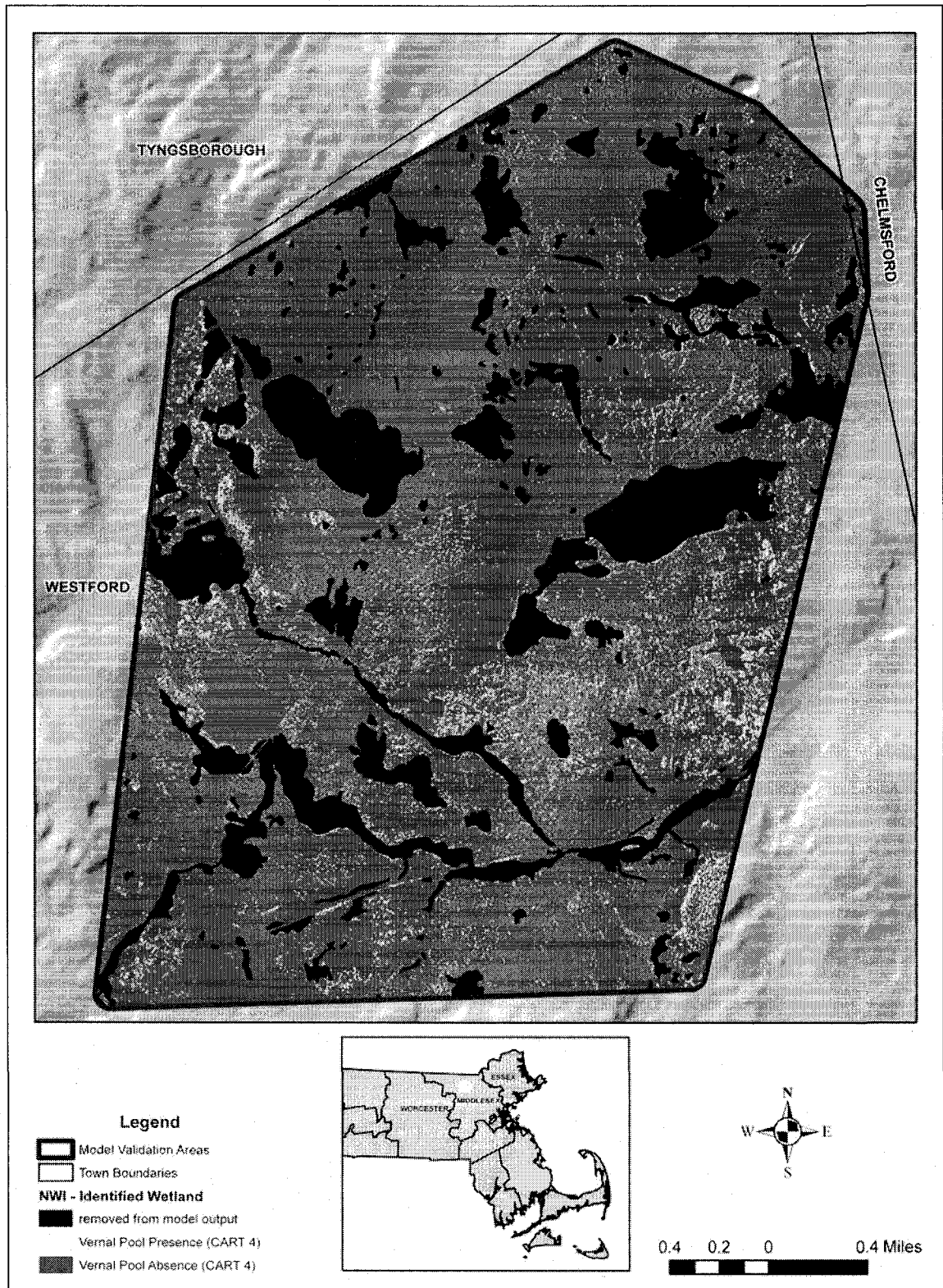**Figure 19: CART4 Cartographic Model: Bolton Study Site.**

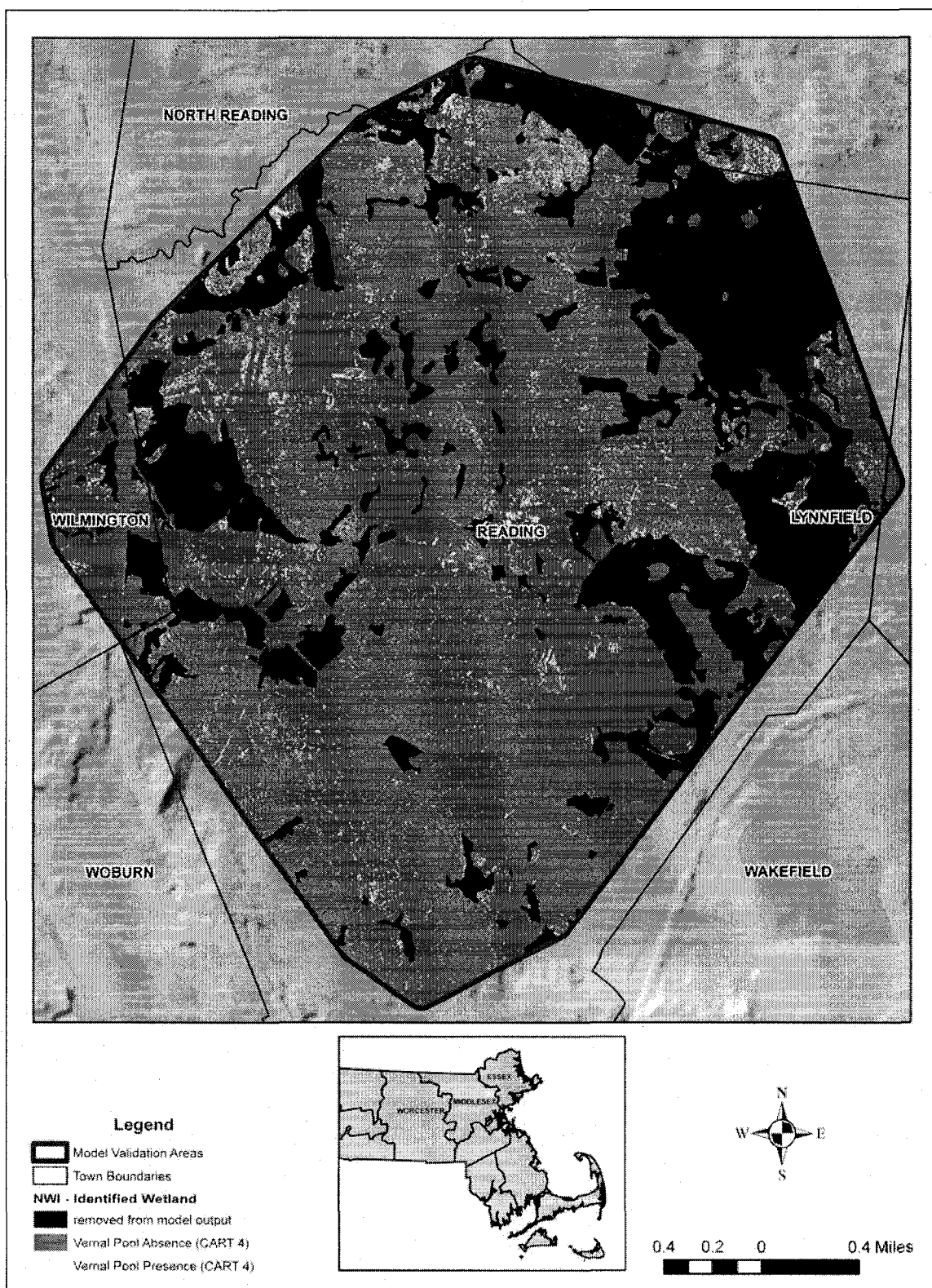**Figure 20: CART4 cartographic model: South Westford study site.**

93

Figure 21: CART4 cartographic model: Reading study site.
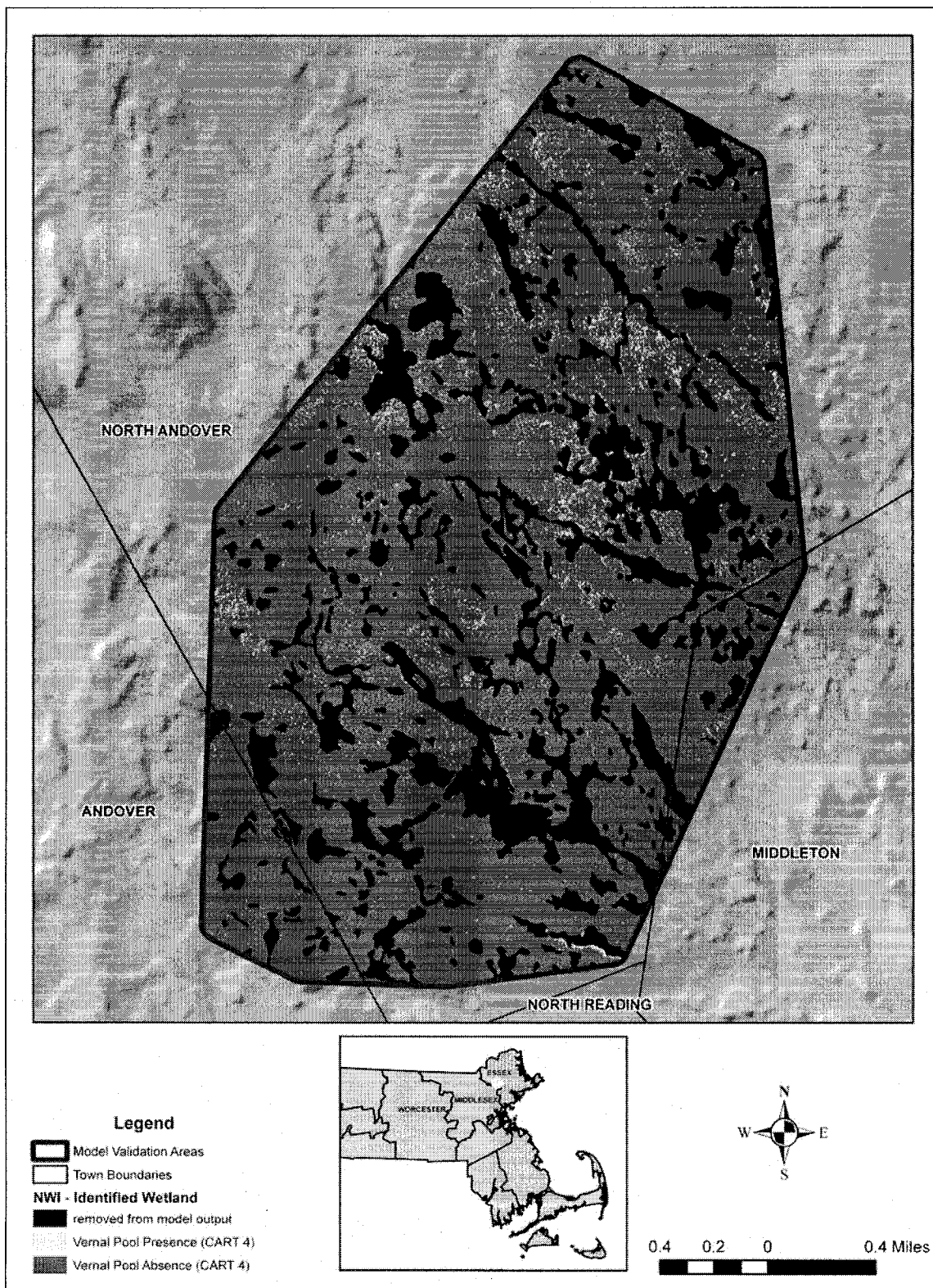
94

Figure 22: CART4 cartographic model: North Andover study site.

95

A similarly structured set of rules was generated for the CART5 model. The rules were as follows (the number of points statistically predicted by each rule set and the percentage of those points that represented correct predictions are found in parentheses):

1. (6, 100%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is north, northeast, or south; slope is less than 3.7 degrees, or;

2. (16, 81.25) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is north, northeast, or south; slope is greater than or equal to 3.7 degrees; Red band reflectance is greater than or equal to one; Land use is urban, forest, residential, or wetland, or;

3. (62, 100%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is flat, east, southeast, southwest, west, or northwest; Blue band reflectance is less than 1.5, or;

4. (17, 100%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is flat, east, southeast, southwest, west, or northwest; Blue band reflectance is greater than or equal to 1.5; Green band reflectance is less than 4.5; Red band reflectance is less than 5.0, or;

5. (21, 90.48%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is flat, east, southeast, southwest, west, or northwest; Blue band reflectance is greater than or equal to

96

1.5; Green band reflectance is less than 4.5; Red band reflectance is greater than or equal to 5.0; Blue band reflectance is less than 8.5; Soil is fine sandy loam, loamy sand, muck, or rock outcrop, or;

6. (43, 100%) Green band reflectance is less than 88.5; Red band reflectance is less than 37.5; Aspect is flat, east, southeast, southwest, west, or northwest; Blue band reflectance is greater than or equal to 1.5; Green band reflectance is greater than or equal to 4.5, or;

7. (14, 100%) Green band reflectance is less than 88.5; Red band reflectance is greater than or equal to 37.5; Aspect is flat, south, southwest, or northwest; Green band reflectance is less than 42.5, or;

8. (6, 100%) Green band reflectance is less than 88.5; Red band reflectance is greater than or equal to 37.5; Aspect is flat, south, southwest, or northwest; Green band reflectance is greater than or equal to 42.5; soil is loamy sand or rock outcrop, or;

9. (6, 83.33%) Green band reflectance is greater than or equal to 88.5; Red band reflectance is less than 118.5; Land use is forest; Aspect is northeast, east, southeast, or northwest.

Again, merging the resulting grids produced a "vernal pool presence" model prediction layer for each validation study area (Figures 23 - 26).

97
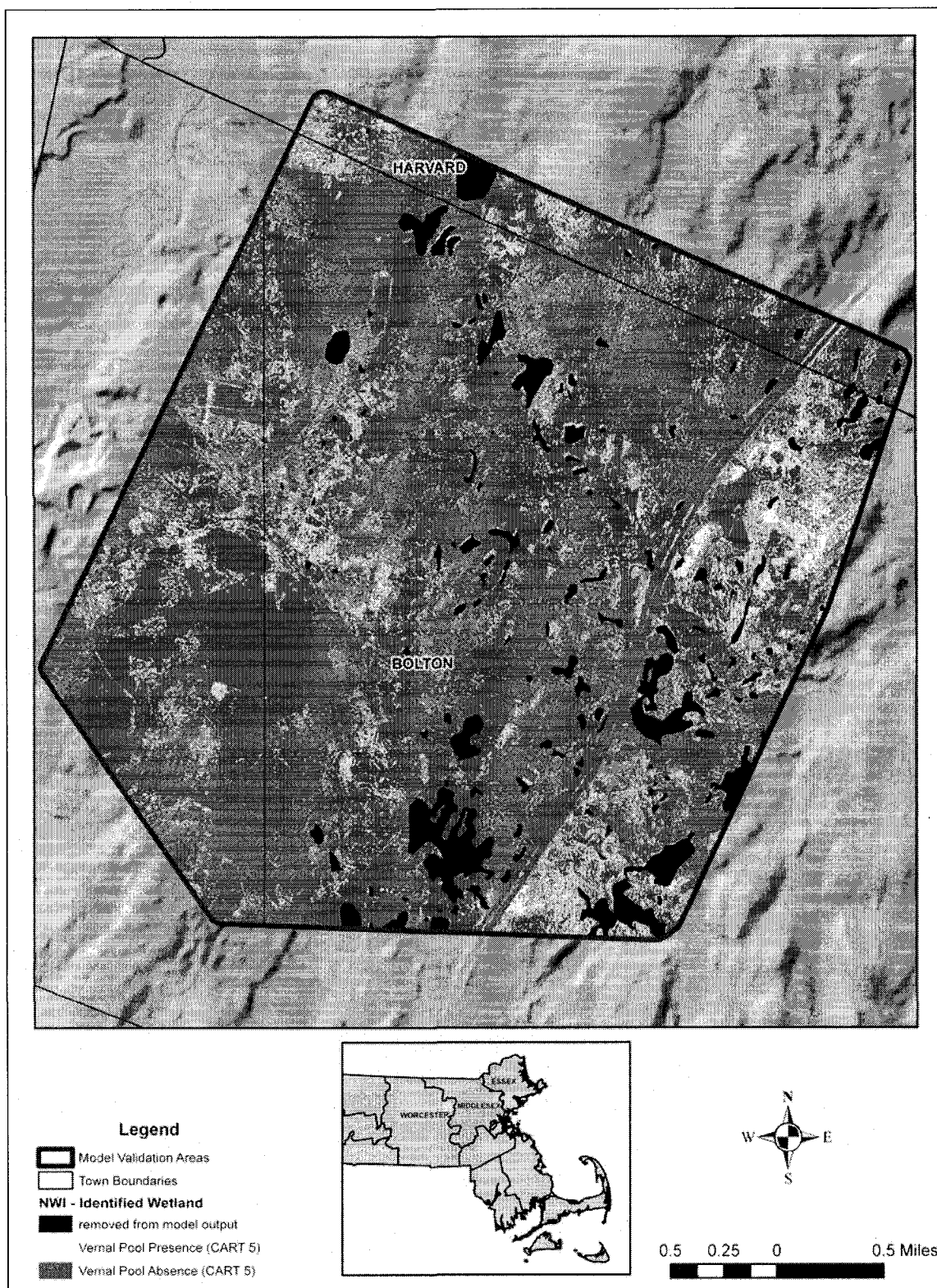
Figure 23: CART5 cartographic model: Bolton study site.

98
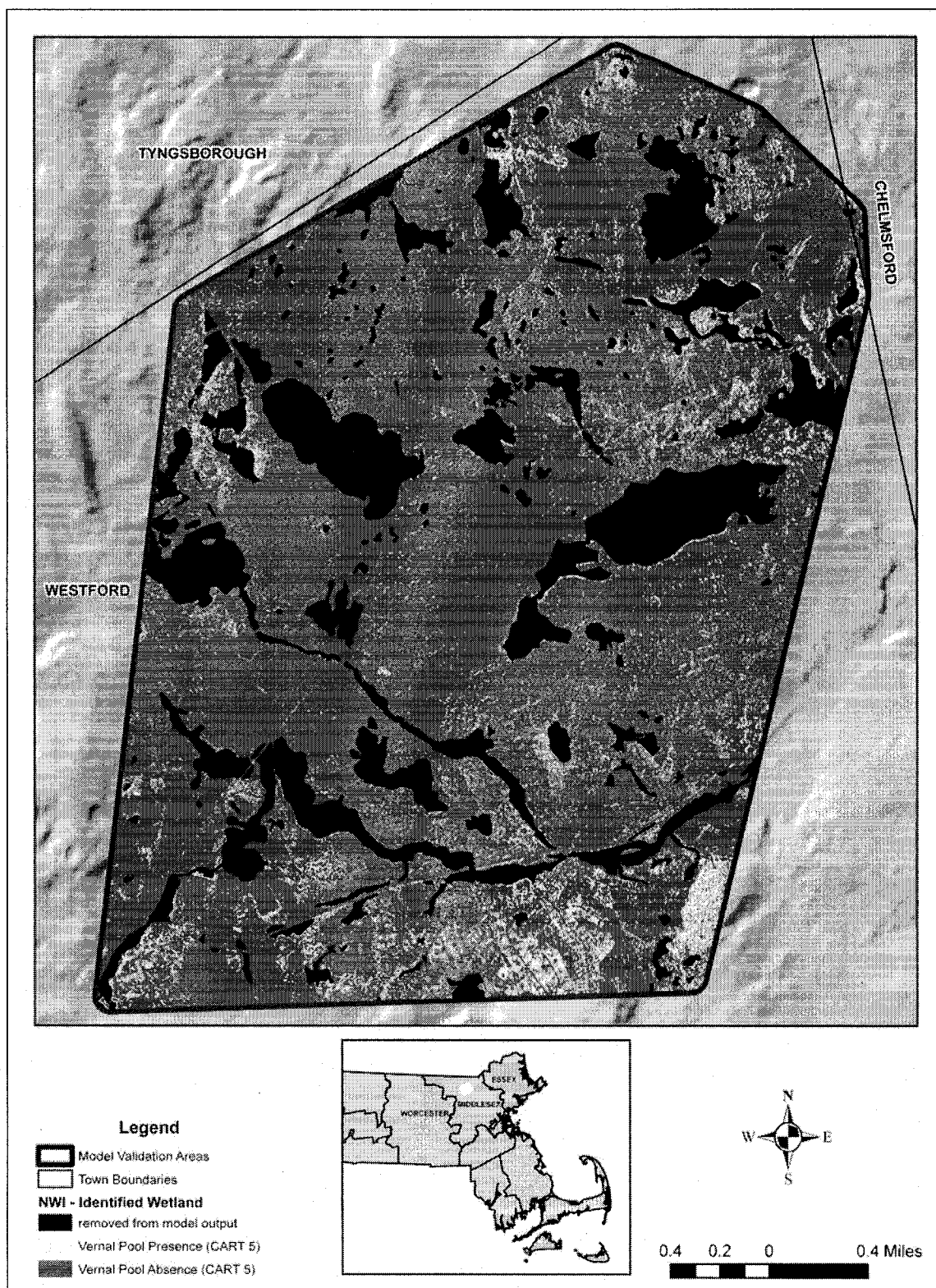
Figure 24: CART5 cartographic model: South Westford study site.

Legend

Model Validation Areas

Town Boundaries

NWI - Identified Wetland

removed from model output

Vernal Pool Presence (CART 5)

Vernal Pool Absence (CART 5)

99
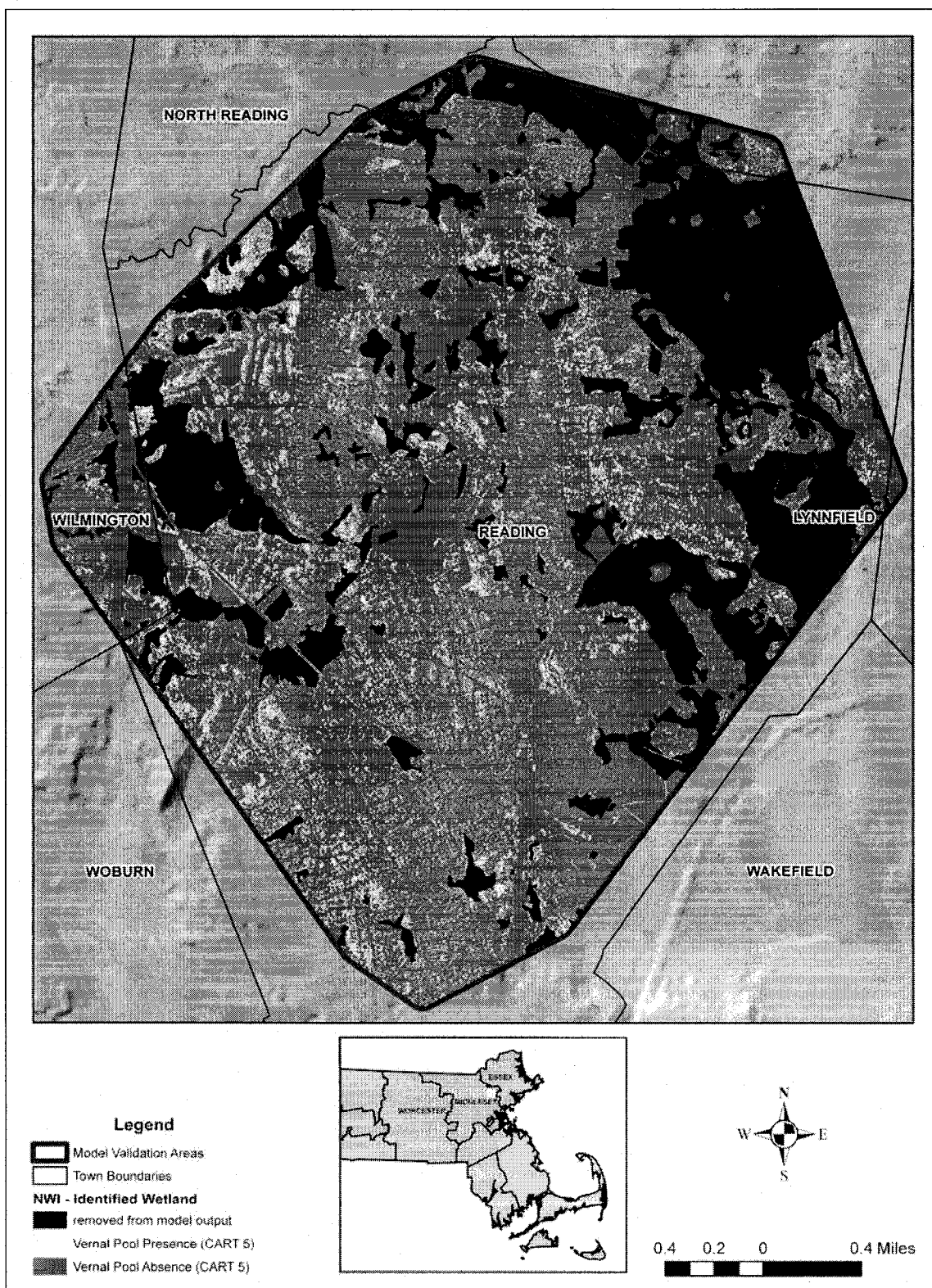
Figure 25: CART5 cartographic model: Reading study site.
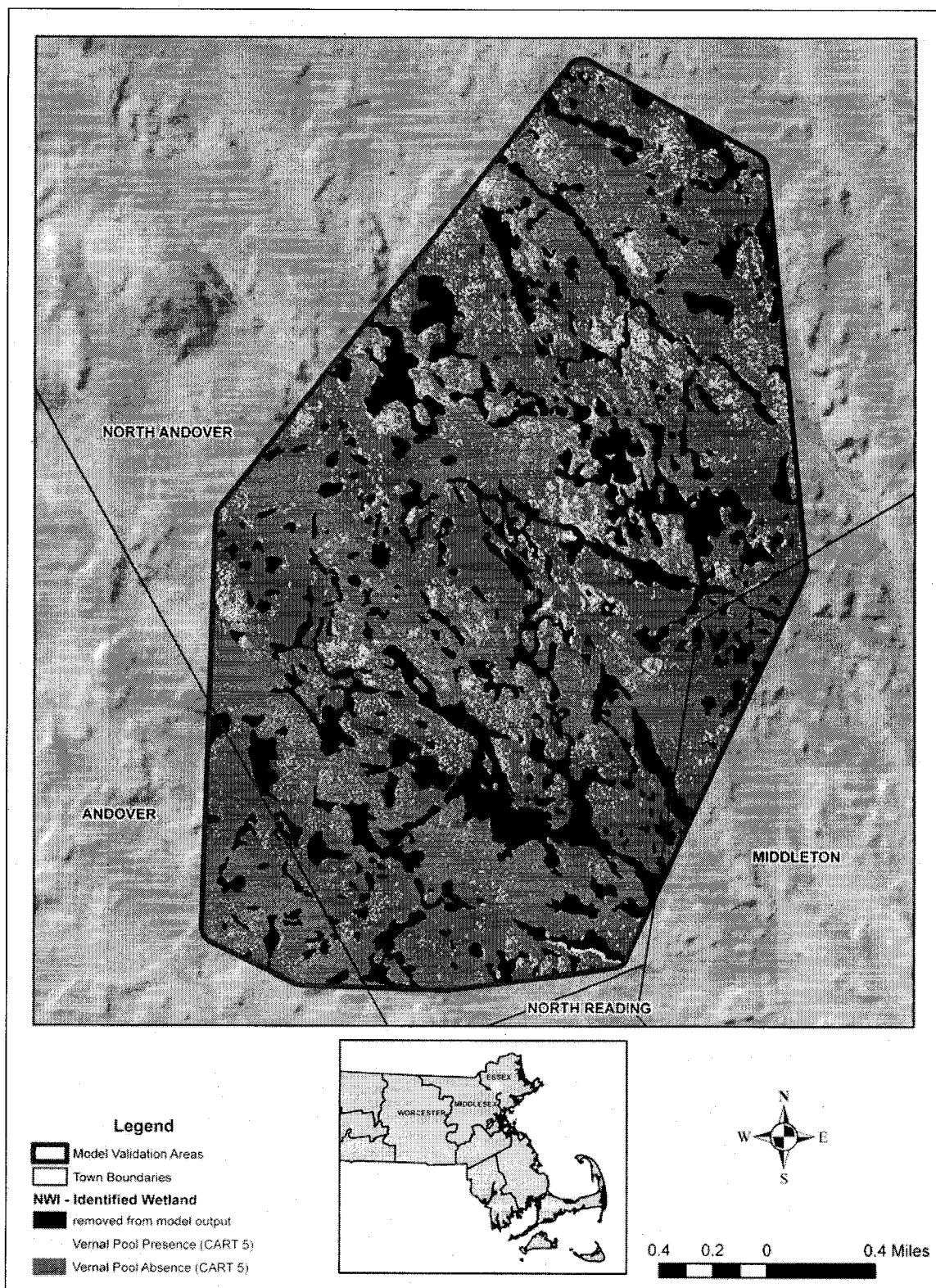
Figure 26: CART5 cartographic model: North Andover study site.

Legend

- Model Validation Areas
- Town Boundaries

NWI - Identified Wetland
- removed from model output
- Vernal Pool Presence (CART 5)
- Vernal Pool Absence (CART 5)

0.4   0.2   0   0.4 Miles

101

Cartographic Model Accuracies – Producer's Accuracy

Analysis of the cartographic models first involved examining each validation point and recording which of the models, if any, predicted it correctly (Figure 9). This exercise was equivalent to determining how well the vernal pool class was mapped when compared to the validation set (Congalton and Green 1999). Producer's accuracy helps to determine errors of omission. At this point in the project, the conservative logistic regression cartographic model was removed from the analysis; visual observation of its performance indicated that it was an extremely inferior model and did not warrant further analysis. The four remaining models (liberal logistic regression, weighted logistic regression, CART4, and CART5) were first evaluated by how well they were able to predict the validation set of Certified Vernal Pools (Table 10). The liberal logistic regression model correctly predicted 68/205 validation pools (33.2%) and 199/200 absent points (99.5%). The weighted (probability) logistic regression model correctly predicted 111/205 vernal pool points (54.2%) and 190/200 absent points (95%). For this model, "moderately high" and "high" probabilities were considered correct predictions (>50% probability). The CART4 analysis correctly predicted 179/205 validation pools (87.3%) and 197/200 of absent pools (98.5%). Finally, CART5 was able to correctly predict 199/205 CVP validation pools (97.1%) and 188/200 absent points (94%).

102

Table 10: Cartographic accuracies: A comparison of how each model performed when predicting validation points.

Model Validation - Based upon present/absent validation points.

Model	Number Correct		Number Incorrect		% Correct		% Incorrect		Total Accuracy (%)	Total Omission Error (%)
	Present	Absent	Present	Absent	Present	Absent	Present	Absent		
CART5	199	188	6	12	97.07	94.00	2.93	6.00	95.56	4.44
CART4	179	197	26	3	87.32	98.50	12.68	1.50	92.84	7.16
Liberal LOGREG	68	199	137	1	33.17	99.50	66.83	0.50	65.93	34.07
Weighted LOGREG	111	190	94	10	54.15	95.00	45.85	5.00	74.32	25.68

Cartographic Model – User's Accuracy

The next step in evaluating model performance was much more involved than the first. In this part of the analysis, each *polygon* in the model output layers (within the 10% subset – 891 ha) was evaluated, rather than each validation *point*. This process helped to determine if the predicted areas on the map were representative of what was on the ground (User's accuracy) (Congalton and Green 1999). Since both of the remaining logistic regression models (equal-weighted liberal model, and the probability model) were not favorable (33% and 54% accuracy in predicting vernal pools, respectively), only the CART models were evaluated in this part of the analysis and considered for intensive field validation.

Understanding what a polygon *means* in this analysis is crucial. For instance, five separate polygons could predict a single vernal pool and all five would be considered correct; in other words, the number of correct polygons is not a proxy for the number of vernal pools detected in the landscape by the models. The same is true of area: the areas reported represent model output and not, for instance, total area of vernal pools within the study area. For example, a vernal pool measuring 1 ha on the ground might be predicted by a total model output of .05 ha and would still be correct in identifying that the pool exists within the landscape. Also, a 1 ha pool could be predicted by a 2 ha polygon, where a portion of the polygon is incorrect, and it would still be considered correct for identifying the vernal pool.

The total number of polygons outputted by the CART4 model within the 10% subset was 9,496 (Table 11). These polygons ranged in area from 0.0006 ha to 7.79 ha, and the total acreage was nearly 82 hectares (Table 12). The model produced 279 polygons that predicted CVPs or PVPs, and an additional 97 "Possible Vernal Pool" polygons were validated in the field (Table 11). The total area of model output over CVPs, PVPs, and field verified Possible Vernal Pools was approximately 3.2 ha. Of the total, 3,013 polygons (about 35 hectares of model output) were identified as NWI wetlands. Finally, 6,063 polygons were incorrectly modeled, as they were actually upland/dry areas. These areas accounted for roughly 43 hectares.

Table 11: CART4 model output polygon counts.

CART4 Model Output Polygons - Model-predicted vernal pool presence. Grey indicates agreement.						
		Photo Interpretation/Field Validation (# polygons)				
Model Predictions (# polygons)		Vernal Pool	Other Wetlands	Dry	Inaccessible	Total
	Vernal Pool	376	3013	6063	44	9496

Table 12: CART4 model output area summary.

CART4 Model Output Area - Model-predicted vernal pool presence. Grey indicates agreement.						
		Photo Interpretation/Field Validation (hectares)				
Model Predictions (acres)		Vernal Pool	Other Wetlands	Dry	Inaccessible	Total
	Vernal Pool	3.216	35.132	43.024	0.277	81.649

With these data, it was possible to estimate commission error. About 4% of the polygons (also 4% of the area) were correctly identified as some form of vernal pool (CVP, PVP, or field verified possible vernal pool). Additionally, 32% of the polygons and 43% of the model output area was identified as NWI wetlands and considered to be partially successful predictions, as they distinguished water presence in the landscape. Combining these two classes into one "fuzzy" correct class revealed that about 36% of polygons and 47% of the output area correctly

predicted water in the landscape. Conversely, about 64% of the polygons and 53% of the output area predicted vernal pool presence were actually some other land cover type. A fractional percentage of the polygons were inaccessible in the field and were not included in the calculations. These statistics mean that 64% of the polygons (Table 13) and 53% of the area (Table 14) were committed to the wrong category, which provides a rough estimate of commission error for the overall model. Visual observation of incorrect predictions of vernal pools revealed that shadows were most often confused with water and accounted for the majority of the errors.

Table 13: CART4 polygon commission error estimate.

CART4 - *commission error (polygons)	
% Correct	3.96
% Correct (fuzzy)	35.69
% Incorrect*	63.85

Table 14: CART4 area commission error estimate.

CART4 - *commission error (area)	
% Correct	3.94
% Correct (fuzzy)	46.97
% Incorrect*	52.69

The CART5 model, within the 10% subset area, produced a total of 12,286 polygons (Table 15). The polygons ranged in size from 0.0006 hectares to 8.97 hectares and totaled 168.83 hectares (Table 16). Of the total, 358 were identified as CVPs or PVPs, and an additional 230 were found in the field (9.08 ha). Additionally, 2,284 polygons were identified as NWI wetlands, which represented about 45 ha. There were 9,393 polygons that were incorrectly identified as vernal pools; these areas covered a total of 115 hectares.

106

Table 15: CART5 model output polygon counts.

CART5 Model Error Estimation - Model-predicted vernal pool presence. Grey indicates agreement.						
		Photo Interpretation/Field Validation (# polygons)				
Model Predictions (# polygons)		Vernal Pool	Other Wetlands	Dry	Inaccessible	Total
	Vernal Pool	588	2284	9393	21	12286

Table 16: CART5 model output area summary.

CART5 Model Error Estimation - Model-predicted vernal pool presence. Grey indicates agreement.						
		Photo Interpretation/Field Validation (hectares)				
Model Predictions (acres)		Vernal Pool	Other Wetlands	Dry	Inaccessible	Total
	Vernal Pool	9.075	44.649	114.928	0.178	168.829

Commission error was estimated from the above statistics. About 5% of the model output polygons (also 5% of the area) represented vernal pools, and an additional 19% were identified as NWI wetlands (26% of the output area). These two classes, when merged together, mean that 23% of the polygons and 32% of the area were correctly identified as water in the landscape. The "fuzzy" error report indicated that the remaining 76% of the polygons (Table 17) and 68% of the area (Table 18) that were described as vernal pools were, in reality, a different land cover and had been committed to the incorrect category. Visual assessment of the incorrect polygons again revealed a high percentage of them to be shadows.

Table 17: CART5 polygon commission error estimate.

CART5 - *commission error (polygons)	
% Correct	4.79
% Correct (fuzzy)	23.38
% Incorrect*	76.45

Table 18: CART5 area commission error estimate.

CART5 - *commission error (area)	
% Correct	5.38
% Correct (fuzzy)	31.82
% Incorrect*	68.07

CHAPTER IV

DISCUSSION

Overview

The work presented here has attempted to create a spatial-statistical model that can predict vernal pool locations in the landscape. The goal was to create a cost and time efficient method of inventorying vernal pools by focusing photo interpretation and field efforts in certain areas where vernal pools are likely to exist. The methods chosen to model vernal pool locations were logistic regression and classification and regression tree analysis. These statistical methods were employed and their results were translated into a map output. Both logistic regression and CART had favorable statistical results; however, logistic regression's cartographic results were far inferior to CART's. The CART models showed very low omission error, but tended to have high commission error due to the spectral confusion between water and shadow.

Interpretation Considerations

Interpretation of the descriptive statistics and model-generated rules must be done with special consideration for the fact that the scale and minimum mapping units of the individual GIS layers affect the results. For example, in the description of soil types underlying the training set of vernal pools, it is reported that a significant number of the pools are found atop rock outcrops (27%). In reality, it is unlikely that vernal pools are actually forming over rocks, but the soils

108

layer is highly limited by its 3 acre minimum mapping unit. A more likely situation is that the vernal pools are occurring on soil units covering less than 1.21 ha (3 acres) within these areas of large rock outcrops.

Similar types of observations can be made within the rule-based CART classifications; at times, the rules may seem counterintuitive for vernal pool prediction, and the scale of the inputs may be what is responsible. In addition to scale and minimum mapping unit considerations, however, recognition that CART is able to reveal/predict exceptions to the most common sets of predictive characteristics is necessary. In some cases, the coarseness of the data may be the reason for counterintuitive results, and in others, the model may be identifying special cases where pools exist that were not previously known. Neither of these types of results is necessarily bad. They may or may not produce a set of characteristics that are accurate when ground verified (i.e. the soil may not actually be rock outcrop); however, they serve their intended purpose for predicting vernal pools in the landscape based on the data available. In other words, the interest in modeling vernal pools was to find new pools in the landscape, not to accurately define ground-verified physical characteristics at pools. The models are able to sort through the GIS data, accurate or not, and find patterns that distinguish where vernal pools exist from where they do not. In some of the rule-based scenarios, field investigation of the defining physical characteristics is the only way to determine if the rule sets are accurate in identifying the physical parameters at the pools.

Correlates of Vernal Pool Presence

Model creation provides the opportunity for describing the conditions under which the response variable is present. In this study, both the logistic regression and the CART models were in agreement that the green band of imagery, slope, and land use were important variables. Not surprisingly, the logistic regression determined that vernal pools are negatively correlated with green light reflectance and slope. Since water does not reflect green wavelengths of light, woodland seasonal ponds are more likely to be found on the imagery where there is little reflectance. Also, in order for water to pool, there must be little or no slope. So, as slope increases, the likelihood of finding a vernal pool decreases. Finally, vernal pools were positively associated with forested land and open land or fields. As expected, they were negatively correlated with development.

Unlike the logistic regression, the CART models were much more difficult to generalize, as there were no model fit or predictor strength statistics to help summarize the results of the model. The results of the CART analyses were the actual rules generated as trees. In a very general way, however, the variables near the top of the tree tend to be those that work over large geographic areas (i.e. climate). The top two levels of the classification trees in *this* study were green and red light reflectance variables, which were the biggest differentiators between vernal pool presence and absence. At intermediate levels in the CART analyses, aspect* appeared as a variable. Moving down through the tree structure, more site specific/local variables started to be incorporated, such as

110

land use and soil variables. These local variables were able to more finely depict vernal pool presence from absence and defined the final splits in the tree.

Overall, CART is somewhat difficult to generalize; however, the detail it provides makes it more useful to land managers and others who wish to know where vernal pools are likely to be found. Its ability to uncover conditional rules (rules that depend upon the outcome of other variables) and intricacies in the data make CART models more accurate and detailed; therefore, they are able to identify vernal pools under a variety of conditions. As evidenced from the CART analysis, all vernal pools do not occur under the same conditions. Models like logistic regression identify the overall trends in the data and predict a single scenario under which the response variable occurs. In this study, for instance, logistic regression detected three important variables that coalesced into a single statement to predict all vernal pools. In reality, this type of generalization is not possible. The strength and utility of the CART model, unlike logistic regression, is its ability to predict multiple scenarios under which the response variable occurs.

Model Performances and Utility

Logistic Regression Performance and Utility

The logistic regression performed much better statistically than it did cartographically. The overall model had a high R^2 value of 0.85. It was also successful at statistically predicting the validation data set. When using the 0.50 threshold, an accuracy of over 90% was attained; even when using a more stringent threshold for success (0.75), the overall accuracy in predicting the validation set was about 78%, which is adequate in most remote sensing

projects. This model was a significant improvement upon the foundation laid by Grant (2005). Higher accuracies in this study could be attributable to the use of high resolution bands of imagery and the use of other medium to large scale data layers (1:25,000 or larger) as model inputs.

Cartographically, the model did not perform as well. Presumably, the unsatisfactory performance of the equal-weighted models (conservative and liberal interpretations) was due to the difficulty in creating rules by subjective interpretation of the statistical results. Defining concrete rules based upon the odds ratios was not a straightforward process. The subjectivity in defining rules based on logistic regression results makes it difficult to use and to implement consistently and accurately. Explicitly defined rules, especially ones that account for the strength of the predictors, would be much more useful for a landscape scale predictive model of vernal pool locations.

Even with explicit rules, generated from the inverse logistic transformation, the model still did not perform as expected from the statistical results. At the 50% threshold, the model *statistically* predicted 90% of the vernal pools in the validation set. When mapped and evaluated, also at the 50% threshold, the map was able to predict only 54% of the validation pools, and overall, it performed so poorly that it was removed from the remainder of the analysis.

CART Performance and Utility

Statistically, CART4 and CART5 were able to predict an extremely high percentage of validation pools (95% and 93%, respectively). They also had high overall accuracies in predicting both vernal pools and absent points, 93% and

112

92%, respectively. These statistical results are similar to the logistic regression model results at the 0.5 threshold.

While statistically, the CART and logistic regression models performed similarly, overall, the CART routines were far superior to the logistic regression models in the later stages of analysis. Cartographically CART achieved high accuracies when predicting the validation set, with CART5 reaching 97%. These high validation accuracies, which exhibited few errors of omission (in this study, omission must be considered by how many validation pools were misclassified, rather than how many pools in the landscape were not identified by the model, as these data were not available), can likely be attributed to the clearly defined rules resulting from the Classification and Regression Tree. At each node, there was a rule that was directly queried in the GIS. This 1:1 correlation between the statistical output and the cartographic output eliminated the subjectivity involved in generating queries based upon logistic regression results. The direct, explicit rule set produced by the CART routine makes it much more understandable and much more easily incorporated into a GIS model.

The CART models, while extremely successful at predicting vernal pool locations, were not without limitations. Both CART4 and CART5 had very high commission error (low user's accuracy), meaning that non-vernal pools were falsely identified by the model. The vast majority of the confusion was between water and shadows, which, spectrally, appear similar. The fact that these errors were occurring in predictable ways makes them of lesser concern than if they were occurring at random. While high commission error is not desirable, in the

113

interest of conserving vernal pools, it is better to have high commission and low omission error, than to have lower commission error and overlook potentially critical vernal pools (Muñoz and Felicísimo 2004). With this type of conservation model, the goal is to err on the side of caution, rather than misidentify vernal pools.

Further, when viewed within the context of the purpose of modeling vernal pools, the commission error becomes even less problematic. These models should be regarded as tools for preliminary identification of vernal pools in order to facilitate and focus field or other investigations. With this goal in mind, a deeper examination of the severity of making commission errors is possible. Assuming that the results obtained from the validation subset areas (10% of total validation area, 891 ha) are applicable to the entire validation area (approximately 8,911 ha), the model output within the whole validation area can be evaluated. In the CART4 model, the total land area representing predicted vernal pools was 961 ha, which represented 10.8% of the total validation area. The land area eligible for vernal pool presence, according to this model, was reduced by 89.2%. Further, in keeping with the definition of vernal pools as isolated from other surface water, and for most conservation purposes, it is unnecessary/redundant to search places already identified by the National Wetlands Inventory, so removal of those areas from the analysis resulted in a drastic, 93.6% reduction in land area to search (574 ha). Based upon the estimated commission error, much of this land area is likely erroneously predicted to be vernal pools; however, the drastic reduction in searchable land

114

area is a valuable tool for those wishing to efficiently locate vernal pools. Within the 93.6% of the area dismissed as non-vernal pools, there are undoubtedly a few pools that have been omitted by the model (at least 12.7%, as determined by the omission of validation CVPs). Determination of the true omission error of this CART model would require a total enumeration of vernal pools within the study area, which was not possible to complete during this project.

CART5, which had higher commission error and lower omission error than CART4, also results in a dramatic decrease in the total searchable land area. In other words, it had higher accuracy when predicting the validation set of vernal pools, but it also called a higher percentage of areas "vernal pools" that were really other cover types. In this model, the total land area representing predicted vernal pools was 1535 ha, or about 17% of the total validation area. With areas classified as NWI wetlands removed from the analysis, the model area was reduced to 1,106 ha, or 12% of the total validation area. In this scenario, the land area was reduced by a total of 88%. Like the CART4 model, omission of pools within the 88% of the validation area predicted to be non-vernal pools was unavoidable (at least 2.9%, as determined by the omission of validation CVPs in this model); however, determining the exact percentage of omission error was unrealistic.

Again, while a large percentage of the total model output in both CART representations was likely incorrectly committed to the wrong category, the model output itself was only a small fraction of the total validation area. Since the vast majority of the commission errors in this project are attributable to shadows,

simple photo interpretation can eliminate obviously erroneous polygons to further reduce searchable land area.

Sources of Error

There are a number of errors associated with most GIS models. First, it is well known, but not well quantified, that there is a certain degree of error associated with every GIS layer. When several layers are combined, the error from each propagates through the analysis. In this study, seven data layers were used in the overlay analyses resulting from the CART models, including three layers of imagery, slope, aspect, land use, and soils. The logistic regression only used three variables: Band 2 of the imagery, land use, and slope.

There is also error associated with some of the other GIS layers utilized in this study. For example, the CVP layer, which was used as training and validation data, originally had an accuracy of +/- 100 meters (Szczebak, personal communication, June 5, 2006), due to the way that the layers were created. To improve this accuracy so that accurate information about these points could be collected, the points were manually corrected using photo interpretation an on-screen digitizing. Even though this was done very methodically and with rules governing when, how, and where to move the points, there is error inherently associated with this method. Without visiting the ground for each of those points (405), it is impossible to know with certainty if the points are representative of actual vernal pool locations. Even if they are, they may not be the pool intended by the person who certified it. Since the completion of this project, the methodology for spatially locating Certified Vernal Pools has changed; they are

116

now photo interpreted from high resolution imagery and digitized at approximately a 1:25,000 scale (National Heritage and Endangered Species Program 2007a).

Finally, the analysis was not divided based upon geographic location. One of the assumptions regarding this modeling effort was that vernal characteristics do not vary significantly over the geographic range of the study area. Since all four sets of training and validation data were in different locations throughout the northeastern part of the state, there may have been important characteristics related to each specific location that would have helped determine vernal pool location. To determine if this is true, "study site" would have to become a variable in each of the models. Geographic differences in vernal pool locations would be an interesting topic for further investigation.

Overall Model Improvements

The most important improvement to the CART models would be to establish a method of reducing errors of commission caused by shadows. An attempt was made to decrease shadow-water confusion by using remote sensing and statistics. Using the natural color 0.5 m ortho photos (RGB), the raw bands and all possible ratios between them, a spectral pattern analyses were created to try to distinguish water from shadow (Figure 27). Spectrally, with these bands and ratios, it was virtually impossible to tell the two apart; better spectral resolution may have helped with this problem. With little or no assistance from the imagery, a second attempt was made to differentiate shadows from water by including the ancillary data layers. In this trial, CART was employed with three

117

categories instead of only two: vernal pool, dry, and shadow. It, too, was unable to find any distinguishing characteristics between the two groups. At this point in the study, the problems with shadows were irresolvable with the available data and tools.
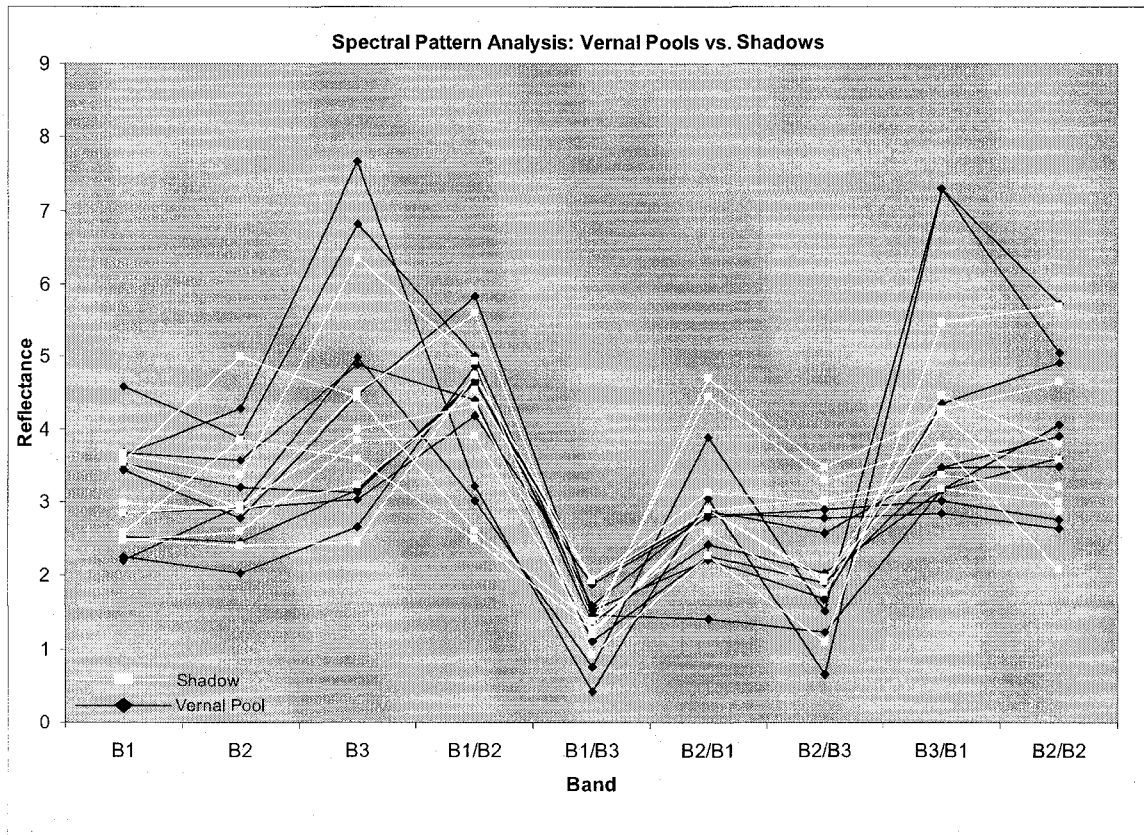
Figure 27: Spectral pattern analysis showing confusion between shadows and water.

Analyzing a time series of images that have different sun angles may be able to reduce shadow interference by changing the locations of shadows within the landscape, while the pools would remain constant. Also, greater spectral or radiometric resolution may provide new information or greater detail by which to tell the two categories apart. For instance, hyperspectral information (greater spectral resolution) may offer some separation between water and shadows by offering additional wavelengths for study. Further, increased radiometric

118

resolution (i.e. 11 bit imagery), which is able to detect more detailed reflectance information, may supply necessary distinction between the two spectrally similar objects. Another way of reducing the effect caused by shadows is to decrease the spatial resolution. The 0.5 m imagery helps to identify small objects in the landscape; however, such detail naturally makes shadows a problem. By slightly decreasing the spatial resolution to 1 m, 4 m, or even 10 m, the effects of shadows would be minimized. With decreasing resolution, however, more of the smaller vernal pools are at risk for omission, so this method would need extensive analysis to determine success or failure. This alternative deserves attention, as Sperduto and Congalton (1996) were able to successfully predict a rare orchid's habitat using 30 m Landsat imagery, illustrating that small patches of habitat can be predicted using models with coarse resolution.

In addition to improving commission error, a deeper investigation into omission error could be conducted to more fully evaluate the efficacy of using models such as the ones presented here. In this study, omission was defined by the number of validation CVP sites that were not identified by the models. True omission error is calculated by completing a total enumeration of vernal pools in the field, and then determining how many were missed by the model. As part of a study evaluating vernal pool identification using photography of different scales, Calhoun et al. (2003) estimated at least 27% omission error in mixed/deciduous forests (white pine, hemlock, red maple, red oak), much like the forests of northern Massachusetts. They determined that scale and forest cover type were two main limitations to identifying vernal pools using aerial photography. In this

119

study, the scale was fairly large (1:5,000), so the main limitation was likely forest cover type.

Since the study areas chosen in Massachusetts can be generally characterized as mixed/deciduous forest, one of the most unfortunate limitations to this model is that it does not have the ability to identify pools that are beneath a thick tree canopy. For this reason, spring-leaf off imagery is the most effective in this type of analysis. Of course, even with the optimal imagery, those pools beneath dense coniferous canopies are still undetectable, and, as previously asserted, are a significant source of omission errors. Synthetic Aperture RADAR (SAR) may be a plausible solution to this problem (Hess et al. 1990; MacDonald et al. 1981). Resulting from double-bounce reflections between surface water and tree trunks, flooded forest floors appear very bright on RADAR images. Forest structure (specifically basal area and height to the bottom of the canopy) has been shown to affect the accuracy of mapping below-canopy inundation with some types of RADAR, and would have to be considered in this project (Townsend 2002). Also, the output resolution would have to be a consideration, since vernal pools tend to be small water bodies. Overall, RADAR is a viable, explorable option for improved detection of vernal pools below the forest canopy.

Other improvements or adjustments could be made to the models that may decrease classification errors. For instance, in Grant (2005), the underlying geology was an important variable for predicting vernal pool locations. Underlying geology was not utilized in the present study because the scale (1:250,000) would have greatly limited the output resolution; however, at a finer scale, it could

120

have be an important predictor. Hydrologic parameters were also not considered as variable in this project. Seasonal or yearly average precipitation is one of the principal factors in determining the water balance and hydroperiod of vernal pools and may have had important predictive qualities useful for finding pools in the landscape, especially if reported at a local scale (Brooks 2004, Brooks and Hayashi 2002). Some other possible predictors could be proximity to perennial streams and depth to groundwater. While vernal pools, by definition, do not have permanent surface water connections, their interaction with groundwater accepted but not well-understood. Groundwater-surface water connections exist in some pools, and modeling that relationship may provide insight into where they occur (Brooks 2004, Brooks and Hayashi 2002, Hayashi and Rosenberry 2002).

Finally, while these models predicted seasonal forest pond locations, there was no provision for estimating the functional *value* of the pools. Evaluation of most wetland functions relies on field visits to assess such characteristics as water quality (temperature, dissolved oxygen, pH, etc.), hydroperiod, connectivity to other wetlands/pools, soil suitability, refugia, food sources, level of disturbance, canopy cover, vegetation abundance/richness, vegetation structure, condition of adjacent terrestrial habitat, presence and abundance of breeding amphibians, macroinvertebrate richness, etc. GIS is limited in its capability to remotely derive most of these data (Wolfson et al. 2002; Calhoun et al. 2003). Traits such as pool size (surface area, perimeter), connectivity (distance to other pools or wetlands), and density can be extracted from remote sources; however,

121

at the present time, there are no suitable data to serve as surrogates for the other, abovementioned, wetland/seasonal forest pond qualities related to function and value. When possible, many studies use GIS, though field visits are inevitable for identification of some key functional traits (Wolfson et al. 2002; Calhoun et al. 2003). Perhaps, with advances in the resolution of remote sensing products and in GIS data quality, the possibility of remotely determining wetland functions will be more fully realized in the future.

Overall, with more time, resources, and advanced technology, additional variables could be added to the models and potentially enhance the accuracy with which they identify vernal pools. The goal of the models (identification) could also be expanded to include evaluation of pool function, both on an individual basis and within a greater network of pools. The specific goal of *this* study, which was to identify vernal pools using statistics, GIS, and remote sensing with pool conservation in mind (limiting omission errors), was achieved with better-than-expected results.

Conclusion

The results of this study indicate that there is a correlation between vernal pools and the physical characteristics that are present at vernal pool locations: slope, aspect, land use, soil type, and spectral reflectance were investigated. The relationship between these variables and the presence of vernal pools was determined by the use of statistics, Geographic Information Systems, and remote sensing. By combining the power of statistical modeling with the utility of

122

cartographic modeling, a highly accurate representation of vernal pool locations was produced.

Many environmental studies are utilizing ecological modeling to predict spatial phenomenon; however, some do not make the leap from statistical predictions to spatial ones. To terminate a spatial project at the point where only statistical results have been achieved is to leave the project unfinished. The importance of following through and determining if/how the statistical results/accuracies translate into spatial ones should not be overlooked. For instance, in this study, the logistic regression routine produced a strictly statistical accuracy of about 86% (correct predictions of validation pools). Had the modeling process ended at this step, the model would have been considered extremely successful; however, the cartographic model was only about 33% accurate in predicting the validation pools. Had this vital second step been excluded from the study, valuable time and resources may have been spent trying to implement this model in a real world application. In the end, the model was discarded because it did not perform as well as initially expected considering the statistical results.

With the highlighted importance of translating statistics into some usable product (i.e. cartographic representation), choosing a modeling technique that has an output that is easily converted into a spatial model is critical. In this study, two modeling techniques were tested. As already discussed, the logistic regression performed well statistically, but did not produce an intuitive set of rules that could be easily converted into GIS queries; therefore, its spatial model accuracy was much lower than expected. Likely, this low accuracy was a result

123

of the subjective interpretation and creation of GIS queries from inexplicit results. The CART models, however, were much more conducive to cartographic modeling. By their very nature, they produced a specific rule set that was directly queried in a GIS; therefore, there was no subjective interpretation of the results for the spatial model because the statistical model specifically defined the rules for the spatial one. The high accuracies of the CART models reflect their ease of translation. Overall, the cartographic outputs of the CART models had the highest accuracies both statistically and cartographically, and have the potential to be used in similar geographic areas for the detection of vernal pools. CART5 had the lowest omission error and is therefore most appropriate for conservation purposes.

LITERATURE CITED

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model Fitting. *Biometrica*, 66(2): 237-242.

Andersen, M. C., Watts, J. M., Freilich, J. E., Yool, S. R., Wakefield, G. I., McCauley, J. E., and Fahnestock, P. B. (2000). Regression-tree modeling of desert tortoise habitat in the central Mojave Desert. *Ecological Applications*, 10(3): 890-900.

Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E. (1976). Land use and land cover classification system for use with remote sensor data. Geological Survey Professional Paper 964. Washington D.C.

Babbitt, K. J. (2005). The relative importance of wetland size and hydroperiod for amphibians in southern New Hampshire, USA. *Wetlands Ecology and Management*, 13: 269-279.

Babbitt, K. J., Baber, M. J., and Tarr, T. L. (2003). Patterns of larval amphibian distribution along a wetland hydroperiod gradient. *Canadian Journal of Zoology*, 81: 1539-1552.

Babbitt, K. J., and Tanner, G. W. (2000). Use of temporary wetlands by anurans in a hydrologically modified landscape. *Wetlands*, 20(2): 313-322.

Bian, L. and West, E. (1997). GIS modeling of elk calving habitat in a prairie environment with statistics. *Photogrammetric Engineering and Remote Sensing*, 63(2): 161-167.

Bonney, G. E. (1987). Logistic regression for dependent binary observations. *Biotmetrics*, 43: 951-973.

Brooks, R. T. (2004). Weather-related effects on woodland vernal pool hydrology and hydroperiod. *Wetlands*, 24(1): 104-114.

Brooks, R. T. and Hayashi, M. (2002). Depth-area-volume and hydroperiod relationships of ephemeral (vernal) forest pools in southern New England. *Wetlands*, 22(2): 247-255.

Brooks, R. T., Stone, J., and Lyons, P. (1998). An inventory of seasonal forest ponds on the Quabbin Reservoir watershed, Massachusetts. *Northeastern Naturalist*, 5(3): 219-230.

Burne, M. R. (2001). *Massachusetts aerial photo survey of Potential Vernal Pools*. NHESP: Westborough, MA.

Burne, M. R. and Griffin, C. R. (2005). Protecting vernal pools: a model from Massachusetts, USA. *Wetlands Ecology and Management*, 13: 367-375.

Calhoun, A. J. K., Walls, T. E., Stockwell, S. S., and McCollough, M. (2003). Evaluating vernal pools as a basis for conservation strategies: A Maine case study. *Wetlands*, 23(1): 70-81.

Carroll, C., Zielinski, W. J., and Noss, R. F. (1999). Using presense-absence data to build and test spatial habitat models for the fisher in the Klamath Region, U.S.A. *Conservation Biology*, 13(6): 1344-1359.

Clark Labs (2007). *Statistics*. Retrieved September 6, 2007 from: http://www.clarklabs.org/products/statistics.cfr.

Colburn, E. A. (2004). *Vernal pools: Natural history and conservation*. Blacksburg, VA: McDonald and Woodward Publishing Co.

Commonwealth of Massachusetts Division of Fisheries and Wildlife. (2001). *Guidelines for the Certification of vernal pool habitat*. Westborough, MA: National Heritage and Endangered Species Program.

Compton, B. W., Rhymer, J. M., and McCollough, M. (2002). Habitat selection by wood turtles (*Clemmys insculpta*): An application of paired logistic regression. *Ecology*, 83(3): 833-843.

Congalton, R. G., and Green, K (1999). *Assessing the accuracy of remotely sensed data: Principles and practices*. New York: Lewis Publishers.

Cowardin, L. M., Carter, V., Golet, F. C., and LaRoe, E. T. (1979). Classification of wetlands and deepwater habitats of the United States. U.S. Fish and Wildlife Service Report No. FWS/OBS/-79/31. Washington, D.C.

De'ath, G. and Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, 81(11): 3178-3192.

Downing, D. M., Winer, C., and Wood, L. D. (2003). Navigating through the Clean Water Act jurisdiction: A legal review. *Wetlands*, 23(3): 475-493.

Gibbons, J. W. (2003). Terrestrial habitat: A vital component for herpetofauna of isolated wetlands. *Wetlands*, 23(3): 630-635.

Gibbs, J. P. (1998). Distribution of woodland amphibians along a forest fragmentation gradient. *Landscape Ecology*, 13: 263-268.

Gibbs, J. P. (2000). Wetland loss and biodiversity conservation. *Conservation Biology*, 14(1): 314- 317.

Gibson, L. A., Wilson, B. A., Cahill, D. M., and Hill, J. (2004). Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. *Journal of Applied Ecology*, 41: 213-223.

Godwin, K. S., Shallenberger, J. P., Leopold, D. J., and Bedford, B. L. (2002). Linking landscape properties to local and hydrogeologic gradients and plant species occurrence in minerotrophic fens of New York state, USA: A hydrogeologic setting (HGS) framework. *Wetlands*, 22(4): 722-737.

Grant, E. H. C. (2005). Correlates of vernal pool occurrence in the Massachusetts, USA landscape. *Wetlands*, 25(2): 480-487.

Guisan, A., Edwards, T. C., and Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modeling*, 157: 89-100.

Guisan, A. and Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modeling*, 135: 147-186.

Hayashi, M. and Rosenberry, D. O. (2002). Effects of groundwater exchange on the hydrology and ecology of surface water. *Groundwater*, 40(3): 309-316.

Herrmann, H. L., Babbitt, K. J., Baber, M. J, and Congalton, R. G. (2005). Effects of landscape characteristics on amphibian distribution in a forest-dominated landscape. *Biological Conservation*, 123: 139-149.

Hess, L. L., Melack, J. M., and David, S. S. (1990). RADAR detection of flooding beneath the forest canopy: a review. *International Journal of Remote Sensing*, 11(7): 1313-1325.

Hirzel, A. H., Helfer, V., and Metral, F. (2001). Assessing habitat-suitability models with a virtual species. *Ecological Modeling*, 145: 111-121.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wiley.

Iverson, L. R. and Prasad, A. M. (1998). Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs*, 68(4): 465-485.

Jenness, J. (2004). Convex hulls around points (conv_hulls_pts.avx) extension for ArcView 3.x, v. 1.21. Jenness Enterprises. Available at http://www.jennessent.com/acrview/convex_hulls.htm.

Keeley, J. E., and Zedler, P. H. (1998). Characterization and global distribution of vernalpools. Pages 1-14 *in:* Witham, C. W., Bauder, E. T., Belk D., Ferren Jr., W. R., and Ornduff, R. (Eds). Ecology, Conservation, and Management of Vernal Pool Ecosystems - Proceedings from a 1996 conference. California Native Plant Society: Sacramento, CA.

Lathrop, R. G., Montesano, P., Tesauro, J., and Zarate, B. (2005). Statewide mapping and assessment of vernal pools: A New Jersey case study. *Journal of Environmental Management*, 76: 230-238.

Lawrence, R. L. and Wright, A. (2001). Rule-based classification systems using Classification and Regression Tree (CART) analysis. *Photogrammetric Engineering and Remote Sensing*, 67(10): 1137-1142.

Lehmann, A. (1998). GIS modeling of submerged macrophyte distribution using Generalized Additive Models. *Plant Ecology*, 139: 113-124.

Leibowitz, S. G. (2003). Isolated wetlands and their functions: An ecological perspective. *Wetlands*, 23(3): 517-531.

Leica Geosystems. (2006). Image Analysis extension for ArcGIS 9.1. Available at http://gi.leica-geosystems.com/LGISub2x236x127.aspx.

Lek, S. and Guégan, J. F. (1999). Artificial neural networks as a tool in ecological modeling, an introduction. *Ecological Modeling*, 120: 65-73.

Lichko, L. E. and Calhoun, A. J. K. (2003). An evaluation of vernal pool creation projects in New England: Project documentation from 1991-2000. *Environmental Management*, 32(1): 141-151.

"Logistic Regression" (2007). Retrieved June 4, 2007 from http://www2.chass.ncsu.edu/garson/pa765/logistic.htm.

MacDonald, H. C., Waite, W. P., and DeMarcke, J. S. (1981). Use of Seasat satellite RADAR imagery for the detection of standing water beneath forest vegetation. *Rainbow 80; Fall Technical Meeting, ASP Technical Papers*, R-S-3-B-1 - R-S-3-B-13.

Mace, R. D., Waller, J. S., Manley, T. L., Ake, K., and Wittinger, W. T. (1999). Landscape evaluation of grizzly bear habitat in western Montana. *Conservation Biology*, 13(2): 367-377.

Marsh, D. M. and Trenham, P. C. (2001). Metapopulation dynamics and amphibian conservation. *Conservation Biology*, 15(1): 40-49.

MassGIS (2001). *Color Ortho Imagery*. Retrieved January 4, 2006 from http://www.mass.gov/mgis/colororthos2001.htm.

MassGIS (2002a). *Community Boundaries (Towns)*. Retrieved January 9, 2006 from http://www.mass.gov/mgis/towns.htm.

MassGIS. (2002b). *Land Use*. Retrieved January 9, 2006 from http://www.mass.gov/mgis/lus.htm.

MassGIS. (2003). *Digital Terrain Model (DTM) files*. Retrieved September 19, 2005 from http://www.mass.gov/mgis/dtm.htm.

MassGIS. (2005a). *Digital Elevation Model*. Retrieved September 19, 2005 from http://www.mass.gov/mgis/img_elev5k.htm.

MassGIS (2005b). Soils. Retrieved January 9, 2006 from http://www.mass.gov/mgis/soi.htm.

Mladenoff, D. J., Sickley, T. A., Haight, R. G., and Wydeven, A. P. (1995). A regional landscape analysis and prediction of favorable gray wolf habitat in the northern Great Lakes Region. *Conservation Biology*, 9(2):279-294.

Muñoz, J. and Felicísimo, A. M. (2004). Comparison of statistical methods commonly used in predictive modeling. *Journal of Vegetation Science*, 15: 285-292.

National Heritage and Endangered Species Program (2000). *Potential Vernal Pools data layer*. Retrieved September 19, 2005, from http://www.mass.gov/mgis /pvp.htm.

National Heritage and Endangered Species Program (2002). *Certified Vernal Pools data layer*. Retrieved September 19, 2005, from http://www.mass.gov/mgis/cvp.htm.

National Heritage and Endangered Species Program (2007a). *Certified Vernal Pools data Layer*. Retrieved September 28, 2007 from http://www.mass.gov/mgis/cvp.htm.

Natural Heritage and Endangered Species Program (2007b). Massachusetts Forestry Conservation Management Practices for MESA-listed mole salamanders. Version 2007.1. Natural Heritage and Endangered Species Program, Massachusetts Division of Fisheries and Wildlife, Westborough, Massachusetts, USA.

National Oceanic and Atmospheric Administration; National Climatic Data Center (2005). *Climate of Massachusetts.* Retrieved June 1, 2007 from http://cdo.ncdc.noaa.gov/cgi-bin/climatenormals/climatenormals.pl?directive=prod_select2andprodtype=CLIM60andsubrnum=.

North, M. P., Franklin, J. F., Carey, A. B., Forsman, E. D., and Hamer, T. (1999). Forest stand structure of the northern spotted owl's foraging habitat. *Forest Science*, 45(4): 520-527.

Palik, B., Batzer, D. P., Buech, R., Nichols, D., Cease, K., Egeland, L., and Streblow, D. E. (2001). Seasonal pond characteristics across a chronosequence of adjacent forest ages in northern Minnesota, USA. *Wetlands*, 21(4): 532-542.

Preisser, E. L., Kefer, J. Y., and Lawrence, J. D. (2000). Vernal pool conservation in Connecticut: An assessment and recommendations. *Environmental management*, 26(5): 503-513.

Rittenhouse, T. A. G. and Semlitsch, R. D. (2006). Grasslands as movement barriers for a forest-associated salamander: Migration behavior of adult and juvenile salamanders at a distinct habitat edge. *Biological Conservation*, 131: 14-22.

Rothermel, B. B. and Semlitsch, R. D. (2002). An experimental investigation of landscape resistance of forest versus old-field habitats to emigrating juvenile amphibians. *Conservation Biology*, 16(5): 1324-1332.

Rothermel, B. B. and Semlitsch, R. D. (2006). Consequences of forest fragmentation for juvenile survival in spotted (*Ambystoma maculatum*) and marbled (*Ambystoma opacum*) salamanders. *Canadian Journal of Zoology*, 84: 797-807.

"S+ SpatialStats Product Features." Retrieved September 27, 2007 from http://www.insightful.com/products/spatial/features.asp

Semlitsch, R. D. (1998). Biological delineation of terrestrial buffer zones for pond-breeding salamanders. *Conservation Biology*, 12(5): 1113-1119.

Semlitsch, R. D. and Bodie, J. R. (1998). Are small, isolated wetlands expendable? *Conservation Biology*, 12(5): 1129-1133.

Semlitsch, R. D. and Bodie, J. R. (2003). Biological criteria for buffer zones around wetlands and riparian habitats for amphibians and reptiles. *Conservation Biology*, 17(5): 1219-1228.

Semlitsch, R. D., Ryan, R. J., Hamed, K., Chatfield, M. Drehman, B., Pekarek, N., Spath, M., and Watland, A. (2007). Salamander abundance along road edges an within abandoned logging roads in Appalachian forests. *Conservation Biology*, 21(1): 159-167.

Skidds, D. E., and Golet, F. C. (2005). Estimating hydroperiod suitability for breeding amphibians in southern Rhode Island seasonal forest ponds. *Wetlands Ecology and Management*, 13: 349-366.

Skidmore, A. K., Gauld, A., and Walker, P. (1996). Classification of kangaroo habitat distribution using three GIS models. *International Journal of Geographical Information Systems*, 10(4): 441-454.

Smith, M. A. and Green, D. M. (2005). Dispersal and the metapopulation paradigm in amphibian ecology and conservation: are all amphibian populations metapopulations? *Ecography,* 28: 100-128.

Smith, D. W. and Verrill, W. L. (1998). Vernal pool-soil-landform relationships in the Central Valley, California. Pages 15-23 *in:* Witham, C. W., Bauder, E. T., Belk D., Ferren Jr., W. R., and Ornduff, R. (Eds). Ecology, Conservation, and Management of Vernal Pool Ecosystems - Proceedings from a 1996 conference. California Native Plant Society: Sacramento, CA.

Sperduto, M. B. and Congalton, R. G. (1996). Predicting rare orchid (small whorled pogonia) habitat using GIS. *Photogrammetric Engineering and Remote Sensing*, 62(11): 1269-1279.

Thurston, J. (2002). GIS and artificial neural networks: Does your GIS think? Retrieved March 29, 2006 from http://www10.giscafe.com/link/display_detail.php?link_id=10124

Tiner, R. W. (2003). Estimated extent of geographically isolated wetlands in selected areas of the United States. *Wetlands,* 23(3): 636-652.

Townsend, P. A. (2002). Relationships between forest structure and the detection of flood inundation in forested wetlands using C-band SAR. *International Journal of Remote Sensing.* 23(3): 443-460.

United States Army Corps of Engineers. (1987). *Corps of engineers wetlands delineation manual* (Technical report Y-87-1). Vicksburg, MS

United States Census Bureau (2006). *State and County Quick Facts*. Retrieved June 1, 2007 from http://quickfacts.census.gov/qfd/states/25000.html

Vasconcelos, D. and Calhoun, A. J. K. (2004). Movement patterns of adult and juvenile *Rana sylvatica* (LeConte) and *Ambystoma maculatum* (Shaw) in three restored seasonal pools in Maine. *Journal of Herpetology*, 38(4): 551-561.

Vogiatzakis, I. N. (2003). GIS-based modeling and ecology: A review of tools and methods. Geographical Paper No. 170.

Winter, T. C. and Rosenberry, D. O. (1995). The interaction of groundwater with prairie pothole wetlands in the Cottonwood Lake area, east-central North Dakota, 1979-1990. *Wetlands*, 15(3): 193-211.

Wolfson, L., Mokma, D., Schultink, G., and Dersch, E. (2002). Development and use of a wetlands information system for assessing wetland functions. *Lakes & Reservoirs: Research and Management*, 7: 207-216.

Zedler, P. H. (2003). Vernal pools and the concept of "isolated wetlands." *Wetlands*, 23(3): 597-607.

Zhou, W. (1999). Verification of the nonparametric characteristics of backpropagation neural networks for image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2): 771-779.

APPENDIX

CLASSIFICATION AND REGRESSION TREES
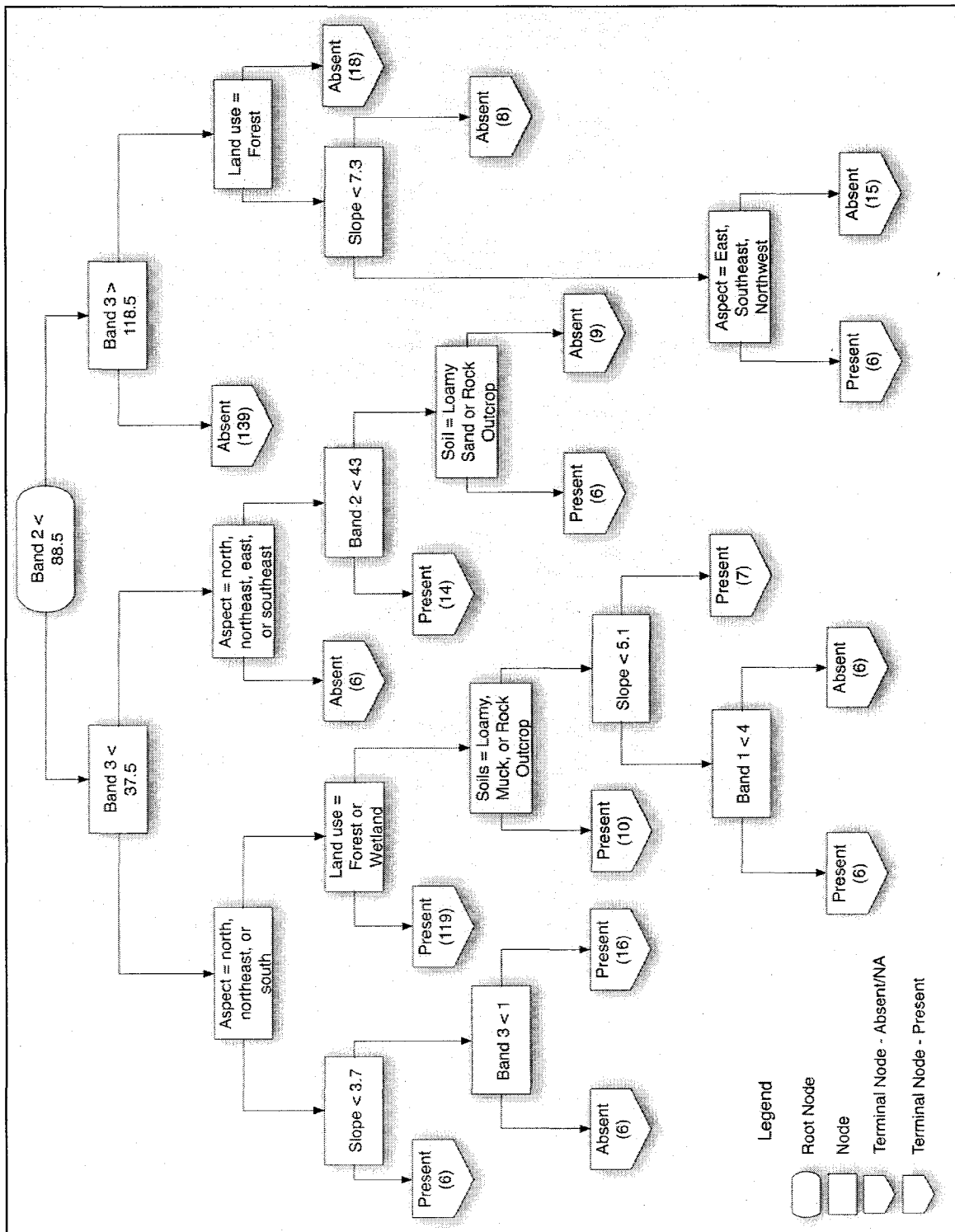
Figure A - 1: Representation of CART4 model. The tree is read by beginning at the root node and extending through each decision point until a terminal node is reached. If the condition presented at an individual node is correct for a given point, the tree proceeds to the left. Conversely, if the condition at a node is incorrect, the tree proceeds to the right. At each terminal node, the number in parenthesis represents the number of points predicted by that particular rule set.
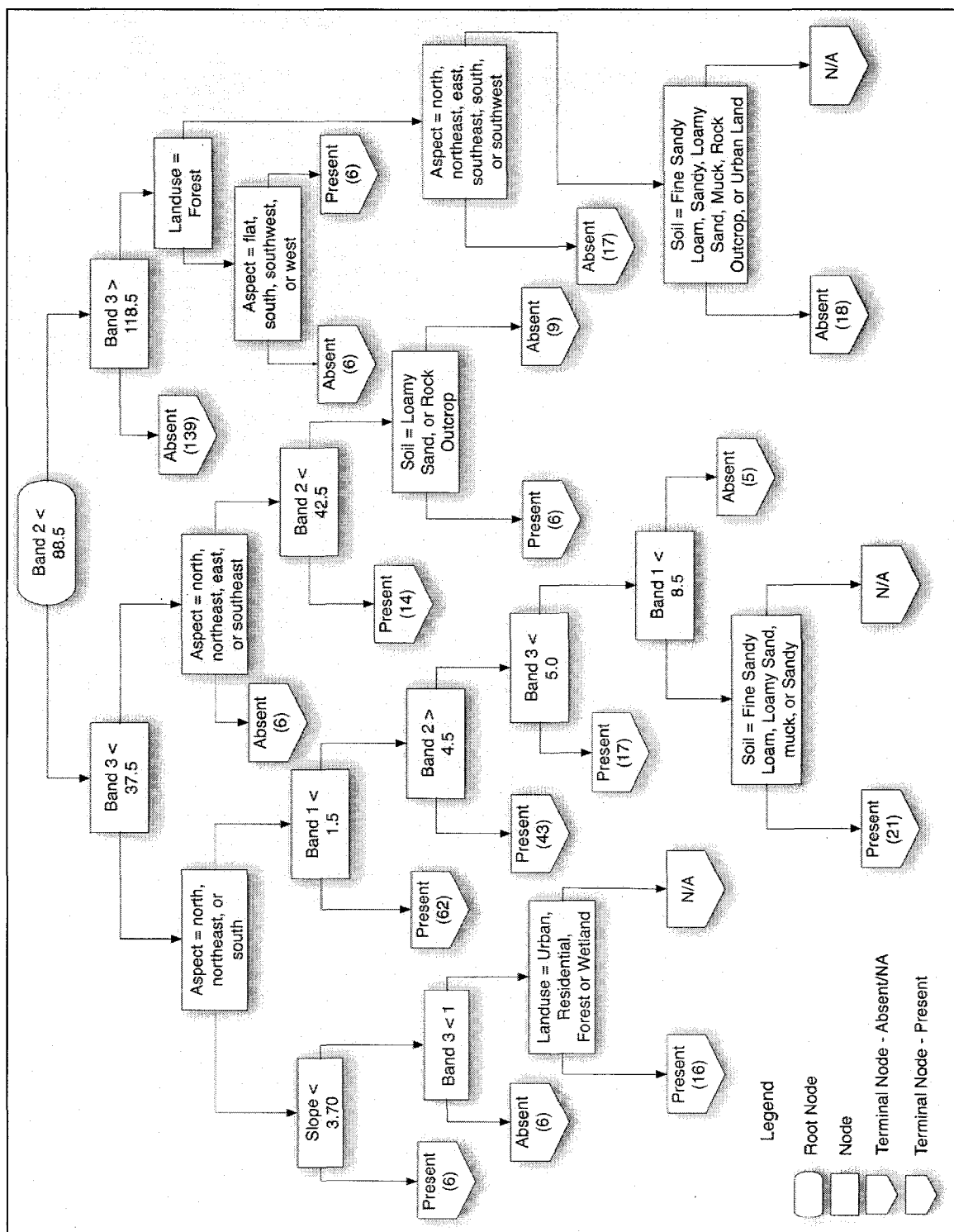
134

Figure A - 2: Representation of CART5 model. The tree is read by beginning at the root node and extending through each decision point until a terminal node is reached. If the condition presented at an individual node is correct for a given point, the tree proceeds to the left. Conversely, if the condition at a node is incorrect, the tree proceeds to the right. At each terminal node, the number in parenthesis represents the number of points predicted by that particular rule set.

135