PENNSYLVANIA DEPARTMENT OF TRANSPORTATION
ARCHAEOLOGICAL PREDICTIVE MODEL SET

TASK 6: Study Regions 7, 8, 9, and 10

CONTRACT #355I01

ARCHAEOLOGICAL PREDICTIVE MODEL SET

Category #05 – Environmental Research

December 2014

URS

PENNSYLVANIA DEPARTMENT OF TRANSPORTATION ARCHAEOLOGICAL PREDICTIVE MODEL SET TASK 6: STUDY REGIONS 7, 8, 9, AND 10

CONTRACT #355I01

Prepared for

Pennsylvania Department of Transportation
Bureau of Planning and Research
Keystone Building
400 North Street, 6th Floor, J-East
Harrisburg, PA 17120-0064

Prepared by

Matthew D. Harris, Principal Investigator
Susan Landis
and
Andrew R. Sewell, Hardlines Design Company

URS Corporation
437 High Street
Burlington, NJ 08016-4514

December 2014

ABSTRACT

This report is the documentation for Task 6 of the Statewide Archaeological Predictive Model Set project sponsored by the Pennsylvania Department of Transportation (PennDOT). This project was solicited under Contract #355I01, Transportation Research, Education, and Technology Transfer ITQ, Category #05 – Environmental Research. The goal of this project is to develop a set of statewide predictive models to assist the planning of transportation projects. PennDOT is developing tools to streamline individual projects and facilitate Linking Planning and NEPA, a federal initiative requiring that NEPA activities be integrated into the planning phases for transportation projects. The purpose of Linking Planning and NEPA is to enhance the ability of planners to predict project schedules and budgets by providing better environmental and cultural resources data and analyses. To that end, PennDOT is sponsoring research to develop a statewide set of predictive models for archaeological resources to help project planners more accurately estimate the need for archaeological studies.

The objective of Task 6, discussed in the following report, is to create a series of archaeological predictive models for Regions 7, 8, 9, and 10. In total, this area covers 13,701 square miles, which is 30% of the state. These three regions cover much of eastern Pennsylvania, including the Ridge and Valley Province, New England Province, Piedmont Province, Atlantic Coastal Plain Province, and part of the Appalachian Plateau Province. A total of 7,297 prehistoric archaeological components were incorporated into this modeling effort. Two hundred and sixty-four individual candidate models were created to cover these four regions. The final ensemble is created from 66 models selected for their representation of the archaeological sensitivity of each of the subareas. This final model correctly classifies 98.5% of known site-present cells within 26.8% of the study area, for a Kg of 0.726 and an average hold-out sample prediction error of RMSE = 0.122.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1
INTRODUCTION

The purpose of this project is to use the existing Pennsylvania Archaeological Site Survey (PASS) file database to produce a baseline model for the sensitivity of prehistoric site-presence throughout the entire Commonwealth using Archaeological Predictive Modeling (APM). The resulting assessments of archaeological sensitivity will be used by transportation, planning, and other Cultural Resources Management (CRM) practitioners to make better-informed and more consistent assessments of prehistoric archaeological sensitivity, with the ultimate goal of saving time, money, and sparing cultural resources.

Building from the previous tasks in this project—a review of APM literature (Harris 2013a), designation of study regions (Harris 2013b), the creation of a pilot model for central Pennsylvania (Harris 2014), and modeling six regions in western and central Pennsylvania (Harris et al. 2014a, 2014b), this report documents the final in a series of three tasks that apply the modeling methodology to the entire state. This report details the creation, findings, and conclusions of predictive models created for Regions 7, 8, 9, and 10 (Figure 1). These regions comprise a total of 13,701 square miles, 30% of the entire state. Covering almost the entirety of eastern Pennsylvania, this process involved creating 66 individual models from a dataset of over 7,000 prehistoric archaeological sites.

The process reported below consisted of the development of proportionally weighted models and three statistical models (Logistic Regression [LR], Multivariate Adaptive Regression Splines [MARS], and Random Forest [RF]) for each of 66 subareas. Each of these model types is discussed and detailed in the previous Task 3 report (Harris et al. 2014a). The final model selected to represent the Regions 7, 8, 9, and 10 is a composite of each of the three different statistical model types: one LR model, 11 MARS models, and 54 RF models. The selection of a model for each subarea was based the quality, quantity, and representativeness of the known data, the model metrics and error rates, and the distribution of site-present cells versus background cells summed up by the Kvamme Gain (Kg) statistic (Kvamme 1988). The end result of this process is the classification of a high, moderate, and low sensitivity model that covers the entirety of each of the four regions. The report below documents the model building process, as well as the breadth of previous modeling attempts in the regions, the prehistoric context of the area, an assessment of PASS data quality, and special topics of concern for the modeling process.

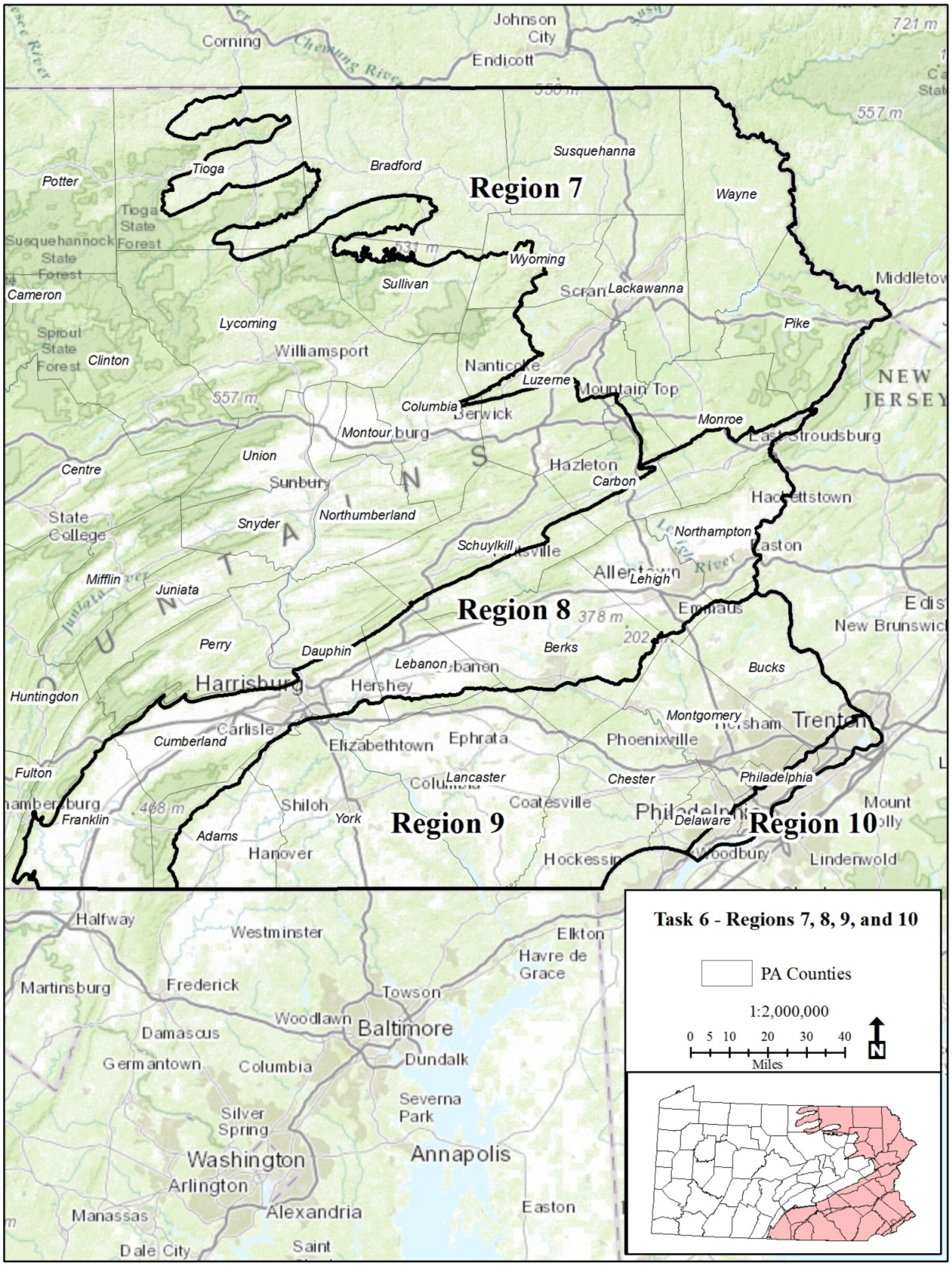


Figure 1 - Overview of Regions 7, 8, 9, and 10

PREDICTIVE MODELING IN REGIONS 7, 8, 9, AND 10

Only a few predictive models were located for Regions 7, 8, 9, and 10, and most were associated with compliance-related projects. Because of this association, the models often focused on an area determined by the location of the specific project rather than being generated to answer questions about settlement patterns. Some of these models simply used environmental factors combined with analysis of site locations from PASS data to generate areas of high and low probability for locating a prehistoric site in general, without attempting to refine the model to predict site types or cultural affiliation (Hay 1993; Mooney et al. 2003; Pan Cultural Associates 2005; Miller and Kodlick 2006; Petyk et al. 2010). The more useful studies found for this report are summarized briefly below.

In 1980 a team from the State University of New York-Binghamton (SUNY-Binghamton) conducted a survey of the archaeological and historical resources of the Delaware National Historic and Scenic River for the National Park Service (Dekin 1980). While most of the area studied by the SUNY-Binghamton team was in New York, the Pennsylvania counties of Pike and Wayne were also included (located in Region 7). The survey applied a model that essentially depicts the study area as a surface showing peaks where various factors overlap to show the likeliest locations for prehistoric occupations to be found. With this model, the landscape is divided into hexagons covering 214 acres each and ranked using scored variables based on factors including stream rank, confluence, slope, and various physiographic landforms (Dekin 1980). The highest scoring hexagons would be considered the most probable areas to contain prehistoric sites. Scores above 13 were considered to represent very high probability for site locations. The system was developed for areas with major floodplains, and may not be applicable to upland locations. Additionally, the model did not attempt to predict sites by type or by time period.

Another early attempt at characterizing site distribution in eastern Pennsylvania was conducted by Edward Wilson and W. Fred Kinsey in 1981 and 1982, when they surveyed a region in the Great Valley region of the Ridge and Valley physiographic province, between the Schuylkill and Lehigh drainage systems (Wilson and Kinsey 1982). Their study area was a hilly, upland region, removed from major river systems. The results of this study suggest proximity to water was the most important factor in determining site locations in the study area, with large sites most commonly found within 500 feet of a water source. Site locations did not seem to be associated with stream order. Sites were more common on floodplains and in upland depressions. Locations with southern exposures and less than 8% slopes were preferred for site locations as well. The types of sites that are probably represented by the survey results are transient resource-procurement camps associated with groups based in the Schuylkill and Lehigh river valleys that bracket the study area. The cultural periods represented in their survey were largely Late Archaic, with few instances of Woodland occupation.

One model constructed by Hunter and Burrows (1990) for a modification project at the F. E. Walter Dam in Luzerne and Monroe Counties was particularly well-constructed, although limited to a specific upland setting. In fact, Affleck et al. (1994) adapted the model constructed by Hunter and Burrows (1990) to their site area because of similarities between the two project areas. Hunter and Burrow's model was built on Binford's (1980) ideas on foraging and collecting societies and their site types. The authors noted that factors influencing site location in northeast Pennsylvania included:

- Distance to water
- Distribution of well-drained, low-relief areas
- Distribution of game habitat
- Distribution of zones of maximum habitat overlap
- Distribution of high-order streams
- Distribution of lithic sources
- Distribution of areas with maximum sunlight exposure
- Distribution of cultivatable land

They also referenced Stewart and Kratzer's 1989 work on the Unglaciated Allegheny Plateau in Pennsylvania, which identified the following settings as having probability for archaeological sites:

- Broad upland flats overlooking stream valleys
- Saddles between drainage divides and upland flats
- Locations at heads of active drainages
- Locations near heads of inactive drainages
- Upland flats adjacent to first-order streams and proximal to stream confluences

The authors also noted that areas with broad floodplains had greater potential for high densities of sites, as they would contain a greater diversity of habitats as well as landforms suitable for occupation.

Using the above factors and applying them to the F.E. Walter Dam project area, the authors found that there were few areas of high probability for archaeological sites (Hunter and Burrows 1990:5–8). They observed that the floodplain in the project area was narrow to intermediate in width. The lack of chert-bearing limestone in their area meant that prehistoric occupants of the area would have had to travel outside the immediate region to obtain high-quality lithic material. Finally, they noted that soils in the floodplains and terraces were not characterized as high fertility. These environmental factors led the authors to predict the project area would have a low probability for large base camps and Late Woodland villages, although the area would likely have been used during short-duration resource procurement forays by groups based outside the project area.

Botwick and Wall (1994) conducted a set of surveys in the uplands of the Delaware Water Gap area of Pennsylvania and New Jersey. As part of their work, they developed a predictive model for uplands that was based on previously documented site locations in relation to landforms in the region, and also on analysis of relict hydrological systems. The authors identified landforms that were likely to contain prehistoric archaeological sites, such as rock outcrops that provided either lithic raw material or shelter, ridgetops and other similar overlooks, edges of wetlands, and stream terraces, among others. The authors did not attempt to refine the model beyond location to address site types or temporal periods. Using this model in subsequent surveys, they found that within the uplands of the Delaware Water Gap, most of the sites they identified were along stream terraces or otherwise close to water, while areas located farther from water or along low-order streams tended to produce far fewer sites.

2
STUDY AREA – REGIONS 7, 8, 9, AND 10

PHYSICAL CHARACTER

Regions 7, 8, 9, and 10 occupy the easternmost part of the state and cover wide-ranging physiographic settings. Portions of Regions 7 and 8 are located within the Ridge and Valley physiographic province, which is characterized by long, even ridges punctuated by long valleys that run in a southwesterly to northeasterly direction through the central and eastern portions of the state. One section of the Ridge and Valley province falls within Region 7 (Anthracite Valley), while three sections are within Region 8 (Blue Mountain, South Mountain, and Great Valley). The remainder of Region 7 is located within the Appalachian Plateaus physiographic province. Two sections of the Appalachian Plateaus fall within Region 7 (Glaciated Low Plateau and Glaciated Pocono Plateau). A small portion of Region 8 is located within the New England physiographic province, in the Reading Prong section. Region 9 is located entirely within the Piedmont physiographic province, in three sections (Gettysburg-Newark Lowland, Piedmont Lowland, and Piedmont Upland). Region 10 is also contained within one physiographic province (Atlantic Coastal Plain) and just one section (Lowland and Intermediate Upland) (Table 1; Figure 2)

Table 1 - Physiographic Provinces and Sections for Modeling Regions 4, 5, and 6

Modeling Region	Physiographic Province	Physiographic Section
7	Appalachian Plateaus	Glaciated Low Plateau
		Glaciated Pocono Plateau
	Ridge and Valley	Anthracite Valley
8	Ridge and Valley	Blue Mountain
		South Mountain
		Great Valley
	New England	Reading Prong
9	Piedmont	Gettysburg-Newark Lowland
		Piedmont Lowland
		Piedmont Upland
10	Atlantic Coastal Plain	Lowland and Intermediate Upland

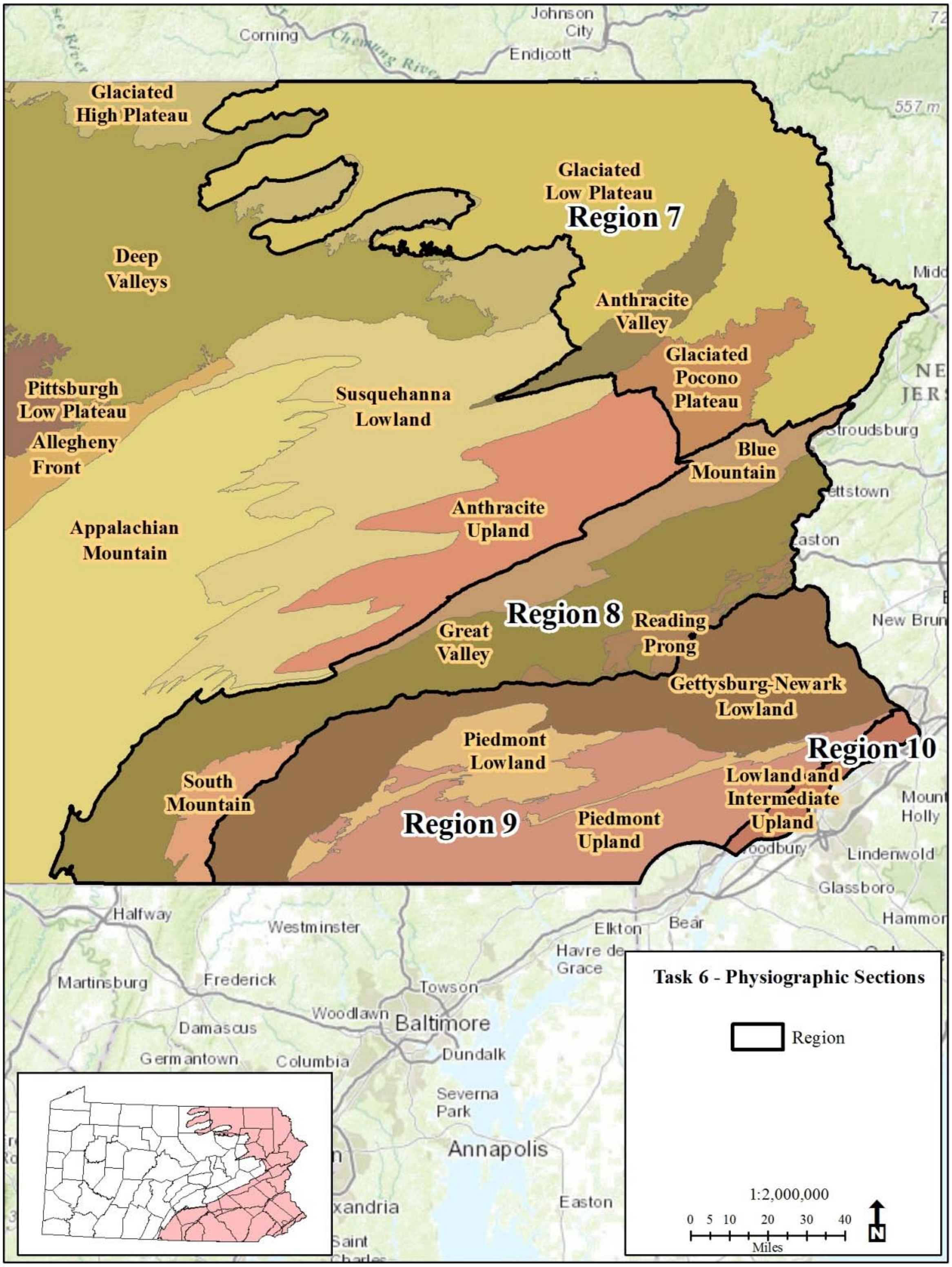


Figure 2 - Regions 7, 8, 9, and 10 physiographic sections.

Appalachian Plateaus

Glaciated Low Plateau Section

The Glaciated Low Plateau section is located in the northeast of Pennsylvania along the New York and New Jersey state borders. The section abuts six sections: Glaciated High Plateau, Deep Valleys, Susquehanna Lowland, Anthracite Valley, Glaciated Pocono Plateau, and Blue Mountain. The delineated boundary was defined by the base of the escarpments of adjacent uplands and the base of the Pocono escarpment. The escarpments refer to long steep slopes at the edge of plateaus. Portions of the boundary not based on the escarpments were arbitrarily made. The dominant topographic forms found in the section include rounded hills and valleys. The elevation of the section ranges from a minimum of 440 feet to a maximum of 2,690 feet. The local relief for the area is labeled low to moderate (101–600 feet). The underlying rock types include sandstone, siltstone, and shale. The geologic structure of the Glaciated Low Plateau section consists of low amplitude folds. The origin of the section's topography and landscape is attributed to the processes of fluvial and glacial erosion as well as glacial deposition. Glacial deposition refers to the melting of the glaciers, which in turn deposited sediments and minerals forming different landforms. These high energy occurrences and fluvial/glacial erosion formed the dendritic drainage pattern that can be seen in the section today.

Glaciated Pocono Plateau Section

The Glaciated Pocono Plateau section is smaller than most other sections within the state. The section abuts five sections: Glaciated Low Plateau, Anthracite Valley, Anthracite Upland, Susquehanna Lowland, and Blue Mountain. The boundaries of the Glaciated Pocono Plateau are defined by the base of the Pocono escarpment to the south and east. The north is demarcated by the crest of drainage divides, and the west boundary was arbitrarily delineated. The origin of the topography and the characteristics of the section are attributed to movement and sculpting by fluvial and glacial erosion along with glacial deposition. The high energy molding of the landscape produced a deranged drainage pattern in which the waterways are governed by the topography of the land. The dominant topographic form in the section includes broad, undulating upland surfaces with dissected margins. The upland surface's margins have been dissected by severe erosion over time, including the aforementioned fluvial and glacial erosion. The geologic structure of the section consists of beds having a low north dip and also incorporates some small folds. The underlying rock type in the section is made up of sandstone, siltstone, shale, and some conglomerate minerals. With the topographic description being made up of upland surfaces, the elevation would naturally be higher with less extreme changes in elevation or local relief. The section's elevation ranges between 1,200 and 2,320 feet above sea level.

Ridge and Valley

Anthracite Valley Section

The Anthracite Valley section is crescent shaped and runs through portions of Wayne, Susquehanna, Lackawanna, Luzerne, and Columbia Counties. The section's delineated boundary is a natural barrier around its perimeter—the outer base of the surrounding mountain. With the outer rim being mountain, the dominant topographic form of the section is a narrow to wide, canoe-shaped valley with irregular to linear hills enclosed by a steep-sloping mountain rim. The maximum elevation on the ridge tops reach 2,368 feet above sea level, and the minimum in the valley is approximately 500 feet above sea level with a low to moderate local relief. The section's shape is directly related to its origin in fluvial and glacial erosion and some glacial deposition. The way the valley was cut through explains the trellis and parallel drainage patterns that are found in the section. When the glaciers moved through the area they cut and gouged waterways parallel to one another; where the ice moved downwards through the valley it created trellis-like patterns with channels meeting at right angles. The underlying rock types found in the Anthracite Valley section include sandstone, siltstone, conglomerate stone, and anthracite. The geologic structure of the area is defined as a broad, doubly plunging syncline with faults and smaller folds.

Blue Mountain Section

The Blue Mountain section is a relatively thin, long strip running northeast to southwest between the Great Valley section and a large portion of the Anthracite Upland section. The southeastern boundary is defined by the base of the slope change on the southeast side of Blue Mountain, while the northwestern boundary is the base of the mountain and the base of the Pocono escarpment. The northeastern border of the section is an arbitrary delineation. The topography of the area is variable. The south is lined with a linear ridge and then valley to the north. The valley widens eastward and includes low linear ridges and shallow valleys. Within these valleys, the trellis drainage pattern has formed, with a larger river and smaller tributaries pouring into it at right angles. The topography of the section was created by a combination of fluvial (river/stream) erosion with some glacial erosion and deposition in the northeast of the section. The underlying rock types of the section include sandstone, siltstone, and shale with some limestone and conglomerate inclusions. The geologic structure of the section is again variable but is typified in the southwest by a south limb of broad fold and in the northeast by small folds north of Blue Mountain. The elevation of the section varies between 300 feet above sea level in the low valleys and 1,680 feet above sea level at its highest elevation along the ridge.

Great Valley Section

The Great Valley section is a long strip that stretches from New Jersey through Northampton County down to the southwest until crossing into Maryland through Franklin County. The boundaries of the section are defined to the north ast the base of slope change on the southeast side of Blue Mountain.

The southern border is at the base of slope change to the adjacent uplands. The dominant topography of the section is made up of a very broad valley with the northwest half being a dissected upland while the southeast half is a low karst terrain. The section has undergone fluvial erosion, solution of carbonite rocks, and some periglacial mass wasting. The dissected upland has been severely eroded, leaving an undulating or sharp relief. The underlying rock type of the northwest includes shale and sandstone with slate at the east end. These less dense minerals explain the dissected upland and how the terrain of the area was formed. The low karst terrain contains underlying rock types including limestone and dolomite. Both of these minerals are soluble bedrock meaning they are weathering resistant rocks. The drainage patterns are directly related to the rock types within the section. The areas with the permeable stone consist of dendritic patterns or branch-like tributaries. The areas containing dolomite and limestone created karst patterns that are underground caverns and waterways. The geologic structures of the Great Valley section include thrust sheets, nappes, overturned folds, and steep faults. The section also incorporates many third- and fourth-order folds. The elevation of the area has a minimum of 140 feet above sea level and a maximum elevation of 1,100 feet above sea level.

South Mountain Section

The South Mountain section is a very small section located in portions of Cumberland, Franklin, York, and Adams Counties before continuing across the border into Maryland. The section is sandwiched between the Great Valley section and the Gettysburg-Newark Lowland section. The boundary is defined as the base of slope changes to the adjacent lowlands. The dominant topographic forms include linear ridges, deep valleys, and flat uplands with moderate to high local relief. The section’s drainage pattern is defined as dendritic. The landscape of the section was sculpted by fluvial erosion of highly variable rocks and some periglacial mass wasting. The underlying rock types include metavolcanic rocks, quartzite, and some dolomite. The geologic structure of the section has major anticlinorium with second- and third-order folds. The anticlinorium refers to folds that are convex, with the oldest beds at the core of the landform. Unlike the elevation of the neighboring Great Valley section, the elevation is much higher with a minimum of 450 feet above sea level and a maximum of 2,080 feet above sea level.

New England

Reading Prong Section

The Reading Prong section is the only section in the New England physiographic province. The section is patchy and located along the eastern border of Pennsylvania and New Jersey. Most of the section is surrounded by the Great Valley section or abutting the Gettysburg-Newark Lowland section. The boundaries of the section are defined as the base of the slope change to the adjacent lowlands. The dominant topographic form is circular to linear rounded hills and ridges. The origin of the section is attributed to fluvial erosion and some periglacial mass wasting, which explains the dendritic drainage pattern found throughout the section. The geologic structure is made up of multiple nappes that have moved sideways over the neighboring strata or geologic structures. The underlying rock types in the section consist of granitic gneiss, granodiorite, quartzite, and jasper. While all lithic raw material types were considered important, jasper especially was prized by prehistoric peoples who likely settled on the Reading Prong due to its proximity to jasper. The elevation is very similar to the Great Valley section in which the Reading Prong section sits, with a low of 140 feet above sea level and a maximum of 1,364 feet above sea level.

Piedmont

Gettysburg-Newark Lowland Section

The Gettysburg-Newark Lowland section lies in the southeastern portion of Pennsylvania and stretches from New Jersey southwest into Maryland. The boundary of the section is delineated by the base of the slope changes with adjacent uplands and lowlands. The remaining portions of the section, lacking definite slope changes, were arbitrarily set. As the name suggests, the dominant topographic landform includes rolling lowlands, shallow valleys, and isolated hills. The geographic structure of the section is half graben with low, monoclinal, northwest-dipping beds. Graben and monoclinal structures refer to the earth's crust in an area that lies between two faults and is uniformly inclined in the same direction. The geologic structure of the section is directly related to the high energy fluvial erosion of rocks with variable resistance, which shaped the landscape. The underlying rock types within the area consist of mainly red shale, siltstone, and sandstone. There are also some conglomerate and diabase (type of igneous rock) present in the section, however. The drainage patterns throughout the section are defined as two types: dendritic and trellis, both having distinct characteristics. This section has the second lowest elevation within Regions 7, 8, 9, and 10, with a minimum of 20 feet above sea level and a maximum of 1,355 feet above sea level.

Piedmont Lowland Section

The Piedmont Lowland section is also located in the southeast of Pennsylvania. This section is smaller than the Gettysburg-Newark Lowland section, and portions of it are set within the Piedmont Upland section. The boundaries in the south are defined by the base of slope changes of the adjacent uplands and in the north by Mesozoic red rocks. The dominant topographic forms resemble very closely those of the Great Valley section in the Ridge and Valley physiographic province. The topography includes broad, moderately dissected (heavily eroded) karst valleys separated by broad low hills. The karst valleys are created by fluvial erosion that flows through areas with impermeable minerals such as the limestone and dolomite that are dominantly located in this section. Two other underlying rock types located in the area are phyllitic shale and sandstone related to the other origin of the section, periglacial mass wasting. The geologic structure of this section is described as complexly folded and faulted, due quite possibly to the drainage patterns and underlying rock types. The Piedmont Lowland section's drainage patterns include the obvious karst type with underground caverns and waterways, but also dendritic patterns. These are very common in many of the other sections in Pennsylvania. Much like the Gettysburg-Newark Lowland section, the elevation is relatively low compared to the other sections in the state. The minimum elevation is 60 feet above sea level and the maximum is 700 feet above sea level with low local relief.

Piedmont Upland Section

The third section in the Piedmont Province is the Piedmont Upland section, located south of the Gettysburg-Newark Lowland section and encompassing portions of the Piedmont Lowland section. The section also crosses the border into Maryland and Delaware. The boundaries of the section in the east are the base of low to vague fall line escarpment (long steep slopes) and to the north are demarcated at the base of slope change to adjacent lowlands. The dominant topographic forms in the section include broad, rounded to flat topped hills and shallow valleys. The geologic structure of the section is extremely, complexly folded and faulted. The underlying rock types include mainly schist, gneiss, and quartzite with some saprolite. The origin of the section is attributed to the fluvial erosion throughout the area and some occurrences of periglacial mass wasting. These events led directly to cutting in the landscape to create a dendritic drainage pattern. The minimum elevation of the Piedmont Upland section is slightly higher than the other sections in the province. The lowest elevation is 100 feet above sea level and the maximum is 1,220 feet above sea level.

Atlantic Coastal Plain

Lowland and Intermediate Upland Section

The Lowland and Intermediate Upland section is set at the most southeastern extent of Pennsylvania and is within Philadelphia and Delaware Counties. The section is a small sliver along the Pennsylvania border that carries over into New Jersey and Delaware. The boundaries to the section are defined to the northwest as the base of a low to vague fall line escarpment. The boundary to the east is an arbitrary demarcation. The dominant topographic forms in the section are described as a flat upper terrace surface that is cut by shallow valleys and the Delaware River floodplain. The topography and land formation was created by fluvial erosion and the deposition that followed from these high energy events. The underlying rock types are what you would expect in an area that has been built up mostly by flooding deposition. These minerals and sediments include unconsolidated to poorly consolidated sand and gravel. Underlying these deposits are minerals such as schist, gneiss, and other metamorphic rocks. The geologic structures of the area are much like the above described minerals in the section. They are unconsolidated deposits underlain by complexly folded and faulted rocks. The drainage pattern that cuts through the topography of the Lowland and Intermediate Upland section is categorized as the dendritic pattern. The dendritic pattern is the most reoccurring waterway designation in all of the topographic sections in Pennsylvania. The elevation of the section is the lowest in the state with a minimum of 0 feet above sea level and a maximum of 200 feet above sea level.

Study Region Delineation

As described in the report for Regions 1, 2, and 3 (Harris et al. 2014a), the state was divided into 10 modeling regions to ensure uniform modeling within similar landscapes and to help manage the large datasets (Figure 3). The boundaries for the 10 regions are based on grouping similar physiographic sections into regions of very roughly equal size (with the exception of Regions 3 and 10). The current report deals with the Regions 7, 8, 9, and 10. Because Region 10 is so small, it was merged with Region 9 for data management and computing purposes into Region 9/10. Nonetheless, each of the subareas within the combined region was modeled separately.

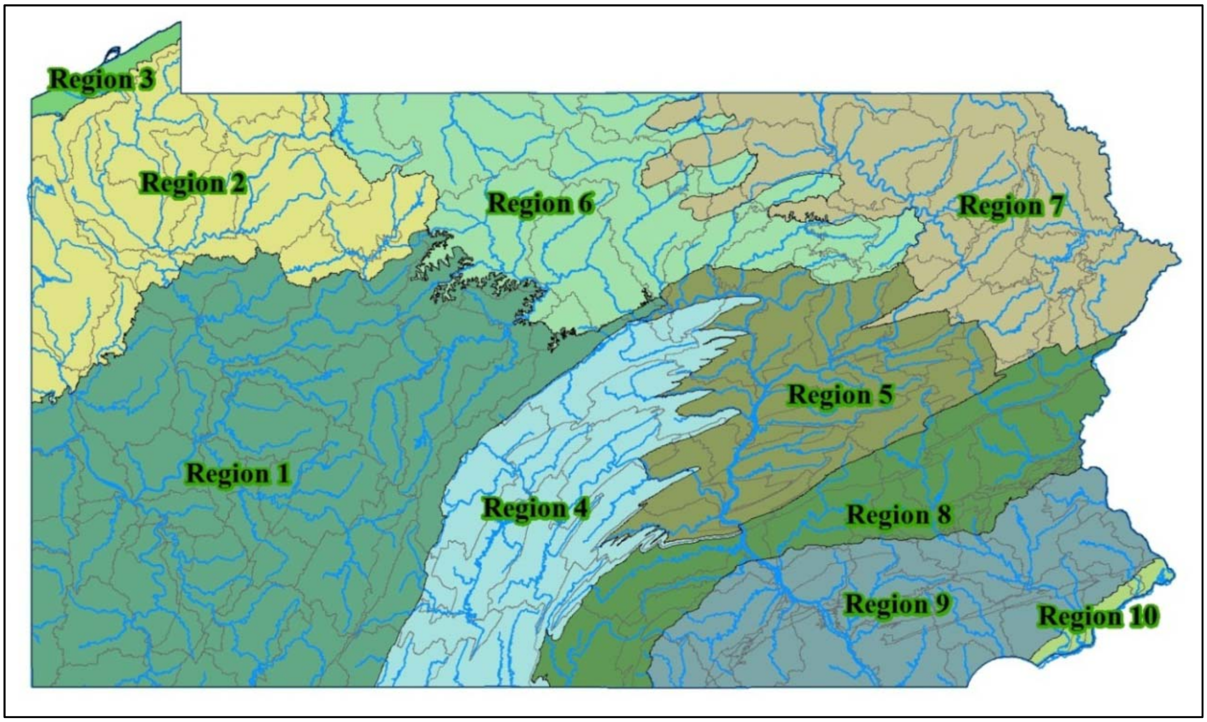


Figure 3 - Modeling regions for the Pennsylvania Model Set project.

In earlier reports, some of the regions were broken down into a small number of zones based on drainage basin boundaries within physiographic province, largely for data management purposes. For this report, Regions 7, 8, and 9/10 did not require division into zones (Table 2, Figure 4). Zones, where used, are further subdivided into units referred to as sections, which are based on watershed boundaries within physiographic sections (sections were referred to as “physio-sheds” in earlier reports for this project). As shown in Table 2, Regions 7 and 8 each contain nine sections, while Region 9/10 contains 15 sections.

Finally, each section was divided into upland and riverine subareas, shown in the final column in Table 2. Each subarea represents the study area for a single model, meaning that each subarea was run through the entire modeling process as an individual unit exclusive from the rest. For Regions 7, 8, and 9/10 there are a total of 66 subareas and, therefore, 66 separate model building efforts. The rationale and methodology for dividing the sections into upland and riverine settings is discussed in detail in the Task 4 report (Harris et al. 2014a). The results of various statistical tests and model metrics will be displayed and categorized by the subareas since these are the unit of analysis. Subareas will be differentiated by including other elements of the hierarchy such that the expression "R9/10_all_riverine_section_1" will refer to the riverine subarea of section 1 of Region 9/10. The modeled subareas are shown in Figure 5, Figure 6, and Figure 7.

Table 2 - Relationship between Regions, Zones, Sections, Subareas, and Physiography

Physiographic Province	Region	Zone	Physiographic Section	Section	Subarea
Appalachian Plateaus	7	All	Glaciated Low Plateau	1	riverine section 1
					upland section 1
				2	riverine section 2
					upland section 2
				3	riverine section 3
					upland section 3
				4	riverine section 4
					upland section 4
				5	riverine section 5
					upland section 5
				6	riverine section 6
					upland section 6
				7	riverine section 7
					upland section 7
			Glaciated Pocono Plateau	8	riverine section 8
					upland section 8
Ridge and Valley			Anthracite Valley	9	riverine section 9
					upland section 9
	8	All	Blue Mountain	1	riverine section 1
					upland section 1
			South Mountain	3	riverine section 3
					upland section 3
			Great Valley	4	riverine section 4
					upland section 4
				5	riverine section 5
					upland section 5
				6	riverine section 6
					upland section 6

Physiographic Province	Region	Zone	Physiographic Section	Section	Subarea
				7	riverine section 7
					upland section 7
				8	riverine section 8
					upland section 8
				9	riverine section 9
					upland section 9
New England			Reading Prong	2	riverine section 2
					upland section 2
Piedmont	9 (9/10)	All	Gettysburg-Newark Lowland	1	riverine section 1
					upland section 1
				2	riverine section 2
					upland section 2
				3	riverine section 3
					upland section 3
				4	riverine section 4
					upland section 4
				5	riverine section 5
					upland section 5
				6	riverine section 6
					upland section 6
				7	riverine section 7
					upland section 7
				8	riverine section 8
					upland section 8
			Piedmont Lowland	10	riverine section 10
					upland section 10
				11	riverine section 11
					upland section 11
				12	riverine section 12
					upland section 12
			Piedmont Upland	13	riverine section 13
					upland section 13
				12	riverine section 14
					upland section 14
				15	riverine section 15
					upland section 15
Atlantic Coastal Plain	10 (9/10)	All	Lowland and Intermediate Upland	9	riverine section 9
					upland section 9

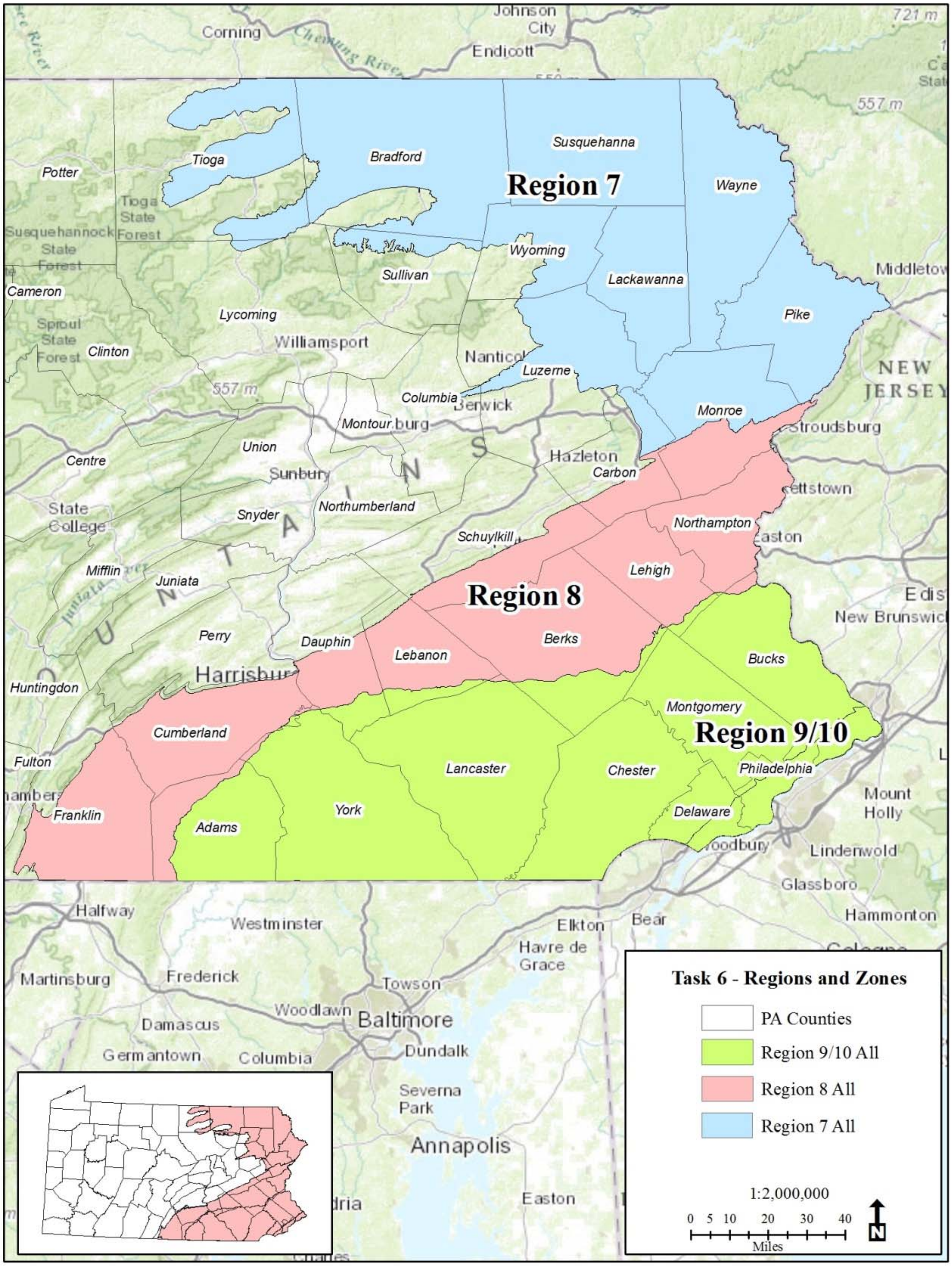


Figure 4 - Task 6 report regions.

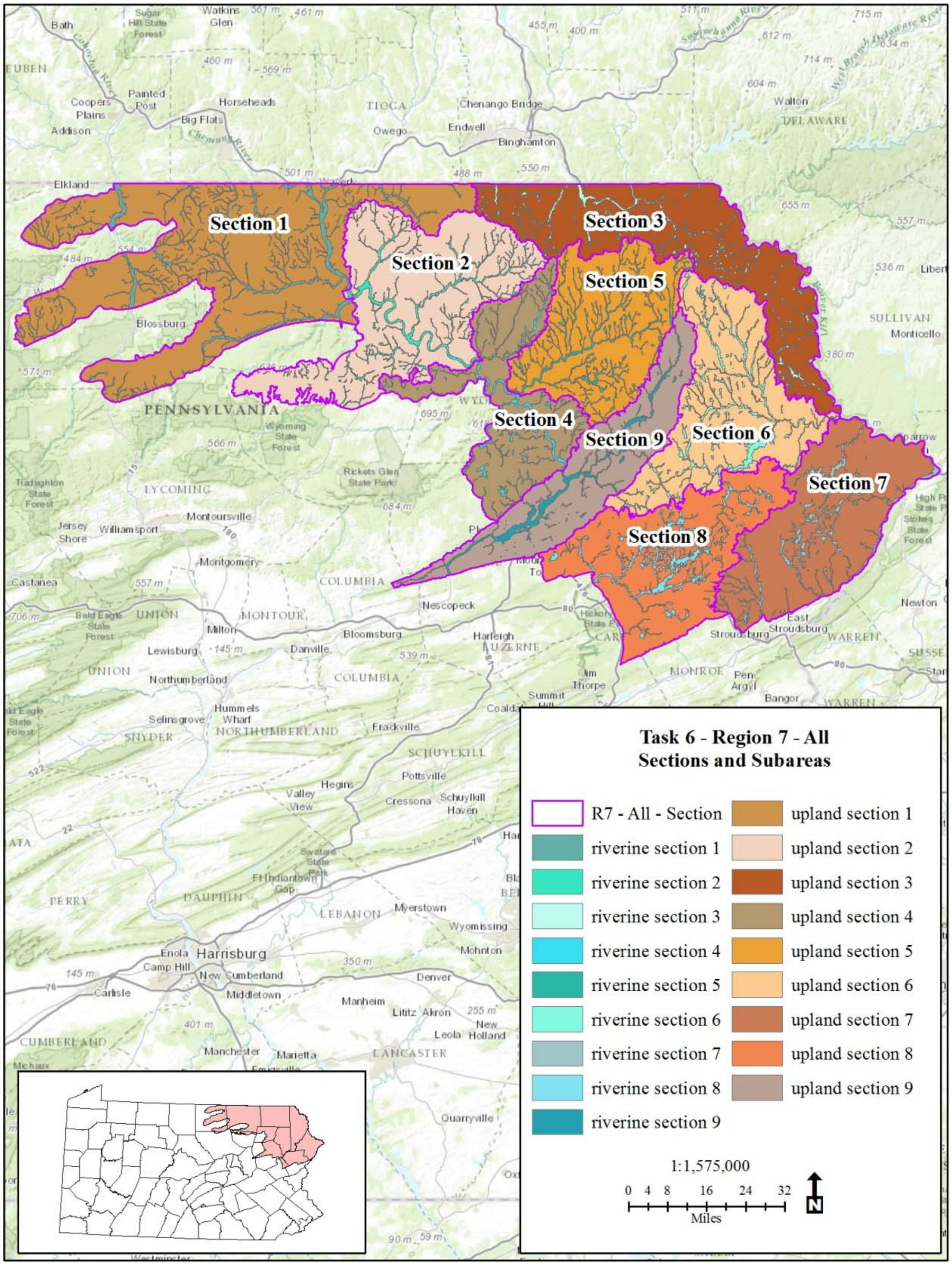


Figure 5 - Modeling subareas of Region 7.

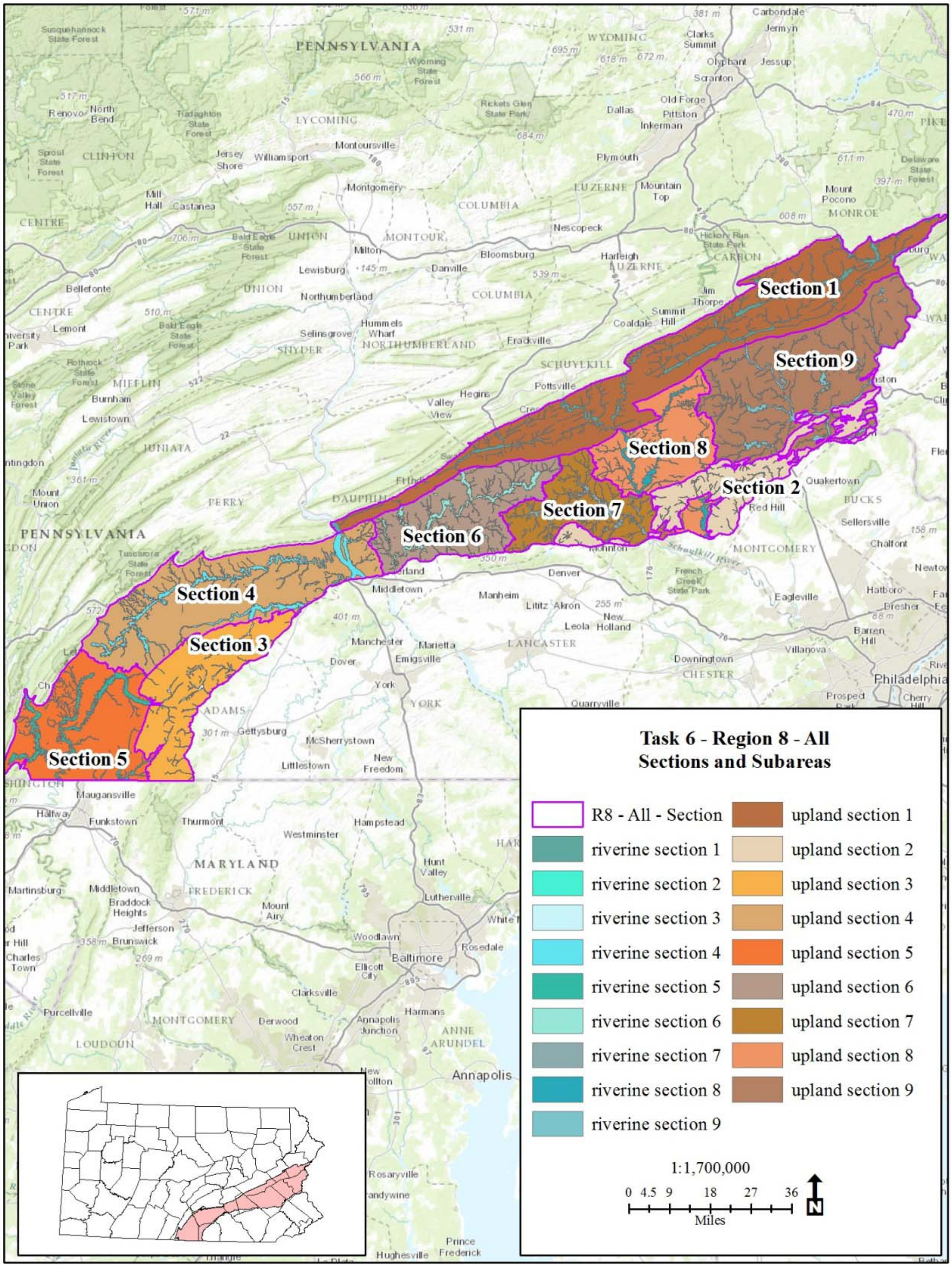


Figure 6 - Modeling subareas of Region 8.

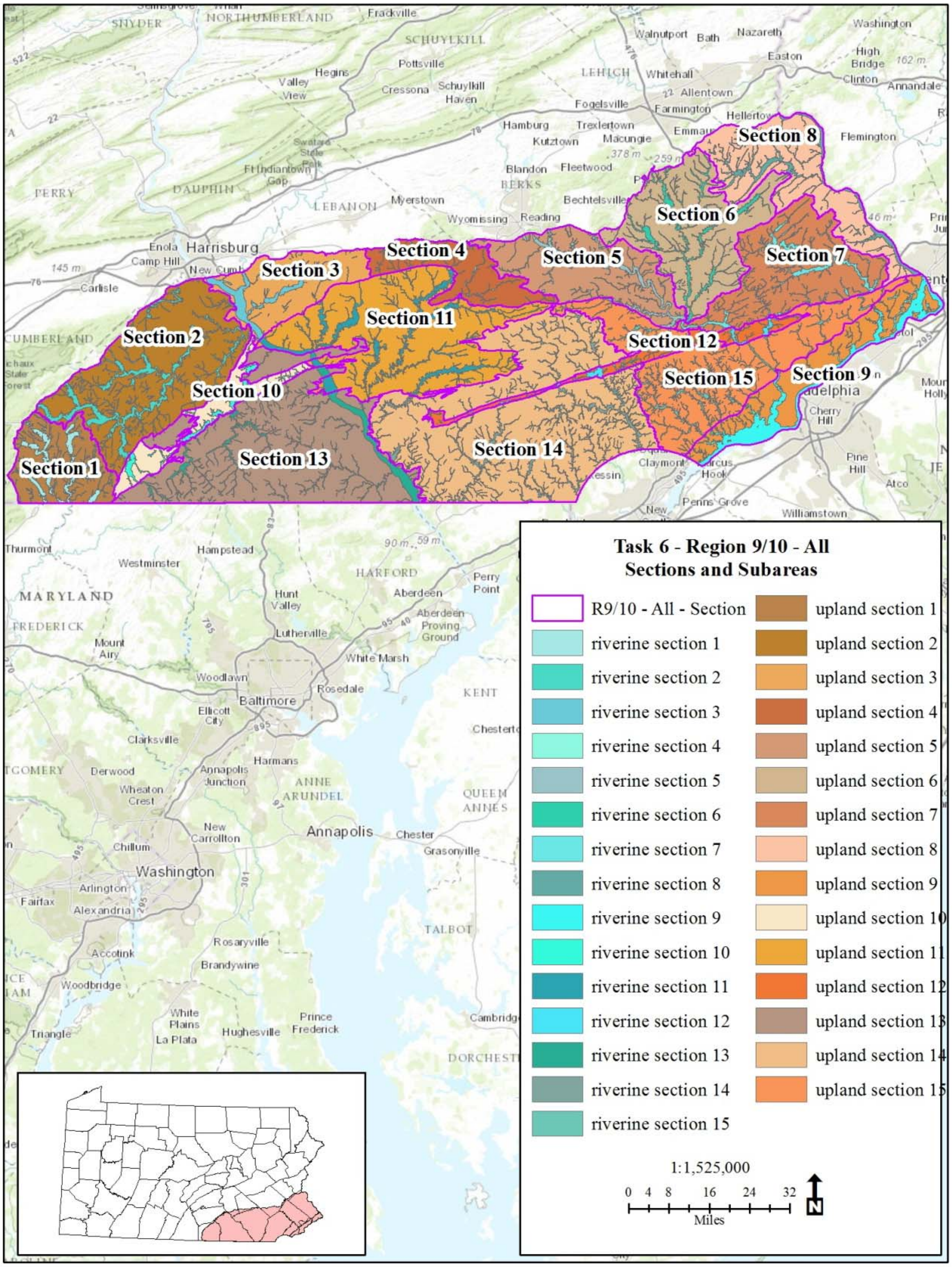


Figure 7 - Modeling subareas of Region 9/10.

PREHISTORIC BACKGROUND

The Peopling of the Americas and the Paleoindian Period

The first humans likely reached North America no earlier than about 30,000 years ago. The chronology of the Paleoindian period in Pennsylvania begins with a period known as Pre-Clovis, dating from about 14,000 to 9500 B.C. (Quinn et al. 1994). This date is largely supported through the extensive research performed at Meadowcroft Rockshelter in southwest Pennsylvania, which has a minimum early date of 9300 B.C., although Carr and Adovasio (2002:7) argue that the average date of the deepest deposits point to a Pre-Clovis occupation by 13,950 B.C. The Pre-Clovis material is marked by a distinct prismatic blade industry at Meadowcroft Rockshelter (Quinn et al. 1994).

Most evidence of early human occupation in eastern North America is associated with the Clovis period (9500–8000 B.C.), which is characterized primarily by its distinctive lithic assemblage. Fluted projectile points, usually produced from high-quality lithic material, are generally considered the diagnostic marker of the time period, along with scrapers and spurred gravers. In eastern Pennsylvania, the Clovis point is the earliest Paleoindian point type, followed by Debert, Mid-Paleo points, and Dalton-Hardaway points by the end of the period (Custer 1996:94). Bergman et al. (1998:84) recorded differing preferences for raw material in the Paleoindian period, based on physiographic locations. In the Piedmont province, jasper was the preferred material; Onondaga chert in the Ridge and Valley province was highly valued, and chert was preferred in both the glaciated plateau and unglaciated plateau sections.

Boyd et al. (2000:38) note that Paleoindians in the eastern United States likely employed a settlement pattern in which a small group would be highly mobile through part of the year, and then practice a semi-sedentary lifestyle the rest of the year, in accordance with the specific seasonally available resources that were the focus of subsistence at that particular time. This pattern results in two basic types of Paleoindian sites: base camps and short-term resource procurement camps. The short-term camp characterization subsumes other specialized site types, such as hunting stations, quarries, and isolated point finds. Boyd et al. (2000:43) also use the same site types for the subsequent Early Archaic period.

Carr and Adovasio (2002:36) provide some data about the location of Paleoindian sites within the various physiographic provinces covered by Regions 7, 8, 9, and 10. They note that 81% of Paleoindian sites in the Piedmont and Ridge and Valley provinces (including portions of Regions 7, 8, and 9) occur in the major stream valleys, close to the active floodplain. However, in the Great Valley section of the Ridge and Valley province in Region 8, Paleoindian sites tend to occur at higher elevations, possibly due to factors such as higher water tables resulting in wetter lowland conditions, or the use of the Great Valley for hunting caribou as they passed through gaps in the ridges.

Carr and Adovasio (2002:40–41) also note a difference between the settlement patterns of Paleoindian groups in the glaciated region of Pennsylvania versus the unglaciated region. Sites in the unglaciated region appear to be smaller, representative of small band territories, and show a focus on a foraging subsistence strategy. These sites are more focused on the floodplains and show a cyclical use of quarries, with low amounts of lithic material that suggests long-distance band movement. In contrast, sites in the glaciated region can be relatively large, exhibiting evidence of both large and small band territories, and focused on glacial features. Lithic procurement strategies in the glaciated region show a serial and cyclical use of quarries, with materials moving long distances from their source. Paleoindian bands in the glaciated region practiced a subsistence strategy that included both foraging and migratory game animal exploitation.

The best-documented Paleoindian occupation in eastern Pennsylvania is likely that of the Shawnee-Minisink site, located on the third terrace above the Delaware River in Monroe County. This stratified, multi-component site has been intensively studied and has revealed new insights about Paleoindian occupations in the Northeast. The Paleoindian components have been radiocarbon dated to ca. 11,000 B.C. and are present approximately 2.4 m below the modern ground surface (Gingerich 2007). The site may represent a repeatedly occupied transient camp, with an assemblage containing Clovis fluted points, end scrapers, and bifaces. However, Gingerich (2007) cautions that not enough is known about the site to conclusively determine if the site was an intensively occupied base camp or a series of overlapping, short-duration resource acquisition camps. Importantly, Shawnee-Minisink is one of the few sites to have contributed data on Paleoindian use of plant resources, which appear to have focused on hawthorn, hickory nuts, blackberry, and hackberry (Gingerich 2007:134).

The Archaic Period

The Archaic period is the longest documented temporal segment of prehistory in eastern North America. In Pennsylvania, it is typically divided into four subperiods: Early Archaic (8500–6000 B.C.), Middle Archaic (6000–4000 B.C.), Late Archaic (4000–1800 B.C.), and Terminal Archaic (1800–1000 B.C.), based on the marked differences in subsistence and settlement patterns (Quinn et al. 1994).

The Early Archaic Period (8500–6000 B.C.)

Small bands of Early Archaic hunter-gatherers appear to have been highly mobile and may have traveled across large territorial ranges and a variety of landforms (Jefferies 1990:150). Raber et al. (1998:121) note that Early Archaic lifeways show a high degree of continuity with the preceding Paleoindian period. Projectile points form a sequence within the Early Archaic period, beginning with Palmer types and ending with Kirk series points. Bergman et al. (1998:84) note a preference for jasper and rhyolite in the Piedmont province, with preferences for chert in the Ridge and Valley province and glaciated plateau and unglaciated plateau sections. MacDonald (2003) observes that in the western part of Region 7, Early Archaic sites are predominately open camps in lowland settings close to water.

Bergman et al. (1998:85) state that Early Archaic groups in the Ridge and Valley and Piedmont provinces showed a preference for riverine settings, although groups in the Appalachian province showed a slight preference for upland settings. The Early Archaic period is not well represented in the archaeological record for Regions 7, 8, 9, and 10. Siegel et al. (2001) note that site types and settlement patterns are essentially the same as the preceding Paleoindian period. Custer (1996) argues that there are only stylistic differences between Paleoindian and Early Archaic groups, who otherwise had very similar settlement and subsistence patterns, and feels the two periods should be combined for analysis.

The Middle Archaic Period (6000–4000 B.C.)

By the Middle Archaic, populations had shifted their movement strategies from high mobility to reduced mobility (Stafford 1994). The period saw a substantial increase in size of the regional population, and marked the transition from the cyclical settlement pattern focused on lithic resources practiced by Paleoindian and Early Archaic groups to one using seasonal base camps situated in floodplains; these transitional camps were aimed at specific resource exploitation in uplands (Harris et al. 2010:21). The appearance of ground stone tools and the related implication of increased plant usage also support the idea that Middle Archaic populations were somewhat more sedentary than those living in the region before them, and possessed greater knowledge of seasonally available resources and the best locations to access those resources.

Several technological innovations took place between the Early and Middle Archaic periods. The bifurcated-base point is typically seen as first occurring in the early Middle Archaic. Projectile point types of this time period in Pennsylvania include MacCorkle, LeCroy, St. Albans, Kanawha, Neville, Otter Creek, and Stanly (Justice 1995; Carr 1998:80). Problematically, triangular points have recently been found to occur in Middle Archaic deposits, which previously were solely associated with the Late Woodland period. Some archaeologists now argue that triangular points found in plow zone deposits cannot be automatically assigned to Late Woodland associations, and some sites previously identified as such based on triangular points may actually represent Middle Archaic occupations (Siegel et al. 2001). Ground stone tools such as axes, pitted stones, pestles, and grinding stones first appeared at this time (Jefferies 1996:48). In addition, archaeological evidence indicates that Middle Archaic people were also familiar with the atlatl, or spear thrower (Jefferies 1996:48). Lithic material preferences varied according to the different physiographic provinces occupied by Middle Archaic groups. In the Piedmont province, the use of jasper, so prevalent in the Paleoindian and Early Archaic periods, was displaced by a preference for locally available quartz (Bergman et al. 1998:84); in the Ridge and Valley province and glaciated plateau and unglaciated plateau sections, however, the preference for chert exhibited by Early Archaic groups continued with Middle Archaic people.

Middle Archaic sites are characterized by Boyd et al. (2000:50) as represented by the same two basic site types as the preceding periods (base camps and short-term camps), but Carr (1998:81) notes that in the Great Valley section of the Ridge and Valley physiographic province, base camps may represent smaller groups than during other time periods, possibly confined to members of a nuclear family group. Site types may include small base camps on terraces, specialized resource procurement camps in the uplands, and lithic processing camps near quarry locations (MacDonald 2003:63). Bergman et al. (1998:85) note that Middle Archaic groups generally preferred riverine settings. Middle Archaic groups may have differed primarily from their predecessors in exploiting a broader resource base (Stewart and Cavallo 1991).

The Sandts Eddy site (36NM12), located in Region 8, displayed the use of secondary source (river cobble) lithic materials and expedient tool use as important components of the lithic manufacturing process, characteristics that are thought to be representative of Middle Archaic culture in eastern Pennsylvania (Bergman et al. 1998:72).

The Late Archaic Period (4000–1800 B.C.)

Trends first seen in the Middle Archaic, such as the diversification of utilized plant resources and increased sedentism, continued into the Late Archaic period. Raber (2010) notes a general shift from early Middle Archaic residential mobility/foraging to a collecting strategy with base camps occupied for longer periods of time, possibly even for entire seasons, by the Terminal Archaic. The early Late Archaic in the Susquehanna drainage is best represented at the Memorial Park, East Bank, and Raker I sites (Hart 1995; East et al. 2002; Wyatt et al. 2005). The Memorial Park and East Bank sites, both located on broad floodplains of the West Branch, produced numerous artifacts and fire-related features that ranged between 4000 and 2500 B.C.

In eastern Pennsylvania, the Laurentian and Piedmont traditions are associated with the Late Archaic period. The Piedmont Tradition extends across the piedmont physiographic province from the Carolinas to New England and is noted for narrow-stemmed points, usually made of argillite, quartzite, and rhyolite, and a diversity of ground stone tools (Harris et al. 2010:22). The Laurentian Late Archaic lithic assemblage is dominated by a variety of side-notched and corner-notched point types, such as the Brewerton group, as well as hafted scrapers and ground stone tools, including celts and adzes for woodworking (Prufer and Long 1986; Dragoo 1976). Lithic material choices made by Late Archaic people showed they strongly favored jasper, chert, and rhyolite. Some evidence from sites in the southeastern United States indicates that Late Archaic populations began to experiment with fired clay (Sassaman 1993; Milanich 1994), though there is as yet no firm evidence that Late Archaic groups in Regions 7, 8, 9, or 10 were familiar with this technology.

Late Archaic settlement patterns became diversified compared to the preceding Middle Archaic, with numerous upland sites associated with lowland base camps focused on stable water resources. The diversification of site types is likely tied into a need to focus on known, predictable resources during a warming and drying period that coincided with part of the Late Archaic, known as the mid-postglacial xerothermic period. Late Archaic base camps were strategically located to take advantage of resources that could be exploited with minimal expenditures of labor (Raber et al. 1998:126).

The Terminal Archaic Period (1800–1000 B.C.)

The Terminal Archaic, also known as the Transitional period, is thought to be linked with a climatic change that resulted in warmer and dryer conditions (Custer 1996:187). Diagnostic artifacts associated with the Terminal Archaic include the Broadspear type projectile points, such as Lehigh, Susquehanna, and Perkiomen Broad points (Quinn et al. 1994). Other types associated with the Transitional Archaic include the Genesee type and Snook Hill type of the Genesee cluster (Justice 1987:159). An increased use of jasper and rhyolite indicates expansion of trade networks during the Terminal Archaic (MacDonald 2003). Steatite bowls first appear in this period. The earliest occurring pottery in Region 7 was found at the Sunny Side site, dated to the Terminal Archaic at ca. 1900 B.C., and was identified as Selden Island Cordmarked, featuring a steatite temper (MacDonald 2003:108).

The occurrence of fire-cracked rock (FCR) at Terminal Archaic sites appears to sharply increase from preceding periods (Harris et al. 2014c:19), perhaps related to an increased focused on anadromous fish and the use of earth ovens and stone boiling for processing large amounts of fish at the same time. The Lower Black's Eddy site is notable for its pavements of FCR, dating to the Late and Terminal Archaic periods. This site, which is associated with the Piedmont Late Archaic tradition and the Broadspear Terminal Archaic tradition, is thought to represent a heavily utilized seasonal occupation on a levee of the Delaware River, focusing on anadromous fish processing and argillite tool production (Kingsley et al. 1991). The Oberly Island site (36NM140) is similar to the Lower Black's Eddy site, although smaller in size, and demonstrates an apparent continuity of the Piedmont/Broadspear traditions from the Delaware River Valley upstream to the Lehigh River Valley (Siegel et al. 1999).

The Woodland Period

The Woodland Period in the eastern United States is generally associated with increased sedentary lifestyles and the introduction and widespread use of ceramic vessels. In Pennsylvania, the Woodland Period is usually divided into three temporal units: Early Woodland (1000–100 B.C.), Middle Woodland (100 B.C.–A.D. 1000), and Late Woodland (A.D. 1000–1620). A significant decline in Early and Middle Woodland sites occurs in eastern Pennsylvania, but it is unknown whether this reflects an actual demographic change (regional population decreases) or rather a masking effect resulting from difficulties is distinguishing regional variants of Early and Middle Woodland points from similar Late Archaic styles (Wyatt 2003). Raber (2003) notes that in Pennsylvania, especially in the east, there is difficulty in identifying and dating Early and Middle Woodland sites, due in no small part to scarce evidence for the highly distinctive Adena and Hopewell cultural traits in Pennsylvania, and largely to continuity with preceding Archaic cultural adaptations and technologies.

In general, Early and Middle Woodland groups in eastern Pennsylvania employed settlement patterns and basic site types very similar to those of the Late and Terminal Archaic. Custer (1996:237) notes, however, that base camps in the Early and Middle Woodland periods were larger and featured more storage features.

The Early Woodland Period (1000–100 B.C)

Early Woodland sites are rarely identified in eastern Pennsylvania, which may be attributed to populations adapting poorly to climatic downturns. The site identification issue may really be attributable, however, to smaller numbers of projectile points diagnostic to the period in comparison to the diversity of styles associated with the Archaic periods. Additionally, there may have been significant continuity of use of certain stemmed and notched Late Archaic styles into the Early Woodland period, further confusing identification of sites, especially those with Archaic and Woodland materials mixed in plow zone contexts (Custer 1996). The early adoption of domesticated plants is generally associated with the Early Woodland period in the Eastern Woodlands, but the timing of this slight increase in domestication varies regionally and does not occur in some areas until after A.D. 100. In general, evidence for Early Woodland horticulture seems rarely documented in Regions 7, 8, 9, and 10.

Site types represent a continuation of the base camp and short-term resource procurement camp model of seasonal settlement developed in the Late and Terminal Archaic, namely base camps in lowland settings such as floodplains and estuaries, and transient camps and resource-procurement sites away from the base camps in upland settings, such as rock shelters (Custer 1996:236). Settlement strategies appear to have begun maximizing resource acquisition efficiency and the production of surpluses.

The Early Woodland cultural complexes in eastern Pennsylvania include the Bushkill and Bare Island complexes (Custer 1996). Very little evidence for occupations by western cultural complexes, such as Adena and Meadowood, occur in Eastern Pennsylvania. Early Woodland pottery types include Vinette I, Marcey Creek, and Brodhead Net-marked (MacDonald 2003:117). Early Woodland occupations in southeastern Pennsylvania typically employed Vinette I, Dames Quarter, and Marcey Creek types in their assemblages (Harris et al. 2014c:20). Raber cautions that considerable variety occurs within the types associated with Early Woodland cultures (Raber 2003:8). Custer (1996:223) also notes that there is a general trend of decreasing vessel thickness over time with Early Woodland and Middle Woodland pottery types in eastern Pennsylvania. Diagnostic projectile points include Orient Fishtail, Meadowood, Hellgrammite, and to a much lesser degree Cresap Stemmed, Robbins, and Adena Stemmed styles.

The Middle Woodland Period (100 B.C.–A.D. 1000)

The Middle Woodland period in Pennsylvania was largely a continuation of cultural trends of the previous Early Woodland period, though regional differences occur, such as an increasing focus on maritime resources along the Coastal Plain (Raber 2003:12). Extensive trade networks are a hallmark of Middle Woodland cultures across the eastern United States, but Raber (2003) cautions that the degree to which individual Middle Woodland groups participated in trade networks is likely highly variable. Middle Woodland phases in Regions 7, 8, 9, and 10 include the Three Mile Island Complex in the Susquehanna drainage and the Abbott Complex in the Upper Delaware drainage (Custer 1996). Contemporary with early Middle Woodland Adena-influenced and Middlesex-affiliated cultures in the larger Mid-Atlantic is the Black Rock phase, associated with the Indian Point site (Kingsley et al. 1990). It is argued that this phase would appear to represent more of a continuation of Late Archaic-type lifestyles, but with Woodland technologies added (Harris et al. 2010:31).

Similar to the Early Woodland, Middle Woodland sites occur in lower frequency than those of the preceding Archaic period. Large sites occur on the Atlantic Coastal Plain, where groups may have aggregated to take advantage of maritime resources; these sites may have served as base camps from which smaller groups departed into areas away from the coastal plain for seasonal resource procurement forays, well into the Piedmont physiographic province (Raber 2003:19). Indeed, large Middle Woodland sites are largely absent in the Piedmont province. Very rarely identified in the PASS data of the Delaware River drainage, Middle Woodland components, when present, tend to be contained within plow zone deposits, even at stratified multi-component sites (Harris et al. 2014c:21). The elaborate mound and earthwork-building practices of Midwestern Middle Woodland cultures are not present in eastern Pennsylvania; rather only a few burial mounds are associated with Middle Woodland cultural groups, mainly in the Susquehanna drainage.

Very little is known about how this period differs from the preceding Early Woodland period in this region, with the exception that ceramic technology appears to have experienced a flowering of experimentation, with numerous different ceramic types identifiable to the period. Some early types include the rock-tempered Popes Creek, Wolfe Neck Net-impressed, and Broadhead Net-impressed. Later in the Middle Woodland period, a variety of shell-tempered types was introduced, represented by a number of different Abbot Farm types and net-impressed Mockley wares. Ceramic types from the preceding Early Woodland also persisted into the Middle Woodland, such as Vinette I. Custer (1996:239) notes that while in general ceramic technology in the Middle Woodland was similar to the preceding Early Woodland period, there does appear to be a significant increase in large storage vessels at Middle Woodland sites. Diagnostic lithic artifacts of the Middle Woodland period in eastern Pennsylvania include Rossville, Fox Creek, Levanna, and Jack’s Reef projectile point types. Fox Creek points are associated with the early part of the Middle Woodland period, while Jack’s Reef types represent the latter part.

The Late Woodland Period (A.D. 1000–1550)

The Late Woodland period in general is marked by a move toward nucleated, fortified settlements and the emergence of maize-based agricultural groups (Griffin 1967). The Late Woodland Period in eastern Pennsylvania is characterized by an apparent population expansion or large-scale movement of people, with several times the number of sites identified than in the preceding period. Late Woodland cultures in eastern Pennsylvania include the Minguannan complex in the southeastern portion of the state (including parts of Regions 8, 9, and 10), along with the early Clemson Island, Overpeck, and the subsequent Shenks Ferry complexes in parts of the northeastern portion of the state, in Regions 7 and 8. By the end of the Late Woodland period, the Shenks Ferry groups may have developed into the historical Susquehannock nation, which had migrated down the Susquehanna valley into eastern Pennsylvania. Alternatively, the Susquehannock people may have replaced the Shenks Ferry culture, either through forcible replacement or expansion to occupy territory when the Shenks Ferry groups migrated out of the region. However, the appearance of fortified villages during the later phases of the Shenks Ferry culture strongly suggests conflict played a major part in the disappearance of the culture at the end of the Late Woodland period, while the lack of Contact-period trade goods at Shenks Ferry sites strongly suggests that this cultural expression was no longer present at the end of the Late Woodland in eastern Pennsylvania. Meanwhile, the Minguannan culture is thought to be ancestral to the Contact Period Lenape culture (Harris et al. 2014c:24). The main diagnostic projectile point associated with the Late Woodland period is the small triangular arrowhead, representing the widespread adoption of bow-and-arrow technology. Early ceramic types associated with the Late Woodland include the Owasco/Clemson Island and Shenks Ferry series in the Susquehanna valley and Minguannan series in the Delaware valley; these two series appear to correlate to Iroquoian language groups in the Susquehanna valley and Algonkian language groups in the Delaware valley, respectively (Custer 1996:270). By about A.D. 1300, Iroquoian-like ceramics began to appear in Late Woodland assemblages (Custer 1996:266).

Late Woodland site types include villages, agricultural hamlets, and special-purpose short-duration camps. Very early Shenks Ferry sites in the Susquehanna River Valley do not appear to include villages as a site type, however (Custer 1996:276). Villages only start to appear in the Shenks Ferry complex after about A.D. 1300. Late Woodland villages tend to be circular and surrounded by a stockade, some exhibiting a regular planned placement of houses, while with others, the house placement appears more haphazard. Custer (1996:281) notes that the term “stockade” is probably misleading, and these villages would rather appear fenced. Clemson Island sites in the Susquehanna River Valley included villages with associated burial mounds, hamlets or small villages lacking mounds, and specialized camps (Miller et al. 2007:60). In contrast to Clemson Island and Shenks Ferry, the early Minguannan complex of the Delaware River Valley was a continuation of preceding Woodland settlement patterns, consisting of seasonal base camps and short-term resource procurement sites, and completely lacked either hamlets or villages (Custer 1996:289).

REGION 7 SITES

There are 1,033 archaeological sites in the PASS database with prehistoric components in Region 7. (Table 3 shows a breakdown of the Region 7 sites by site type and landform; individual tables for each of the time periods are included in Appendix B). A total of 467 prehistoric sites in the PASS database did not possess diagnostic material and were not assigned to a temporal period. In addition, there are 63 Archaic-period site components that could not be more specifically assigned to one of the Archaic sub-periods, and 78 Woodland-period site components with a similar issue.

A total of 982 sites in Region 7 had landform associations recorded in the PASS database. Site locations in Region 7 appear to show a strong trend for lowland settings, with 81.3% of all sites with landform information in the PASS database located in lowland settings (n = 798). The floodplain landform alone accounts for 48.5% of all site locations with landform data in Region 7 (n = 476). The only site types unique to lowland settings in Region 7 are the Village site type and the Cemetery site type. The two most commonly occurring site types, Open habitation, prehistoric (n = 565) and Open prehistoric site, unknown function (n = 155), both predominately occur in lowland settings. The apparent trend toward site location in lowland settings in Region 7, however, may simply reflect survey bias rather than an actual prehistoric landform preference.

Paleoindian

Within Region 7, there have been 16 sites identified with Paleoindian components, according to the PASS database. Eleven sites with Paleoindian components also contain one or more components dating to later time periods. Paleoindian sites in Region 7 have mainly been identified in lowland settings, primarily on floodplains. Single-component Paleoindian sites include two isolated finds of fluted points and three Open habitation, prehistoric sites. One notable Paleoindian site, the Trojan site (36BR149), produced a variety of tool types associated with the Paleoindian component, including fluted points, scrapers, prismatic blades, and drills. Lithics in the assemblage came from as far west as Ohio. The site is interpreted as a hunting camp location that was repeatedly occupied by a Paleoindian group practicing a highly mobile foraging strategy (McCracken 1989).

Early Archaic

The PASS database records 18 sites with Early Archaic components in Region 7. Early Archaic sites in Region 7 are almost exclusively found in landform settings that are close to water sources. There are only two single-component Early Archaic sites in the PASS data for Region 7: one Open habitation, prehistoric site and one Rock shelter/cave site. The paucity of Early Archaic sites in Region 7 does not allow for meaningful analysis of site functions in relation to topography.

Table 3. Region 7 Site Types by Landform

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge /Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Burial Mound	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	3
Cemetery	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	3
Earthwork	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Isolated Find	0	3	0	0	0	1	0	0	0	0	0	1	0	0	1	1	7
Lithic Reduction	0	5	0	0	0	2	0	0	0	3	0	0	1	2	1	6	20
Open Habitation, Prehistoric	6	301	13	5	77	90	20	7	5	4	4	2	3	14	2	12	565
Open Prehistoric Site, Unknown Function	2	62	9	1	11	33	1	7	0	7	4	1	0	5	1	11	155
Other Specialized Aboriginal Site	0	13	0	0	1	0	0	0	0	1	0	0	0	0	0	1	16
Quarry	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Rock Shelter/Cave	0	0	0	0	8	9	2	49	1	6	6	0	0	3	0	1	85
Petroglyph/Pictogram	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unknown Function Open Site Greater than 20 m Radius	0	2	0	0	0	3	0	0	0	0	0	0	0	0	0	1	6
Unknown Function Surface Scatter Less than 20 m Radius	1	6	0	0	2	2	0	0	0	2	0	0	0	0	0	2	15
Village	0	24	4	0	0	4	0	0	0	0	0	0	0	0	0	0	32
(blank)	5	59	1	2	11	15	0	3	1	5	4	0	0	2	1	16	125
Total	14	476	28	8	112	160	24	66	7	29	18	4	4	26	6	51	1033

Middle Archaic

The PASS database includes 75 sites with Middle Archaic components in Region 7, a significant rise in site frequency from the preceding Early Archaic period. Middle Archaic sites in Region 7 are primarily located in lowland physiographic settings. This large increase in site frequency may represent population growth, or perhaps more likely the expansion of Middle Archaic group territories into Region 7 from elsewhere in the Northeast. As with the Early Archaic period, most Middle Archaic sites compose part of a multi-component site, with only six single-component sites in the region: two Open habitation, prehistoric sites; three Open prehistoric site, unknown function sites; and one Rock shelter/cave site. The lack of single-component Middle Archaic sites in Region 7 makes analysis of site types in relation to landform association untenable for this study.

Late Archaic

The PASS database includes 228 sites with Late Archaic components in Region 7, a tripling in site frequency from the preceding Middle Archaic. A total of 221 sites had landform data associated with their records in the PASS database. The increase in the number of recorded sites may indicate a population expansion within existing groups in the area, or alternately a population movement into the area of Region 7. Late Archaic sites in Region 7 show a strong emphasis on lowland settings for Late Archaic site distribution, with 85.5% of all sites with landform data found in lowland settings (n = 189). Floodplain settings alone account for 49.7% of all sites with landform information in the PASS database. Single-component Late Archaic site types that may represent the likeliest candidates for seasonal occupation sites, such as base camps and short-term resource extraction camps, are Open habitation, prehistoric; Open prehistoric site, unknown function; and Unknown function open site, greater than 20 m radius. The landforms that contain the greatest number of Open habitation, prehistoric sites, which likely include a number of base camps, are typically lowland settings, with only four such sites identified in upland landforms in Region 7. Only eight Open prehistoric site, unknown function sites were identified in Region 7, and slightly more than half were found in lowland settings.

Terminal Archaic

The PASS database includes 175 sites with Terminal Archaic components in Region 7. There are 114 Terminal Archaic multi-component sites possessing components from either or both the Late Archaic and Early Woodland periods, representing 65.2% of the total population of Terminal Archaic sites. The fact that Terminal Archaic site components are frequently located at sites with preceding Late Archaic and subsequent Early Woodland components suggests group continuity within Region 7 between the Late Archaic and Early Woodland periods.

Terminal Archaic sites in Region 7 show a similar focus toward lowland physiographic settings as with the preceding Late Archaic period, with 88.1% of all Terminal Archaic sites with landform information in the PASS database located in lowland settings (n = 20, out of 168 sites with landform data). Single-component Terminal Archaic site types that may represent the likeliest candidates for seasonal occupation sites, such as base camps and short-term resource extraction camps, are Open habitation, prehistoric; Open prehistoric site, unknown function; and Unknown function open site, greater than 20 m radius. These site types are almost exclusively found in lowland settings. Multi-component sites with Terminal Archaic components occur in a greater number of different upland settings in comparison to the single-component sites.

Early Woodland

The PASS database includes 105 sites with Early Woodland components in Region 7, a drop in site frequency from the Terminal Archaic by 40%. There are 75 Early Woodland multi-component sites possessing either one or both Terminal Archaic and Middle Woodland components, representing 71.4% of the total population of Early Woodland sites, suggesting strong group continuity within Region 7 between the Terminal Archaic and Middle Woodland periods.

Lowland physiographic settings account for nearly all site locations with Early Woodland component. There are only two single-component Early Woodland sites in Region 7: one Lithic reduction site located on a terrace and a Rock shelter/cave site located in a middle slope setting. It seems probable that site types associated with Early Woodland groups represent a continuum of activities from the Archaic through the Woodland periods in east-central Pennsylvania.

Middle Woodland

The PASS database includes 75 sites with Middle Woodland components in Region 7, apparently representing a continuation of decreasing site frequency that began with the Terminal Archaic period in Region 7. There are 64 Middle Woodland multi-component sites possessing either or both Early and Late Woodland components, representing 85.3% of the total population of Late Woodland sites. The fact that Middle Woodland site components are strongly associated with preceding Early Woodland and subsequent Late Woodland components suggests group continuity within Region 7 between the three Woodland periods. Middle Woodland sites in Region 7 show a marked focus toward lowland physiographic settings, with 90.5% of all Middle Woodland sites with landform information located in lowlands. Middle Woodland groups may have preferred flood plain settings, with 67.6% of all Middle Woodland sites with landform data found in that setting. There are only five single-component Middle Woodland sites in Region 7, and they likely represent seasonal occupation sites such as base camps and short-term resource extraction camps rather than year-round occupations such as hamlets or villages.

Late Woodland

The PASS data for Region 7 includes 307 sites with Late Woodland components, a significant increase in site frequency from the Middle Woodland period. There are 60 Late Woodland multi-component sites possessing Middle Woodland components, representing 19.5% of the total population of Late Woodland sites, likely a result of there being far more Late Woodland components than Middle Woodland components at sites in Region 7.

Late Woodland sites in Region 7 show a strong focus toward lowland physiographic settings. Village sites are perhaps the defining site type for the Late Woodland, with 20 such sites identified in Region 7. Single-component Late Woodland Open habitation, prehistoric sites, which likely includes a number of base camp sites, are primarily found in lowland settings (93.7%). Rock shelters or caves also may have served as short-term resource extraction camps or seasonal base camps during the Late Woodland: all but three single-component Late Woodland Rock shelter/cave sites are found in upland settings in Region 7.

REGION 8 SITES

There are 2,526 archaeological sites with prehistoric components in Region 8 (Table 4 shows a breakdown of the Region 8 sites by site type and landform; individual tables for each of the time periods are included in Appendix B). A total of 1,398 sites in the PASS database did not possess diagnostic material and were not assigned to a temporal period. In addition, there are 175 Archaic site components that could not be specifically assigned to one of the Archaic sub-periods, and 77 Woodland site components with a similar issue.

Site types in Region 8 are largely found in lowland settings, with 72.4% of all sites (n = 1,673) located in lowland settings. Stream benches (n = 612) and terraces (n = 606) were the landforms most commonly associated with site locations, followed by floodplains (n = 419). Only one site type is restricted to lowland settings in Region 8, the Cemetery type, while the Rock shelter/cave site type is exclusive to upland settings. The most commonly occurring site type is Open habitation, prehistoric (n = 1,765), which occurs predominantly in lowland settings. Additionally, Region 8 contains the Hardystown Jasper Prehistoric District, which consists of numerous upland jasper quarries that were used throughout prehistory, such as the Vera Cruz site (36LH12; Walker et al. 2012) and the Kings Quarry site (36LH2; Stewart and Schindler 2008).The jasper from these quarries was apparently in especially high demand during the Late Archaic, Middle Woodland, and Late Woodland periods (Hatch 1994).

Table 4. Region 8 Site Types by Landform

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge /Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Burial Mound	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cemetery	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Earthwork	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Isolated Find	0	2	0	0	1	1	1	0	0	0	0	0	0	2	0	0	7
Lithic Reduction	1	6	3	0	13	22	0	11	1	7	0	0	2	12	0	76	154
Open Habitation, Prehistoric	1	330	4	6	516	461	160	49	29	11	9	10	10	112	4	53	1765
Open Prehistoric Site, Unknown Function	0	42	7	1	30	58	10	16	6	11	3	8	1	26	4	33	256
Other Specialized Aboriginal Site	0	1	0	0	4	2	0	3	0	0	0	0	0	1	0	0	11
Quarry	0	1	0	0	17	2	1	10	2	0	1	3	1	1	0	8	47
Rock Shelter/Cave	0	0	0	0	0	0	0	8	0	0	1	0	0	0	1	1	11
Unknown Function Open Site Greater than 20 m Radius	0	1	0	0	0	1	0	0	0	3	0	0	0	2	0	2	9
Unknown Function Surface Scatter Less than 20 m Radius	0	4	1	0	0	7	4	0	0	7	1	2	0	2	0	0	28
Village	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(blank)	5	32	5	2	30	52	7	11	3	9	10	6	5	17	1	42	237
Total	7	419	20	9	612	606	183	108	41	48	25	29	19	175	10	215	2526

Paleoindian

Within Region 8, there have been 31 sites identified with Paleoindian components, according to the PASS database. Twenty sites with Paleoindian components also contain one or more components dating to later time periods. Paleoindian sites in Region 8 are predominately found in lowland settings, with nine sites each identified in floodplain and terrace settings, and another six sites located on stream benches. Upland locations include hill slopes, middle slopes, ridgetops, and upland flats.

The 11 single-component Paleoindian sites in the PASS data for Region 8 include four Isolated findspots, six Open habitation, prehistoric sites, and one Open prehistoric site, unknown function site type. The Shawnee-Minisink Site (36MR43) is located within Region 8 at the confluence of Brodhead Creek with the Delaware River. This important stratified, multi-component site is primarily known for its Paleoindian components showing a focus on local resource exploitation, a variety of plant remains recovered from hearth features, and the use of fish as animal protein (Dent 2002:56).

Early Archaic

The PASS database records 51 sites with Early Archaic components in Region 8. Similar to the preceding Paleoindian Period, Early Archaic sites in Region 8 are predominately found in lowland physiographic settings. Nearly all of the Early Archaic sites in the PASS database were part of a multi-component site. The single-component Early Archaic sites include one site each of the following types: Open habitation, prehistoric; Open prehistoric site, unknown function; Quarry; Rock shelter/cave; and Unknown function surface scatter less than 20 m radius. In addition, one single-component Early Archaic site had no recorded site type in the PASS database. Interestingly, only one of the single-component sites (36CU0189, the Stillpond Farm site) was located in a lowland setting This site is an Open prehistoric site, unknown function site type located on a floodplain.

Middle Archaic

The PASS database includes 162 sites with Middle Archaic components in Region 8. The Middle Archaic period in the Susquehanna River Valley was a time of apparent dramatic population increase, with over three times the number of sites recorded with Middle Archaic components in comparison to the preceding Early Archaic period. Continuing an apparent trend in Region 8, most of the Middle Archaic sites are found in lowland settings, primarily on terraces but also frequently on floodplains and stream benches. There are only 18 single-component Middle Archaic sites types recorded in the PASS database: 11 Open habitation, prehistoric sites; 4 Open prehistoric site, unknown function sites; 1 Isolated find; 1 Lithic reduction site; and one Unknown function site greater than 20 m radius site. In addition, six single-component Middle Archaic sites did not have a site type recorded in the PASS database. The single-component sites are fairly evenly split between upland and lowland settings, with slightly more Open prehistoric site, unknown function sites found in upland settings than in lowland settings. This site type likely represents small group seasonal camps and thus may indicate an expansion of Middle Archaic seasonal rounds between upland and lowland resource locations.

Late Archaic

The PASS database includes 683 sites with Late Archaic components in Region 8, continuing a trend of dramatic population increase from the Middle Archaic. Late Archaic sites in Region 8 appear to focus strongly toward lowland physiographic settings, with site locations distributed mainly in floodplain, stream bench, and terrace settings. In the uplands, the hill ridge/toe and upland flat settings are where Late Archaic sites are most commonly found. There are 254 single-component Late Archaic sites. Single-component Late Archaic site types that may represent the likeliest candidates for seasonal occupation sites, such as base camps and short-term resource extraction camps, are Open habitation, prehistoric (n = 184) and Open prehistoric site, unknown function (n = 34). The other single-component site types include Lithic reduction sites (n = 7), Quarry (n = 4), and Unknown function open site greater than 20 m radius (n = 2). In addition, there are 26 single-component Late Archaic sites without identified site types in the PASS database. The single-component sites are mainly found in lowland settings, with the exception of the quarries, which are primarily found in uplands.

Terminal Archaic

The PASS database includes 346 sites with Terminal Archaic components in Region 8, a marked decrease from the Late Archaic period. There are 260 Terminal Archaic multi-component sites also possessing either Late Archaic or Early Woodland components (or both), representing 75.1% of the total population of Terminal Archaic sites. The fact that Terminal Archaic site components are strongly associated with preceding Late Archaic and subsequent Early Woodland components suggests group continuity within Region 8 between the Late Archaic and Early Woodland periods.

Terminal Archaic sites in Region 8 largely occur in lowland physiographic settings, with 68.4% of all Terminal Archaic sites with landform data found in that setting (n = 230). Terraces, stream benches, and floodplains have similar numbers of Terminal Archaic sites in the lowlands, with an apparent preference for the hill ridge/toe setting in the uplands. Single-component Terminal Archaic site types that may represent the likeliest candidates for seasonal occupation sites, such as base camps and short-term resource extraction camps, are Open habitation, prehistoric and Open prehistoric site, unknown function. Open habitation, prehistoric sites, which likely include a number of base camps, are more commonly found in upland settings, with nearly a third of this site type located on the hill ridge/toe landform. The Open prehistoric site, unknown function site type, however, is primarily located on lowland landforms.

Early Woodland

The PASS database includes 123 sites with Early Woodland components in Region 8. There are 84 Early Woodland multi-component sites possessing either Terminal Archaic or Middle Woodland components (or both), representing 68.3% of the total population of Early Woodland sites. Early

Woodland site components are less strongly associated with preceding Terminal Archaic and subsequent Middle Woodland components, possibly attributable to a decline in site frequency by nearly 65% from the preceding Terminal Archaic period. This decline may represent a population decrease, such as through out-migration; alternatively, the decline in site numbers could reflect the difficulty in identifying Early Woodland sites during archaeological survey, especially when considering the hypothesis that certain Late/Terminal Archaic stemmed points may have continued to be manufactured during the Early Woodland period.

Early Woodland sites in Region 8 show a marked focus toward lowland physiographic settings, with 83 Early Woodland sites identified in lowlands (70.9% of all Early Woodland sites). Single-component Early Woodland site types that may represent the likeliest candidates for seasonal occupation sites, such as base camps and short-term resource extraction camps, include Open habitation, prehistoric and Open prehistoric site, unknown function. There are very few single-component examples of either site type for the Early Woodland period in Region 8, with four Open habitation, prehistoric sites and one Open prehistoric site, unknown function site. No single-component ceremonial sites (burial mounds, earthworks) or sites indicative of a more sedentary lifestyle (such as villages or cemeteries) are present in the PASS data.

Middle Woodland

The PASS database includes 103 sites with Middle Woodland components in Region 8, continuing a decline in site frequency in the region that appears to have begun in the Terminal Archaic period. The reason for the decline in site frequency could be populations aggregating at fewer numbers of sites, but the site types identified for both single-component and multi-component Middle Woodland sites do not suggest that hamlets or villages existed in Region 8 during this time period. Additionally, no ceremonial sites attributable to the Middle Woodland are present in Region 8. The idea that Middle Woodland groups in southeastern Pennsylvania began to practice a seasonal settlement pattern, with large base camps in the Coastal Plain and smaller resource procurement camps ranging up the river valleys to the Piedmont, could also explain the drop in site numbers (Raber 2003:19), as the regional population may have aggregated into larger bands that split into smaller groups and occupied the Piedmont primarily on a seasonal basis. There are 77 Middle Woodland multi-component sites possessing either an Early or Late Woodland component (or containing material from all three Woodland periods), representing 74.5% of the total population of Middle Woodland sites. The fact that Middle Woodland site components are very strongly associated with preceding Early Woodland and subsequent Late Woodland components suggests group continuity within Region 8 between the three Woodland periods.

Middle Woodland sites in Region 8 show a marked focus toward lowland physiographic settings, with 73.2% of all Middle Woodland sites located in lowlands. There are only nine single-component Middle Woodland sites in the PASS database: seven Open habitation, prehistoric sites, one Open prehistoric site, unknown function site, and one site without an identified site type. The Open habitation, prehistoric site type may represent the likeliest candidates for seasonal occupation sites, such as base camps and short-term resource extraction camps; these sites in Region 8 are found mainly in lowland settings, although examples are present on middle slopes and upland flats.

Late Woodland

The PASS data for Region 8 includes 359 sites with Late Woodland components in Region 8. There are 62 Late Woodland multi-component sites possessing Middle Woodland components, representing 17.3% of the total population of Late Woodland sites, a reflection of the huge increase in site frequency between the Middle and Late Woodland periods. This increase in site frequency indicates either a large population explosion in local groups, or an influx of Late Woodland groups expanding into the region from elsewhere; the latter explanation seems the likeliest hypothesis for the dramatic increase in site numbers from the Middle Woodland to the Late Woodland, especially with the occurrence of fortified villages late in the period in parts of eastern Pennsylvania (although apparently not in Region 8 particularly).

Late Woodland sites in Region 8 show a strong focus toward lowland physiographic settings, with 73.7% of all Late Woodland sites occurring there. Village sites are often seen as the defining site type for the Late Woodland period, but no sites classified as villages appear in the PASS data for Region 8. The lack of villages could mean that Late Woodland groups were much less sedentary than their neighbors. Lawrence and Albright (2012:7-2) propose as part of their analysis of the Late Woodland River Road site (36BU379) that Late Woodland groups practiced different settlement patterns, with one group focusing on interior drainages in the Piedmont, while another occupied broad terraces in the Delaware River Valley. Both settlement patterns would involve seasonal occupations of large base camps in spring-summer, with smaller groups splitting off in the fall and winter to focus on upland resources.

Late Woodland single-component site types show slightly more diversity than previous time periods, including one cemetery, but otherwise seem to represent similar types of sites and activities as in the preceding Woodland periods. The most commonly occurring single-component site type is the Open habitation, prehistoric type (n = 65), with the other seven types only occurring in single digits each. Much like preceding periods, Late Woodland Open habitation, prehistoric sites occur mainly in lowland settings, and are fairly evenly spread across the floodplain, stream bench, and terrace landforms. These sites likely represent a mix of both macroband base camps and microband procurement camps that were occupied as part of a seasonal fusion-fission strategy. In some cases, such as at the River Road site, a site may represent both site types as temporally separated occupations.

REGION 9 SITES

There are 3,717 archaeological sites with prehistoric components in Region 9. A total of 2,071 sites in the PASS database did not possess diagnostic material and were not assigned to a temporal period (Table 5 shows a breakdown of the Region 9 sites by site type and landform; individual tables for each of the time periods are included in Appendix B). In addition, there are 301 Archaic-period sites that could not be specifically assigned to one of the Archaic sub-periods, and 129 Woodland-period sites with a similar issue.

Site types in Region 9 are almost evenly split between lowland and upland settings, with 49.1% of all sites with landform data located in lowland settings (n = 1,711) and 50.9% of sites with landform data in upland settings (n = 1,772). Two site types are only found in lowland settings in Region 9, including the Burial mound and Petroglyph/pictogram types; no site types are exclusive to upland settings. The Open habitation, prehistoric site type is the most common site type (n = 2,232), and occurs largely in lowland settings (66.0% of Open habitation, prehistoric sites).

The Lower Black's Eddy Site (36BU23) is a good example of a multi-component base camp in Region 9, with a Late/Terminal Archaic midden component, overlaid by less intense Early and Middle Woodland occupations. A dense Late Woodland component was formerly present, but largely destroyed by modern activities before archaeological excavations occurred. The site is thought to have been repeatedly occupied by small groups to exploit fish and nut resources in the Middle Delaware valley during the fall and early winter seasons, while also allowing access to sources of argillite for stone tool manufacturing (Robertson and Kingsley 1994).

Paleoindian

Within Region 9, there have been 38 sites identified with Paleoindian components, according to the PASS database. Twenty-seven sites with Paleoindian components also contain one or more components dating to later time periods. Only two Paleoindian sites in Region 9 with landform data included in the PASS database are found in uplands, including one Open habitation, prehistoric site and one site with no recorded site type in the PASS data. The Open habitation, prehistoric site type likely represents camp locations, and has been identified on floodplains (n = 1), stream benches (n = 1), terraces (n = 2), and upland flats (n = 1).

Early Archaic

The PASS database records 98 sites with Early Archaic components in Region 9. Eighty-three sites with Early Archaic sites also contain one or more components dating to other prehistoric time periods. Early Archaic sites with recorded landform data in the PASS database overwhelmingly occur in lowland settings, representing 79.0% of all such sites (n = 64). There are only 15 single-component Early Archaic sites in Region 9 categorized by site type as follows: Lithic reduction (n = 2); Open habitation, prehistoric (n = 8); Open prehistoric site, unknown function (n = 3); Unknown function open site greater than 20 m radius (n = 1); and one site with no recorded site type. The Open habitation, prehistoric site type, which likely represents transitional camp locations, occurs more often in upland settings in Region 9.

Table 5. Region 9 Site Types by Landform

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge /Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Burial Mound	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Cemetery	0	0	0	0	0	6	0	0	1	1	0	0	0	0	0	0	8
Earthwork	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Isolated Find	0	2	0	0	2	0	0	0	1	1	1	1	0	3	0	0	11
Lithic Reduction	0	15	1	0	14	18	3	1	1	35	8	6	1	10	6	25	144
Open Habitation, Prehistoric	0	365	6	15	612	475	342	38	40	16	10	16	17	174	20	86	2232
Open Prehistoric Site, Unknown Function	0	95	14	6	104	133	4	61	14	79	24	2	7	51	11	38	643
Other Specialized Aboriginal Site	0	2	1	0	2	6	2	0	0	1	0	0	0	1	1	7	23
Petroglyph/Pictograph	0	2	0	5	1	1	0	0	0	0	0	0	0	0	0	0	9
Quarry	0	4	0	0	21	4	15	3	0	0	0	3	0	3	1	9	63
Rock Shelter/Cave	0	4	0	0	6	6	1	44	1	1	5	1	1	0	1	2	73
Unknown Function Open Site Greater than 20 m Radius	0	4	1	0	1	2	0	0	1	1	3	0	1	4	1	2	21
Unknown Function Surface Scatter Less than 20 m Radius	0	3	2	0	5	5	1	0	1	8	5	1	1	8	1	4	45
Village	0	3	0	0	3	8	4	0	0	1	0	0	0	0	0	0	19
(blank)	1	48	2	3	116	55	10	13	17	15	12	7	3	58	4	61	425
Total	1	548	27	29	887	719	382	160	77	159	68	37	31	312	46	234	3717

Middle Archaic

The PASS database includes 263 sites with Middle Archaic components in Region 9. Middle Archaic sites in Region 9 are largely found in lowland physiographic settings. Terrace settings account for 29.0% of all Middle Archaic site locations in Region 9 (n = 74). There are only 23 single-component Middle Archaic sites in Region 9. There are 14 Open habitation, prehistoric sites and seven Open prehistoric site, unknown function sites, with one Isolated find site and one Lithic reduction site. Most of the single-component sites with landform data occur in lowland settings, although the overall low number of single-component sites makes extrapolation of landform preferences untenable for the Middle Archaic in Region 9.

Late Archaic

The PASS database includes 903 sites with Late Archaic components in Region 9, an increase in site frequency by a factor of 3.4. Late Archaic sites in Region 9 show a focus toward lowland physiographic settings, with 73.4% of all Late Archaic sites with landform data occurring in lowlands (n = 631).

Single-component Late Archaic site types that may represent the likeliest candidates for seasonal occupation sites, such as base camps and short-term resource extraction camps, are Open habitation, prehistoric and Open prehistoric site, unknown function. The Open habitation, prehistoric sites, which likely includes a number of base camps, are predominately found in lowlands, with 84.0% of all such sites with landform data in lowland settings. The Open prehistoric site, unknown function sites are more evenly distributed between lowland and upland settings, and may represent short-term resource extraction camps. Specialized sites associated with single-component Late Archaic occupations include a single hilltop Cemetery site, seven Lithic reduction sites, three Quarry sites, two Other specialized aboriginal sites, and three Unknown function open sites greater than 20 m radius. Additionally, there were 26 single-component Late Archaic sites with no site type recorded in the PASS database.

Terminal Archaic

The PASS database includes 433 sites with Terminal Archaic components in Region 9. There are 297 Terminal Archaic multi-component sites possessing either or both Late Archaic and Early Woodland components, representing 68.6% of the total population of Terminal Archaic sites. The fact that Terminal Archaic site components are commonly associated with preceding Late Archaic and subsequent Early Woodland components suggests a certain degree of group continuity within Region 9 between the Late Archaic and Early Woodland periods.

Terminal Archaic sites in Region 9 show a marked focus toward lowlands, with 77.0% of all Terminal Archaic sites located in that physiographic setting. Single-component Terminal Archaic site types that may represent the likeliest candidates for seasonal occupation sites, such as base camps and short-term resource extraction camps, are Open habitation, prehistoric (n = 47) and Open prehistoric site, unknown function (n = 4). The Open habitation, prehistoric sites are mainly found in lowland settings, while three out of the four Open prehistoric site, unknown function sites are in uplands; however, the small number of Open prehistoric site, unknown function sites makes drawing conclusions about their possible function in relation to landform position untenable. Two other single-component site types are present: Lithic reduction sites and Quarry sites. The one Lithic reduction site and both Quarry sites are located in lowland settings. Additionally, there are six single-component Terminal Archaic sites without an associated site type in the PASS database.

Early Woodland

The PASS database includes 185 sites with Early Woodland components in Region 9, showing a decline in site frequency that appears to have begun in the preceding Terminal Archaic period. There are 123 Early Woodland multi-component sites possessing either or both Terminal Archaic and Middle Woodland components, representing 66.5% of the total population of Early Woodland sites in Region 9. Early Woodland site components are well associated with preceding Terminal Archaic and subsequent Middle Woodland components, suggesting group continuity within Region 9 between the Terminal Archaic and Middle Woodland periods.

Early Woodland sites in Region 9 show a marked focus toward lowlands, with 74.7% of all Early Woodland sites with landform data occurring in that physiographic setting. Only eight single-component Early Woodland sites are located in Region 9, including one Lithic reduction site, one Open habitation, prehistoric site, three Open prehistoric site, unknown function sites, and three Rock shelter/cave sites. The scarcity of single-component Early Woodland sites in Region 9 does not allow for analysis of landform associations with site type with any level of confidence in resulting assertions.

Middle Woodland

The PASS database includes 195 sites with Middle Woodland components in Region 9, a slight increase in site frequency from the Early Woodland period. There are 132 Middle Woodland multi-component sites possessing either or both Early and Late Woodland components, representing 71.3% of the total population of Middle Woodland sites with recorded landform data. The fact that Middle Woodland site components are strongly associated with preceding Early Woodland and subsequent Late Woodland components suggests group continuity within Region 9 between the three Woodland periods.

Middle Woodland sites in Region 9 show a similar focus toward lowlands in comparison to the Early Woodland period, with 76.7% of Middle Woodland sites located in that physiographic setting. There are only 20 single-component Middle Woodland sites in Region 9, and they appear to represent similar site functions as in preceding prehistoric periods, such as seasonal camps (Open habitation, prehistoric; Open prehistoric site, unknown function; and Rock shelter/cave) or a specialized function (Lithic reduction and Other aboriginal specialized site). Four single-component Middle Woodland sites did not possess a recorded site type. The scarcity of single-component Middle Woodland sites in Region 9 does not allow for analysis of landform associations with site type with any level of confidence in resulting assertions.

Late Woodland

The PASS data for Region 9 includes 538 sites with Late Woodland components, a dramatic increase in site frequency from the preceding Woodland periods. There are 107 Late Woodland multi-component sites possessing Middle Woodland components, representing 19.9% of the total population of Late Woodland sites. The near-tripling in frequency of site occurrence from the Middle Woodland to the Late Woodland may obscure the relationship between Middle Woodland and Late Woodland groups.

Late Woodland sites in Region 9 show a general focus toward lowlands, with 74.9% of all Late Woodland sites occurring in that topographic setting. Floodplain and terrace settings account for most Late Woodland site locations in lowlands, followed closely by stream benches. Village sites are perhaps the defining site type for the Late Woodland. There are 12 single-component Late Woodland villages in the PASS database, with 8 occurring in lowland settings, 3 villages in upland settings, and 1 village without a recorded landform. There does not appear to be a specific landform selected for village locations more frequently than others during the Late Woodland, with villages occurring on five different landform types. Single-component Late Woodland site types that may represent likely candidates for seasonal occupation sites, such as base camps and short-term resource extraction camps, are Open habitation, prehistoric; Open prehistoric site, unknown function; and Rock shelter/cave. Both the Open habitation, prehistoric site type and the Open prehistoric site, unknown function site type occur twice as frequently in lowland settings than in upland settings. There are six Rock shelter/cave sites, five of which are in upland settings; these sites also likely represent seasonal camps. Three Late Woodland cemeteries occur in Region 9, all in terrace settings.

REGION 10 SITES

There are only 23 archaeological sites with prehistoric components currently recorded in the PASS database in Region 10, primarily due to the fact that Region 10 consists of the City of Philadelphia and its suburbs and is heavily developed, discouraging the archaeological examination of large areas (Table 6 shows a breakdown of the Region 10 sites by site type and landform; individual tables for each of the time periods are included in Appendix B). Kratzner et al. (2008:5) note that while the history of Philadelphia's development since its founding in 1682 has likely resulted in the obliteration of much of the pre-contact archaeological record, the area encompassing Region 10 would likely have been very attractive to prehistoric groups as part of the Coastal Plain, with a variety of estuarine, terrestrial, and riverine resources.

Table 6. Region 10 Site Types by Landform

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge /Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Burial Mound	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cemetery	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Earthwork	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Isolated Find	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lithic Reduction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Open Habitation, Prehistoric	0	0	1	0	0	4	0	0	0	0	0	0	0	0	0	0	5
Open Prehistoric Site, Unknown Function	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	1	4
Other Specialized Aboriginal Site	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Rock Shelter/Cave	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unknown Function Open Site Greater than 20 m Radius	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unknown Function Surface Scatter Less than 20 m Radius	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Village	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(blank)	0	4	0	0	0	7	0	0	0	0	0	1	0	0	0	1	13
Total	0	4	1	0	0	15	0	0	0	0	0	1	0	0	0	2	23

The stabilization of the environment during the Middle Holocene as North America emerged from the last ice age would have included a slowing of sea level rise and maturation of estuary environments. Perhaps not coincidentally, Late Archaic components represent the earliest dated occupations at archaeological sites in Region 10. Woodland period occupation in Region 10 is not well-documented, but Kratzner et al. (2008:6) observe that the pre-contact environment of Region 10 would have supported the same general Woodland settlement patterns as in other similar areas of eastern Pennsylvania and Delaware. The presence of two Lenape villages at contact in Region 10, Passyunk and Shackamaxon, suggests that earlier village sites were also present in Region 10. A total of 10 sites in the PASS database did not possess diagnostic material and were not assigned to a temporal period.

Archaeological projects associated with improvements to Interstate 95 (I-95) in Philadelphia are ongoing as of this writing, and several newly discovered prehistoric sites are in the process of documentation; however, these sites are not currently in the PASS database and are not included in this analysis. Most of the newly identified sites from the I-95 project have been documented on terraces, and are for the most part multi-component sites with components ranging from the Middle Archaic to the Late Woodland.

Site types in Region 10 are primarily found in lowland settings, with one site recorded in the upland ridgetop setting. Most sites with landform data were recorded on terraces (n = 15). This distribution is likely attributable to the small size of Region 10 and lack of landform diversity within it. However, the very small sample size of prehistoric sites within Region 10 makes any assessment of the distribution of sites across landforms untenable.

There are no Paleoindian, Early Archaic, or Early Woodland sites recorded in the PASS database for Region 10. The PASS database records a single site with a Middle Archaic component, 36BU0344, which is an Open prehistoric site, unknown function site located on a terrace. The PASS database records a single multi-component site with a Middle Woodland component, 36DE0034. No other sites in Region 10 are recorded with Middle Woodland components.

Late Archaic

The PASS database includes 9 sites with Late Archaic components in Region 10, six of which are multi-component sites and two of which are single-component sites with no associated site type in the PASS database. The remaining single-component site, 36PH0130, is recorded as belonging to the Other specialized aboriginal site type and is located on a terrace. Late Archaic sites in Region 10 are often found as one component out of many on multi-component sites, and specific data on Late Archaic lifeways is lacking. Late Archaic groups in Region 10 were presumably not very different from groups in adjacent areas, however, and likely shared most, if not all, general characteristics of the period. The Late Archaic period in eastern Pennsylvania is thought to represent a population increase in the region, with an accompanying diversification in exploitation of food resources (Harris et al. 2014c). One multi-component site with a significant Late Archaic/Terminal Archaic occupation is the Bartram's Site (36PH14), which produced a variety of Late Archaic materials and features.

Terminal Archaic

The PASS database includes four sites with Terminal Archaic components in Region 10, including two Open Habitation, Prehistoric sites and two multi-component sites with Terminal Archaic components. All four sites are located on terraces. Terminal Archaic groups likely practiced similar cultural behaviors as with the preceding Late Archaic period, with some changes in technology; markedly, the use of steatite vessels and the beginnings of ceramic vessel production. Additionally, the occurrence of FCR at Terminal Archaic sites appears to increase (Harris et al. 2014c:19).

Late Woodland

The PASS data for Region 10 includes 5 sites with Late Woodland components, four of which are multi-component sites. The one single-component Late Woodland site is 36BU0346, classified as an Open prehistoric site, unknown function and located on a terrace. The Bartram's Site (36PH14), a multi-component site, had a significant Late Woodland occupation with features containing a variety of pottery styles.

3
DATA QUALITY – REGIONS 7, 8, 9, AND 10

INTRODUCTION

PASS forms have been used by submitters to record archaeological site data for more than 65 years. When PASS forms are accurately filled out, they offer the PHMC vital information regarding location and artifact data. Over the past few decades PHMC has been working diligently to get the PASS form data into its CRGIS database, a map-based inventory of the historic and archaeological sites and surveys currently stored in the files of the Bureau for Historic Preservation (BHP). The CRGIS database is designed to include all information on the PASS forms, with the goal of obtaining as much accurate information as possible about Pennsylvania's archaeological and historic sites. Using roughly 23,000 completed PASS forms, PHMC has managed to accurately enter almost all known archaeological sites into the CRGIS database. The CRGIS database has become PHMC's primary tool when attempting to accurately record and map Pennsylvania's historic and prehistoric past.

In order to establish the validity of the data used for the predictive model set project, the CRGIS database and PASS form data were compared for a sample of Pennsylvania's 18,232 prehistoric archaeological sites. Archaeological site forms were analyzed and compared with the data included in the CRGIS database. Site forms from all of Pennsylvania's 67 counties were considered and a 10% random sample was selected from each county. The following conclusions and data are the results of the 10% sample for the counties within Regions 7, 8, 9, and 10.

METHODS

Within Regions 7, 8, 9, and 10, PASS forms and CRGIS data were examined for 704 prehistoric archaeological sites. The following section presents the results of the analysis by region. Location accuracy, artifact data quality, and form completeness were rated for each of the selected sites using information from the PASS forms and CRGIS database. Ratings were assigned numerical values to facilitate comparison between the two data sources and across regions. Table 7 lists the criteria used to derive ratings for each category of data.

Location data were analyzed by manually comparing mapped locations within the CRGIS with maps provided in the original PASS forms. Artifact information was also manually compared between the PASS forms and the CRGIS database. Discrepancies between the two data sets were categorized using the ranking outlined in Table 7.

Table 7 - Rating Criteria for Site Data

Rating	Criterion
	Location Accuracy, PASS Form
1	*No location information.* No location data are present on the site form.
2	*Coordinates only.* Location is documented only by coordinates with no physical description or landmarks.
3	*Poor accuracy.* The only location information is a hand-drawn map with low detail.
4	*Medium accuracy.* The form contains a USGS map with the site location indicated.
5	*High accuracy* The form contains a detailed map with reference points or an aerial photo and the site location is assumed to be accurate.
	How Well Location is Reflected in CRGIS
1	*Not mapped.* The site has not been mapped into the CRGIS system.
2	*Mapped, > 500 m.* The site location is mapped, but is more than 500 m away from the location indicated on the PASS form. Note that in some cases this reflects corrections to the location data in CRGIS, resulting in *increased* accuracy.
3	*Mapped, 250–500 m.* The site location is mapped, but is between 250 and 500 m away from the location indicated on the PASS form (see note above re: accuracy).
4	*Mapped, < 250m.* The site location is mapped less than 250 m away from the PASS form location.
5	*Mapped accurately.* The site location in CRGIS matches the location on the PASS form.
	Artifact Data Quality, PASS Form
1	*No artifacts.* The PASS form contains no artifact information, either because no artifacts were found or because they were not recorded.
2	*Artifacts poorly represented.* No artifacts are listed on the PASS form, but a note indicating that artifacts were found is included indicating that artifacts were found but not recorded.
3	*Poor quality recording.* The PASS form contains poorly hand-drawn artifacts and/or mislabeled items.
4	*Moderate recording.* Few artifacts are listed on the PASS form or only a small selection were drawn; the location of the collection is not indicated.
5	*Good recording.* All artifacts are listed on the form, which also includes high-quality hand-drawn images or photographs; the location of the collection is usually indicated.
	How Well Artifacts are Reflected in CRGIS
1	*No artifacts.* The CRGIS database does not include any artifacts.
2	*Less artifacts.* Fewer artifacts than appear on the PASS form are included in the CRGIS database.
3	*Moderate quality.* Artifacts are listed in the CRGIS database, but not with any detail.
4	*Higher quality.* The CRGIS database contains more artifacts than are listed on the PASS form.
5	*Accurate recording.* Artifacts listed in the CRGIS database match those listed on the PASS form.
	PASS Form Completeness
1	*Name and/or location.* Only site name and/or location are included on the PASS form.
2	*< 25% completed.* The PASS form contains more than just name and location, but is missing at least 25% of data.
3	*25–75% completed.* The PASS form is mostly filled out and contains artifact and location data.
4	*> 75% completed.* The PASS form is filled out completely and contains all required information.
	PASS Form Type
1	*1950–1980 version.* This form has limited room for data; usually only location information and material culture information was collected.
2	*1981–2007 version.* This form has more space for documentation and includes a requirement for sketched images of artifacts.
3	*2008–present version.* This form is several pages in length; it requires artifacts to be categorized and location information to be detailed on attached maps.

REGION 7

Within Region 7, PASS forms and CRGIS data were examined for 96 sites.

Location Data

Of the 96 sites in the Region 7 sample, the majority (54%) are mapped with medium to high accuracy, that is, on detailed map or USGS topographic quadrangles. The remaining 46% of sites are poorly mapped or provide little locational information (Figure 8). By comparison, 93% of the same site sample has accurately mapped locations in the CRGIS database, and another 4% are mapped within 250 m from the location indicated on the PASS forms (Figure 9). Just 2% of sites in the sample remained unmapped, suggesting an increase in mapping accuracy in CRGIS as compared to the PASS forms.

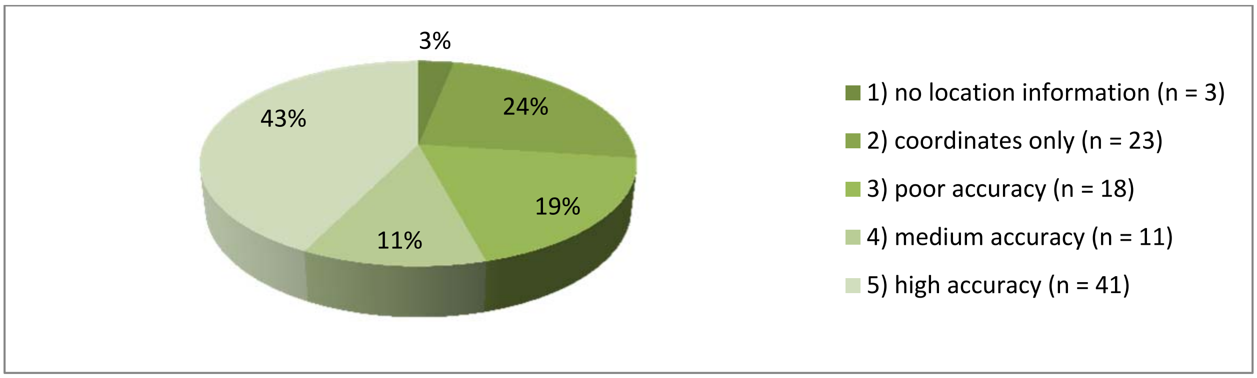


Figure 8 - Quality of location information on PASS forms within Region 7.

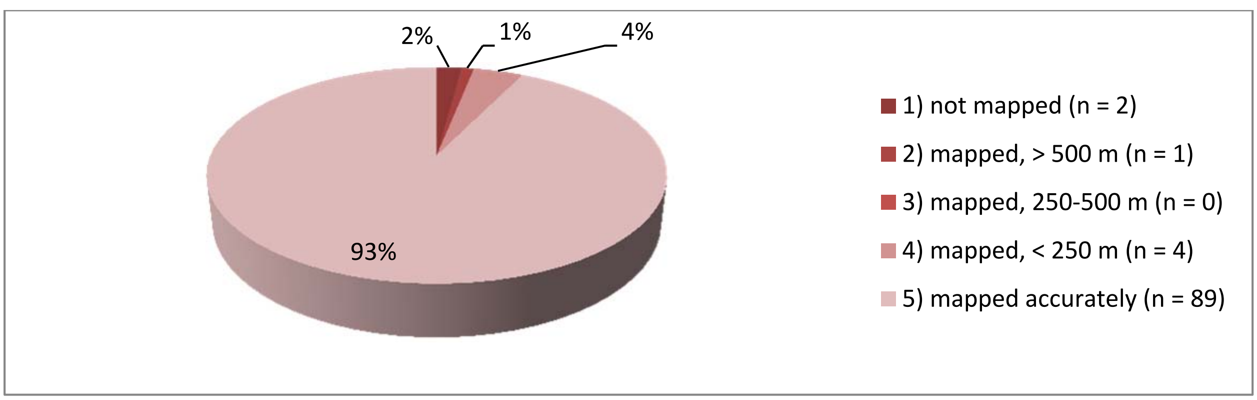


Figure 9 - Quality of location information reflected in CRGIS within Region 7.

Artifact Data

Nearly half (49%) of the site sample in Region 7 has good or moderate artifact descriptions on the PASS forms, while 9% have poor quality data and 40% have no artifact data at all (Figure 10). By comparison, a full 83% of the sites in the Region 7 site sample have moderate to high quality artifact date, while only 4% have poor quality artifact data and 13% have not data (Figure 11), suggesting that data quality was improved in the transition from PASS forms to CRGIS.

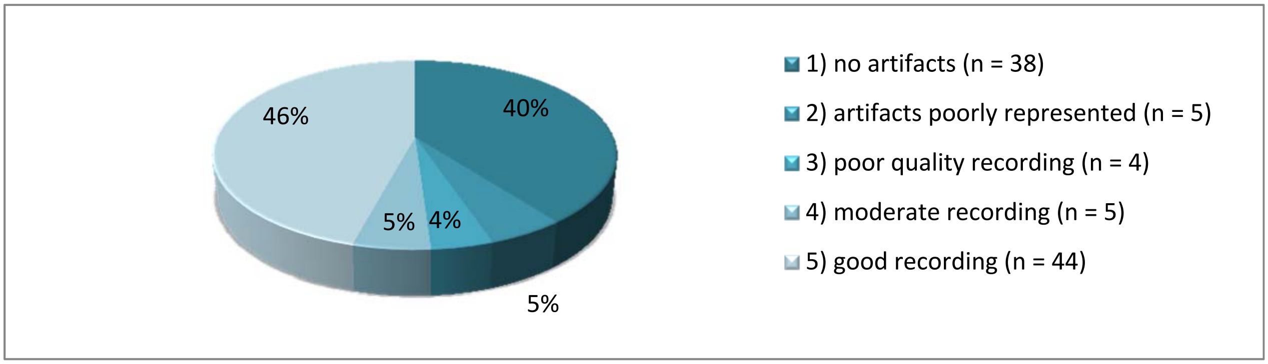


Figure 10 - Original artifact data recorded on PASS forms for Region 7.

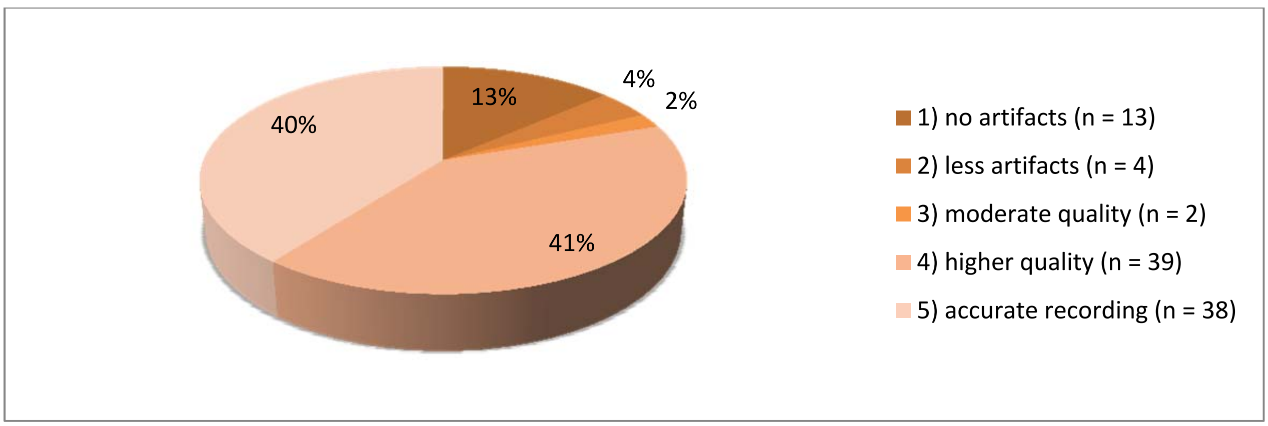


Figure 11 - Artifact data reflected in the CRGIS database for Region 7.

PASS Form Types and Completeness

More than half (61%) of the PASS forms in the site sample from Region 9/10 are up to or greater than 75% complete (Figure 12). The remaining 39% of PASS forms in the site sample contain limited data. Almost all (97% of the site sample for Region 9/10 is recorded on old version or middle version PASS forms, while only 3% are recorded on the newer version of the form that includes detailed artifact information (Figure 13). This suggests that for Region 9/10, the most reliable site information derived from the PASS forms is likely to be locational rather than artifact data.

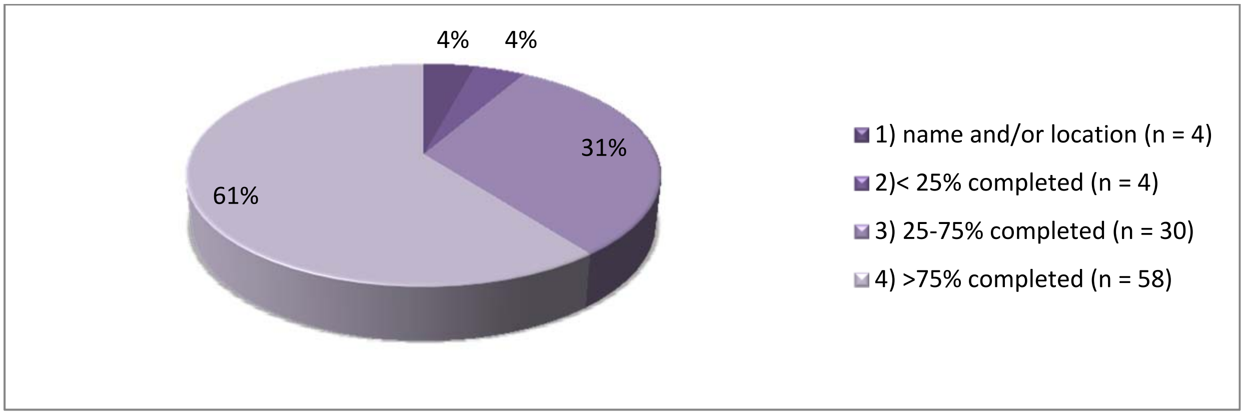


Figure 12 - Completeness of PASS form information in Region 7.

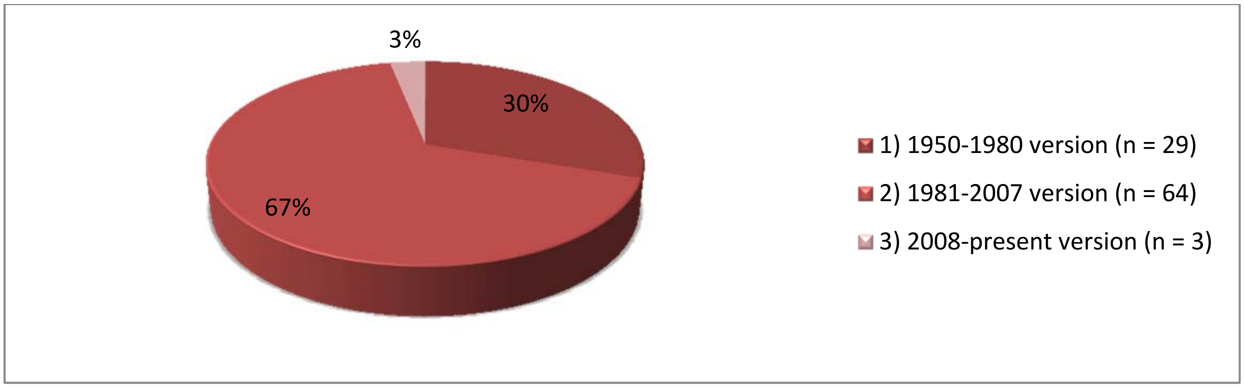


Figure 13 - Distribution of PASS form types in Region 7.

REGION 8

Within Region 8, PASS forms and CRGIS data were examined for 248 sites.

Location Data

Of the 248 sites in the Region 8 sample, 35% are mapped on USGS maps or contain highly detailed maps on the PASS forms. The remaining 65% of forms contain no location data, are only referenced by coordinates, or contain unreliable hand-drawn maps (Figure 14). Within the CRGIS database, almost all (92%) of the site locations match the mapping in the PASS forms. Seventeen sites (7%) were mapped within 250 m of the locations indicated on the PASS forms, and just 2 sites (1%) were not mapped (Figure 15).

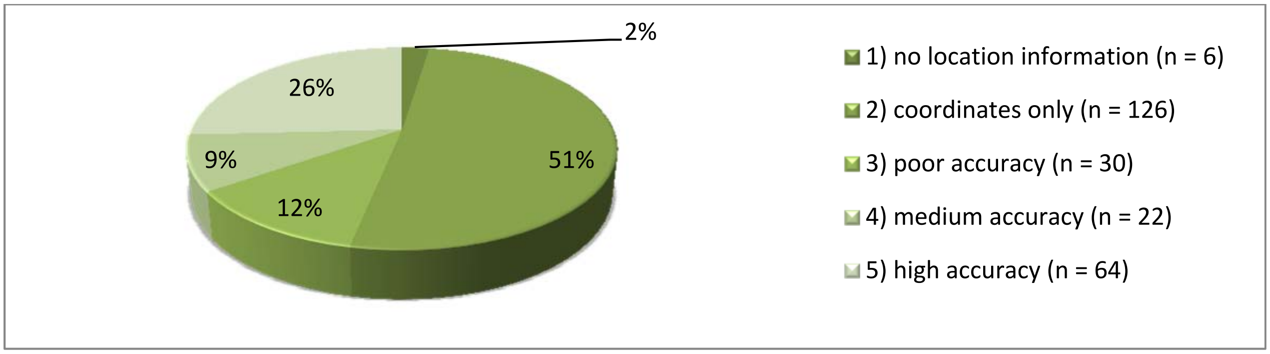


Figure 14 - Quality of location information on PASS forms within Region 8.

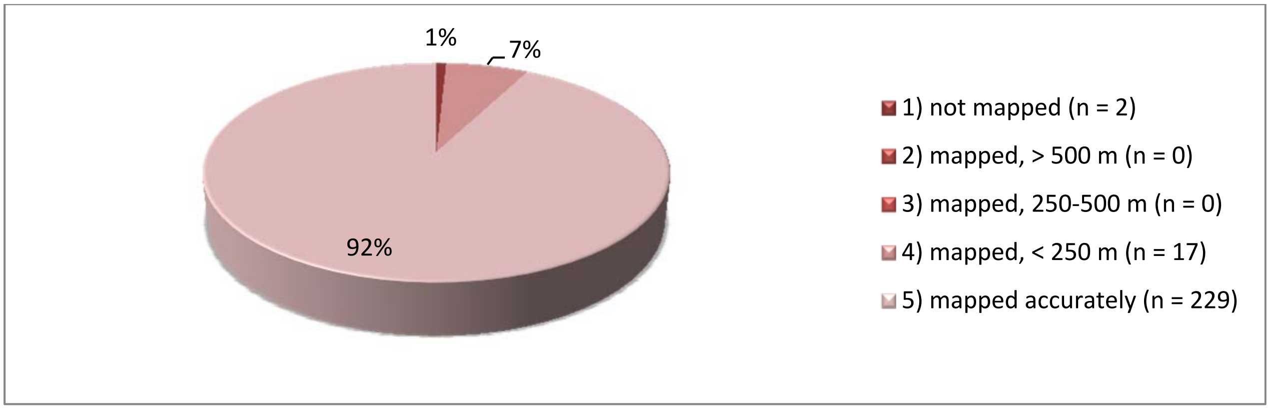


Figure 15 - Quality of location information reflected in CRGIS within Region 8.

Artifact Data

Roughly equal numbers of sites in the Region 8 sample have good artifact data (44%) and no artifact data (43%) on the PASS forms (Figure 16). In between those two extremes are 13% of sites with poor to moderate quality artifact data. The transition to CRGIS appears to have improved the artifacts data quality appreciably, with 76% of sites having high quality or accurate artifact data, while 7% have moderate artifact data quality and 17% have no artifact data (Figure 17).

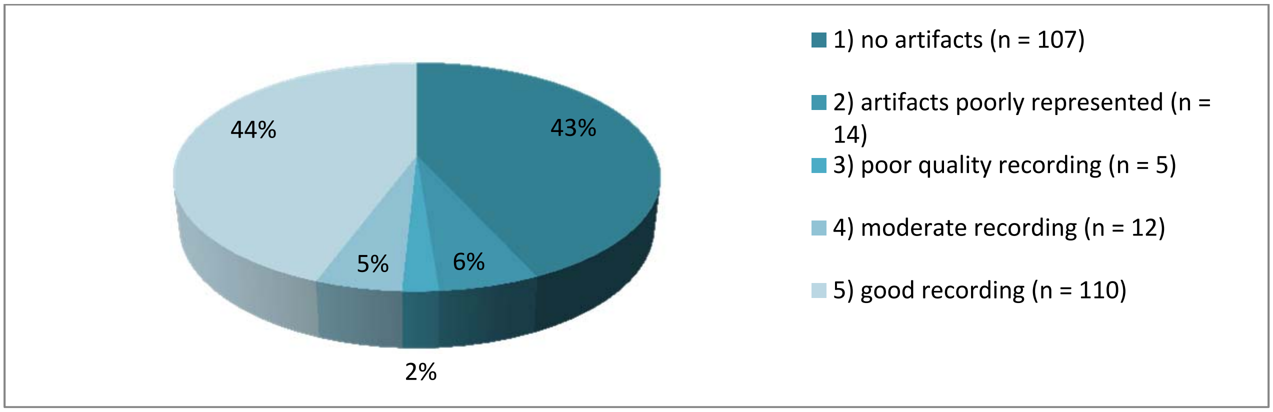


Figure 16 - Original artifact data recorded on PASS forms for Region 8.

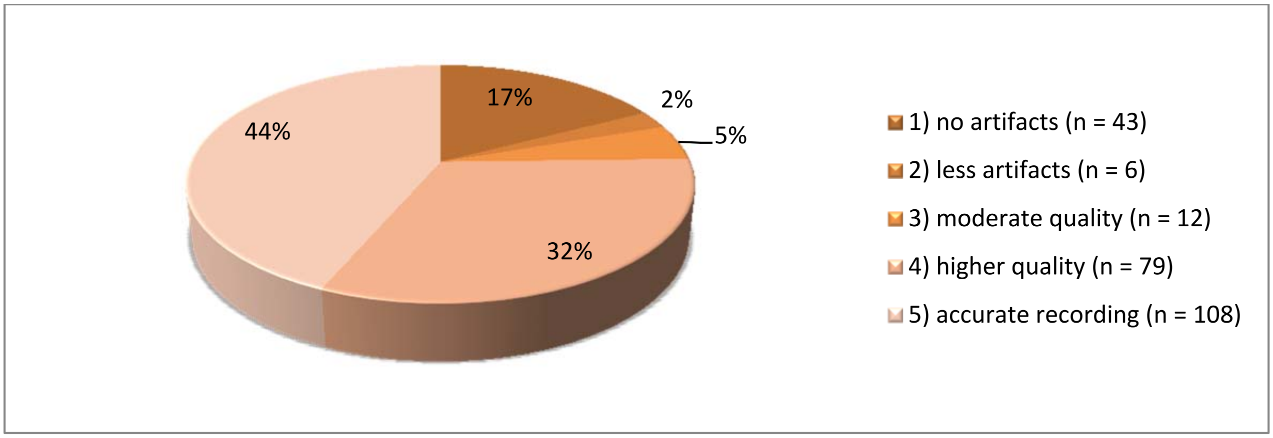


Figure 17 - Artifact data reflected in the CRGIS database for Region 8.

PASS Form Types and Completeness

Of the 248 total sites sampled within Region 8, nearly half (45%) are at least 75% complete. The remaining 55% of the forms contain limited data (Figure 18). The PASS form types for Region 9/10 are almost all (97%) either older or middle versions, with just 3% on new forms with detailed artifact data (Figure 19).

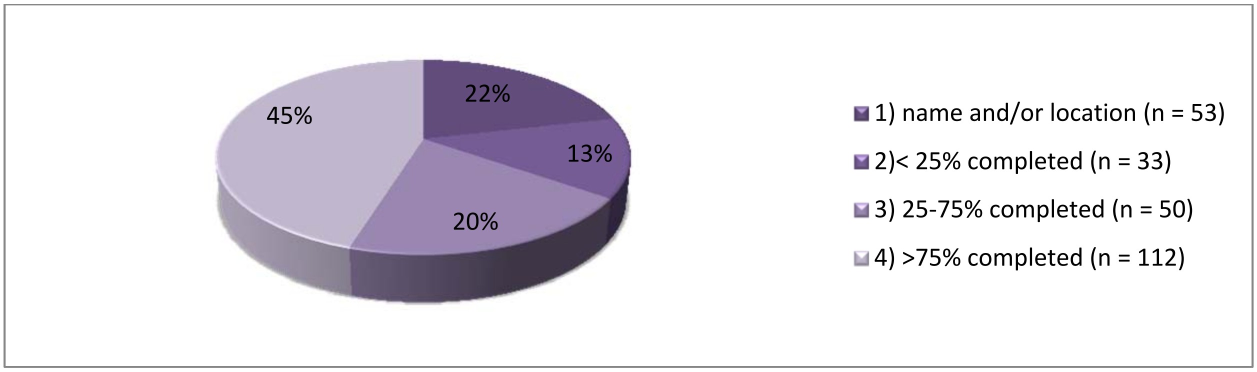


Figure 18 - Completeness of PASS form information in Region 8.

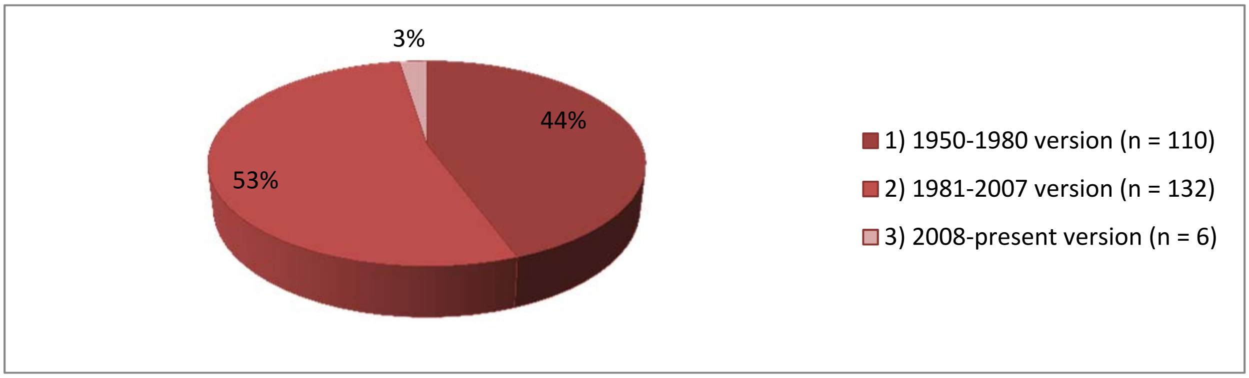


Figure 19 - Distribution of PASS form types in Region 8.

REGION 9/10

For the purposes of analysis, Regions 9 and 10 were combined into one data set. Within these two regions, PASS forms and CRGIS data were examined for a total of 360 sites.

Location Data

Similar to Region 8, the results for Region 9/10 are starkly different for the PASS forms and the CRGIS data. Of the 360 sites in the Region 9/10 sample, just 32% are mapped on USGS maps or contain highly detailed maps on the PASS forms (Figure 20). Nearly one-fifth of the PASS forms (19%) contained unreliable hand-drawn maps, and just about half of the forms (49%) had no locational information at all. By contrast, a full 95% of sites are mapped accurately in the CRGIS database and another 4% are mapped within 250 m of the location provided on the PASS form (Figure 21).

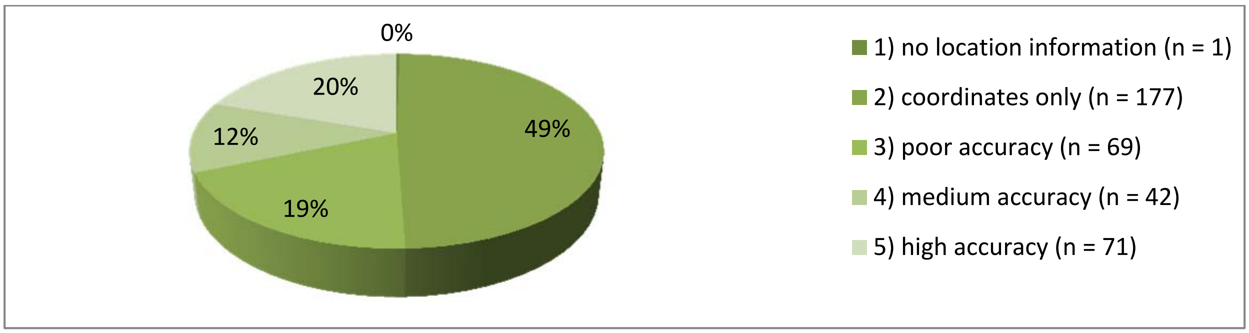


Figure 20 - Quality of location information on PASS forms within Region 9/10.

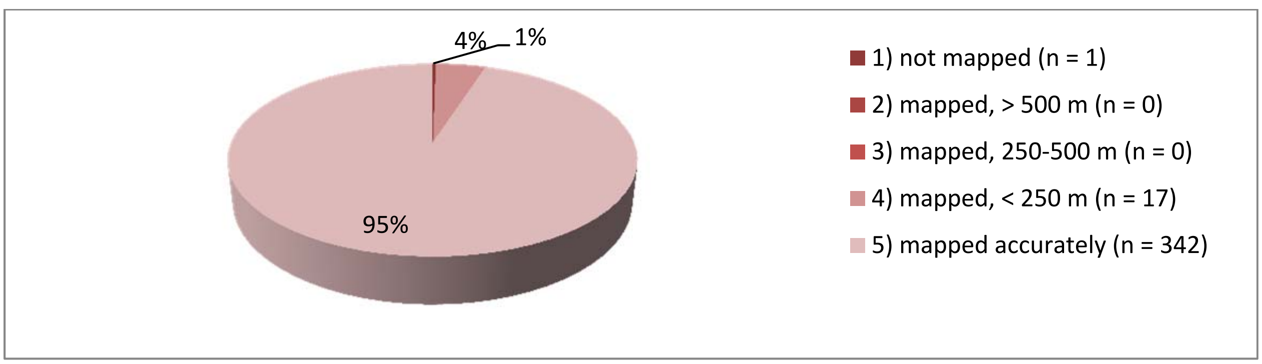


Figure 21 - Quality of location information reflected in CRGIS within Region9/10.

Artifact Data

More than half (57%) of the site sample in Region 9/10 has good or moderate artifact descriptions on the PASS forms, while 18% have poor quality data and 25% have not data at all (Figure 22). By comparison, a full 77% of the sites in the Region 9/10 site sample have moderate to high quality artifact data in the CRGIS database, while only 4% have poor quality data and 19% have no data at all (Figure 23). These results suggest that, overall, artifact data quality in Region 9/10 was improved in the transition from PASS forms to CRGIS.

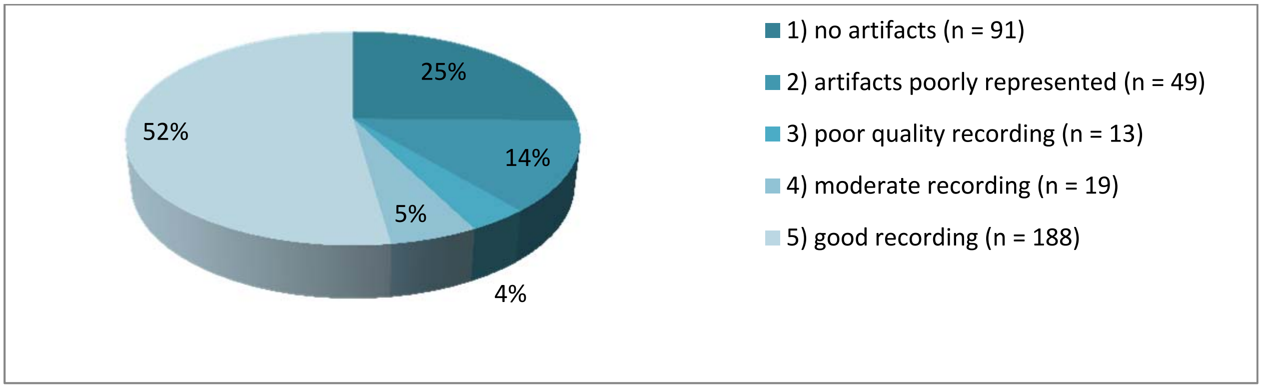


Figure 22 - Original artifact data recorded on PASS forms for Region 9/10.

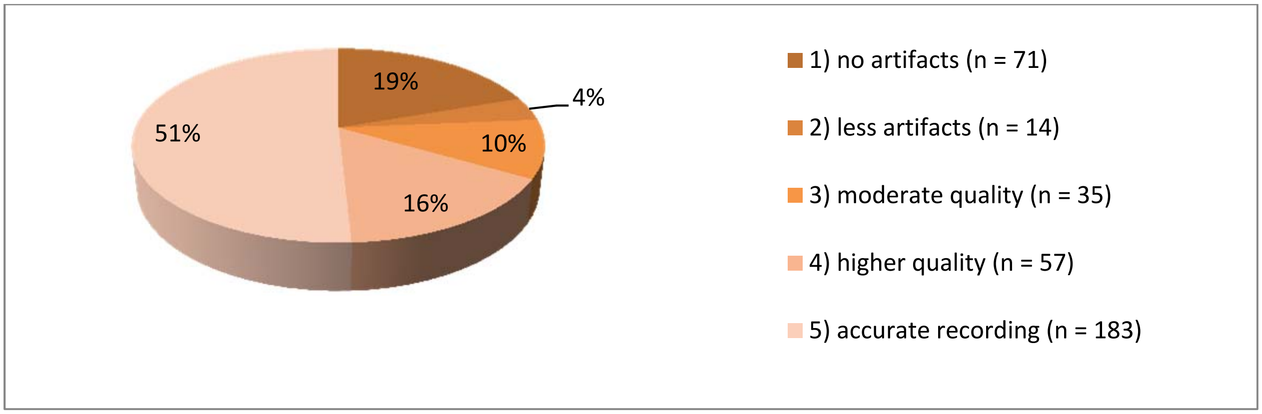


Figure 23 - Artifact data reflected in the CRGIS database for Region 9/10.

PASS Form Types and Completeness

Of the 360 total sites sampled within Region 9/10, almost half (46%) are at least 75% complete. The remaining 54% of the forms contain limited data (Figure 24). The PASS form types for Region 9/10 are overwhelmingly (80%) middle versions, with just 19% on older version forms and 1% on current version forms (Figure 25). The large number of middle version forms, which are often filled out completely or contain very little missing data, probably accounts for the overall completeness of the site sample.

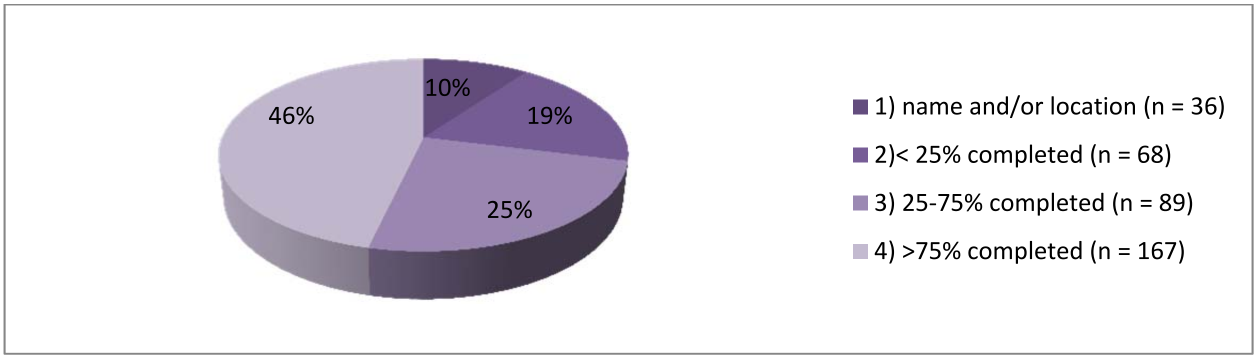


Figure 24 - Completeness of PASS form information in Region 9/10.

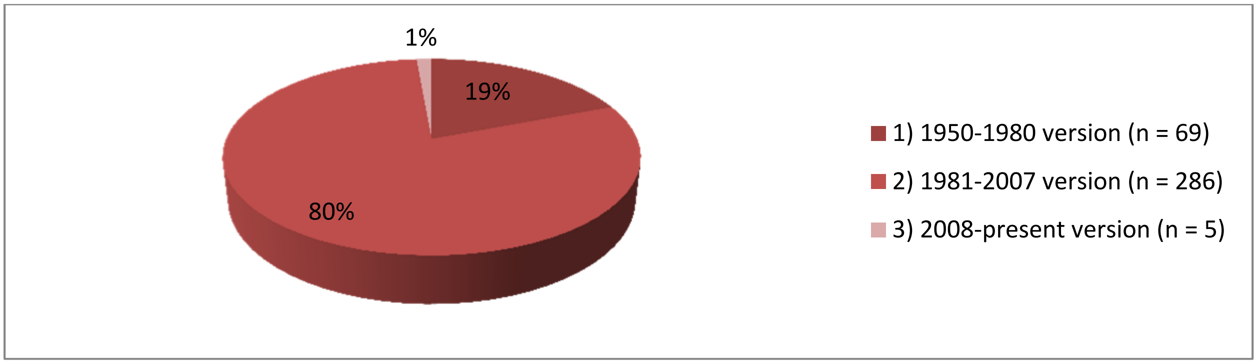


Figure 25 - Distribution of PASS form types in Region 9/10.

CONCLUSIONS

The sample for Regions 7, 8, and 9/10 includes a total of 704 prehistoric archeological sites. Overall, the analysis shows that the data derived from the CRGIS database are at least as complete and accurate as the data included in the original PASS forms, and in some cases, more so. Of the 704 sites in the sample, a total of 5 sites are currently still missing locational information, as compared with the initial 10 sites that contained no location information on the PASS forms. Errors and missing information on the PASS forms were addressed in the transition to CRGIS, and sites that had no mapping were located and plotted. In some cases, CRGIS staffers navigated to the site locations using non-map information provided on the PASS forms, such as landmarks, creeks, road names, or other locational references. Mapping locations in CRGIS diverged very little from locations provided on the PASS forms, reflecting the accurate transcription of data: of the 704 sites in the sample, only 5% (n = 38) sites were mapped 250 m or more from the locations shown on the PASS forms.

Of the 704 PASS forms examined for Regions 7, 8, and 9/10, 49% (n = 342) contain good artifact data, while 33% (n = 236) contain no artifact data, with both categories accounting for 82% of the total site sample. This suggests that most PASS form submitters are recording artifact data thoroughly or not at all. Most of the forms with no artifact data were of the older version that did not provide space for artifact descriptions. Artifact data that was provided on the PASS forms was, overall, accurately transferred into the CRGIS database: artifact information in the CRGIS database matched the information in the PASS form for 47% (n = 329) of the 704 sites. Further, the quality of artifact data was improved upon in the CRGIS data for 25% (n = 175) of the 704 sites. This reflects a successful effort by CRGIS staffers to track down missing artifact information.

PASS forms have changed over time and the current version provides for more thorough recordation of site locations and artifact data. Most of the sites considered for this analysis (68%; n = 482) were recorded on the "middle" version of the PASS form and 48% (n = 337) were considered at least 75% complete. These forms do not include as much information as the newer version and the data in the CRGIS is therefore limited.

4
MODEL METHODOLOGY – REGIONS 7, 8, 9, AND 10

The general approach to modeling Regions 7, 8, 9, and 10 followed the same process used for the previous regions covered in the Task 4 and Task 5 reports. The methodology is documented in detail in the Task 3 report (Harris 2014), with adaptations documented in the Task 4 and 5 reports (Harris et al. 2014a, 2014b). Broadly, the steps leading to the final sensitivity model are as follows:

- delineation of study areas;
- preparation of PASS data;
- creation of environmental variables;
- extraction of variables for each known site and 500,000 background samples;
- statistical comparison of the variables at sites and various background samples;
- selection of variables that are able to discriminate sites from the background;
- parameterization, creation, and validation of statistical models (Logistic Regression, Multivariate Adaptive Regression Splines, and Random Forest);
- application of the statistical models to create study area wide predictions;
- collection of predicted probability distributions from sites and the entire study area background;
- establishment of cut-off values to create high, moderate, and low classes; and
- mosaicking of the selected models into a final assessment of prehistoric site location sensitivity.

The methodology used in this report does not differ in any significant way from the methods used and discussed in the previous reports. There were a number of changes made to the model building code for this task, but these were only done to add efficiency and repeatability to the modeling process. Therefore, the changes are not addressed here as they have no impact on the resulting models aside from creating them faster and with less manual processing.

5
MODEL VALIDATION – REGIONS 7, 8, 9, AND 10

The total number of known archaeological sites within each of the 66 subareas range from as few as 7 sites to as many as 816 sites. The density, measured as the number of sites per square mile, ranges from a low of 0.03 to a high of 6.09, with riverine areas having a higher site density on average (2.53) than upland areas (0.519). With this high variability in the density of known site locations, both the suite of statistical models, Logistic Regression (LR), Multivariate Adaptive Regression Splines (MARS), and randomForest (RF) and the proportionally weighted model (Model 2) were used to try to find the best model to capture the available data. The judgmentally weighted models (Model 1), referenced in the previous task reports, were not used in Regions 7, 8, and 9/10 due to there being at least a single site recorded within each subarea. Proportionally weighted models (Model 2) were created for each subarea within Regions 7, 8, and 9/10. This type of model was initially created to serve as a low-assumption model that could be applied to areas where the number of known sites was low (typically less than 20 sites) or unrepresentative. The theoretical basis and technical components of these models are covered in detail in the Task 3 and Task 4 reports (Harris 2014; Harris et al. 2014a). However, being that it takes little effort to create the model for all subareas once the data are correctly formatted and the code is in place, this model type was created for the entirety of Region 7, 8, and 9/10. In the end, none of the proportionally weighted candidate models were chosen to represent any of the subareas within Regions 7, 8, and 9/10, but final versions of the models were created and will be part of this task’s deliverable.

This model validation section is organized by model type. For each of the 66 subareas for which models were created, a single model was selected as being the best balance between model fit, predictive ability, and the distribution of sensitivity values. The metrics used to assess the most representative model were the same as those used for the other six regions: Root Mean Squared Error (RMSE), Area Under the Curve (AUC), Kvamme Gain (KG), and Kappa (K) at a 0.5 threshold, with the thresholds calculated empirically from final sensitivity raster layers. Each of these metrics was presented and discussed in the Task 4 report (Harris et al. 2014a). Table 8 lists the model type chosen to best represent each subarea. The text that follows will be organized by these model types, beginning with Model 2, followed by LR, MARS, and finally RF.

Table 8 - Selected Model Type for Each Subarea

Region	Zone	Subarea	Model Type	Region	Zone	Subarea	Model Type
7	all	riverine section 1	RF	9/10	all	riverine section 1	MARS
		riverine section 2	RF			riverine section 2	MARS
		riverine section 3	RF			riverine section 3	RF
		riverine section 4	RF			riverine section 4	RF
		riverine section 5	RF			riverine section 5	MARS
		riverine section 6	MARS			riverine section 6	RF
		riverine section 7	RF			riverine section 7	MARS
		riverine section 8	MARS			riverine section 8	RF
		riverine section 9	RF			riverine section 9	RF
		upland section 1	LR			riverine section 10	RF
		upland section 2	RF			riverine section 11	RF
		upland section 3	RF			riverine section 12	MARS
		upland section 4	RF			riverine section 13	RF
		upland section 5	RF			riverine section 14	RF
		upland section 6	MARS			riverine section 15	RF
		upland section 7	RF			upland section 1	RF
		upland section 8	MARS			upland section 2	RF
		upland section 9	RF			upland section 3	RF
8	all	riverine section 1	RF			upland section 4	RF
		riverine section 2	MARS			upland section 5	RF
		riverine section 3	MARS			upland section 6	RF
		riverine section 4	RF			upland section 7	RF
		riverine section 5	RF			upland section 8	RF
		riverine section 6	RF			upland section 9	RF
		riverine section 7	RF			upland section 10	RF
		riverine section 8	RF			upland section 11	RF
		riverine section 9	RF			upland section 12	RF
		upland section 1	RF			upland section 13	RF
		upland section 2	RF			upland section 14	RF
		upland section 3	RF			upland section 15	RF
		upland section 4	RF				
		upland section 5	RF				
		upland section 6	RF				
		upland section 7	RF				
		upland section 8	RF				
		upland section 9	RF				

PREDICTOR VARIABLES

As with the previous models in Task 4 and Task 5, a large number of environmental variables was created and then pared down based on their ability to discriminate site locations from background locations. The ability to discriminate was judged based on the Kolmogorov-Smirnov (K-S) test and Mann-Whitney (MW) U test statistics. Both are non-parametric tests that measure the dissimilarity of two distributions, in this case environmental variables measured at known site locations and those randomly picked from the background. There are specific differences in the tests that contribute information valuable to understanding the way in which the two samples are different. Within each region modeled, each of the 93 variables (including a purely random noise variable) was tested against 100 random samples of 50,000 background values (the variables tested are listed in Appendix C). The results were tabulated and the test statistics and p-values were compared to identify those variables that were most discriminant, as well as detect indications of how site location patterns were expressed within the variable pool. From the list of all variables, those with a K-S D statistic that is higher than the median were selected; typically this was about 35 variables. From this group, the variables that measured the same aspect of the landscape but on a different scale (e.g., range in elevation within 10 cells or 16 cells) were pared down so that only the scale with the highest D statistic was left. Finally, variables that were very highly correlated were removed, resulting in the final selection of predictors, which averaged 17 per subarea.

The inclusion of the soils variables as factors required the models to consider many additional dummy variables. A described in Chapter 4, for each factor variable included in these models, a series of presence/absence variables, referred to as dummy variables, had to be created for each level of the factor. A variable of soil drainage requires the creation of a new dummy variable for each category (e.g., well-drained, moderately well-drained, poorly-drained, etc…). If the drainage variable contains seven different levels (categories) it will be represented within the model as seven separate dummy variables instead of just one. Because of this, if a model includes one of the three soil variables, the total number of soil drainage variables used within each model will include the dummy variables and therefore will be greater than the number of selected variables. As shown in Table 9, Table 10, and Table 11 an additional field is added to show the total number of variables after the inclusion of the dummy variables. The tables included in Appendix D show the variables that were selected to represent each subarea, the K-S D statistic, the MW U statistic, with associated p-values, and the statistics for the variable that represents random noise, for a basis of comparison.

Each of the variables tabulated in Table 9, Table 10, and Table 11 and detailed in the tables in Appendix D was selected to represent the most discriminant version of the particular part of the landscape that it measures. It is understood that many of these variables will be correlated naturally or by the design of what they measure. The previously discussed steps were taken to eliminate highly correlated or redundant variables, but it cannot be assumed that the remaining variables are truly independent. These are simply the facts of dealing with environmentally based variables. However, the LR, MARS, and RF statistical methods have means of dealing with correlated variables and variables that do not contribute to the success of the prediction. For LR, a backwards stepwise routine removes noncontributing variables based on their reduction of the Akaike Information Criterion (AIC) metric. For the MARS algorithm, the backwards elimination routine minimizes the effects of variables that do little to reduce the generalized cross-validation (GCV) metric. Additionally, the *nprune* parameter of the MARS algorithm controls the maximum number of terms within the model. This parameter is optimized to reduce misclassification through 10-fold Cross-Validation (CV). Finally, the RF algorithm reduces the effects of those variables that contribute little to the classification success through repeating predictions for each variable with random data. If the success of the model's classification is changed little by randomizing a given variable, then that variable likely contributes little to the overall success and its effect is minimized. Additionally, RF uses the *mtry* parameter to randomly select a set of variables to try at each node in a tree; the variable that leads to the most successful classification is retained. This serves to reduce the influence of ineffective variables and reduce the influence of variable correlation. Like the *nprune* parameter, *mtry* is also optimized through the use of 10-fold CV as was done and described in the Region 1, 2, and 3 models. These mechanisms are discussed in greater detail in the Task 3 report (Harris 2014) and for RF in Chapter 5 of the Task 4 report (Harris et al. 2014a).

Table 9 - Optimized Number of Variables for Region 7 Models

Subarea	Total Variables	Total w/ Dummy Variables	LR Selected Variables	LR AIC	nprune	mtry
Region 7 All						
riverine_section_1	18	30	28	182904	32	16
riverine_section_2	16	28	24	19530	31	15
riverine_section_3	18	36	31	79276	29	19
riverine_section_4	15	27	24	26116	29	14
riverine_section_5	19	31	30	58747	36	16
riverine_section_6	19	37	32	29724	26	19
riverine_section_7	17	35	29	29931	32	18
riverine_section_8	19	37	31	6444	38	19
riverine_section_9	18	31	24	14319	35	16
upland_section_1	17	29	25	44272	5	15
upland_section_2	19	24	20	7398	29	13
upland_section_3	21	33	30	34557	26	17
upland_section_4	19	19	16	4577	23	2
upland_section_5	21	39	31	34479	38	20
upland_section_6	19	31	28	8263	20	16
upland_section_7	16	28	23	11796	20	15
upland_section_8	14	19	16	3489	23	10
upland_section_9	13	18	14	2209	23	2

Table 10 - Optimized Number of Variables for Region 8 Models

Subarea	Total Variables	Total w/ Dummy Variables	LR Selected Variables	LR AIC	nprune	mtry
Region 8 All						
riverine_section_1	17	35	32	113276	34	10
riverine_section_2	20	38	33	53654	28	11
riverine_section_3	18	29	21	3127	24	8
riverine_section_4	16	34	31	104736	29	18
riverine_section_5	18	36	30	95482	37	19
riverine_section_6	18	36	34	134089	32	19
riverine_section_7	19	37	32	71863	38	19
riverine_section_8	14	25	21	166751	29	7
riverine_section_9	17	28	26	284990	31	15
upland_section_1	14	19	17	94804	22	6
upland_section_2	19	24	22	136437	31	7
upland_section_3	17	28	24	11333	24	8
upland_section_4	13	13	13	75241	18	4
upland_section_5	15	15	14	67129	24	5
upland_section_6	14	19	15	123156	20	6
upland_section_7	17	24	23	62887	23	7
upland_section_8	19	25	23	274681	33	13
upland_section_9	17	28	26	383137	32	15

Table 11 - Optimized Number of Variables for Region 9/10 Models

Subarea	Total Variables	Total w/ Dummy Variables	LR Selected Variables	LR AIC	nprune	mtry
Region 9/10 All						
riverine_section_1	20	38	30	66910	36	20
riverine_section_2	15	33	27	149630	25	17
riverine_section_3	25	44	41	33860	44	23
riverine_section_4	16	27	21	8560	31	14
riverine_section_5	15	26	24	186655	31	14
riverine_section_6	18	29	25	146236	34	15
riverine_section_7	20	38	33	44303	37	20
riverine_section_8	19	30	28	31496	34	16
riverine_section_9	23	36	31	4013	41	19
riverine_section_10	17	35	30	59530	36	18
riverine_section_11	18	29	24	123222	33	15
riverine_section_12	19	32	28	51315	36	17
riverine_section_13	15	27	27	29213	31	14
riverine_section_14	13	18	18	134254	24	10
riverine_section_15	16	34	28	18687	32	18
upland_section_1	16	23	21	61396	28	12
upland_section_2	16	21	18	71777	26	11
upland_section_3	15	20	17	25116	26	11
upland_section_4	22	34	30	20580	38	18
upland_section_5	18	23	22	112210	20	12
upland_section_6	16	21	19	154073	27	11
upland_section_7	19	37	31	37950	32	19
upland_section_8	17	22	21	30756	29	12
upland_section_9	21	34	21	168	32	18
upland_section_10	16	29	23	36719	34	15
upland_section_11	17	22	18	214193	30	12
upland_section_12	22	40	31	44689	34	21
upland_section_13	15	20	20	45706	26	11
upland_section_14	15	20	20	267224	26	11
upland_section_15	17	29	25	32012	32	15

MODEL 3 – SELECTED MODEL TEST SET AND CV ERROR RATES

The final LR, MARS, and RF models were fit on the complete dataset using the selected variables and *nprune* and *mtry* parameter values listed in the tables above. The models were run through 10-fold CV to derive error estimates and the AUC value. The balance between background and site-present data points for model creation was set at a ratio of 3:1, with the background values randomly selected from a pool of 500,000 background values or the entire background sample if there were less than 500,000 cells. The final models were fit using the complete set of data and then calculated for the full population of raster cells within each subarea.

Table 12, Table 13, and Table 14 detail the error estimates and AUC values for each of the selected statistical model types for each subarea. The second column in these tables contains the Root Mean Square Error (RMSE) for the model prediction on a 25% hold-out sample of site-present cells that were not used in fitting the prediction. The third column contains the RMSE (LR model) or Accuracy (MARS and RF models) value for each model calculated as the average error/accuracy from each of the 10 CV out-of-fold samples. As detailed in the Task 3 report, the RMSE is an error estimate that measures the variation and magnitude of errors between the predicted value and the actual value (e.g., site present vs. site absent); simply put, it is the square root of the average of all squared errors. Similarly, Accuracy (for the MARS and RF models) measures the percentage of observations that were correctly classified as either site-present or site-absent. The fourth column is the Coefficient of Variation (CoV) for the error/accuracy expressed as a percentage. The MARS and RF models report Accuracy for the internal CV out-of-fold testing, as opposed to RMSE for the regression based LR model, because these models perform a classification that is measured by how often each observation is correctly classified. The column for AUC presents a single metric that describes the ability of the model to discriminate site-present from site-absent out-of-fold samples averaged across the 10 CV repetitions. This metric was described in detail in the Task 3 report (Harris 2014). Finally, the column for data samples contains the total number of site-present cells for the hold-out and training samples combined.

The RMSE estimate ranges from 0 to infinity and is negatively oriented, so the lower the value, the lower the prediction error. In APM, which has a binary response variable (site present = 1; background = 0), the RMSE is scaled such that 1 is a completely incorrect prediction, 0 is a perfect prediction, and 0.5 is an essentially random prediction. This allows the hold-out test sample RMSE numbers for each of the selected models to be compared relative to each other, but there are factors such as site prevalence and sample size that can influence the RMSE to some degree. For example, upland subareas have a lower RMSE on average than do the riverine subareas (0.277 vs. 0.322 RMSE for all LR held-out samples; 0.236 vs. 0.282 for all MARS held-out samples; and 0.071 vs. 0.127 for all RF held-out samples).

Table 12 - LR Model Prediction Errors from Test Set and 10-Fold CV

Subarea	Test RMSE	CV RMSE	CV RMSECoV	AUC	Data Samples
Region 7 All					
upland section 1	0.177	0.176	1.236	0.987	10270

Table 13 - MARS Model Prediction Errors and Accuracy from Test Set and 10-Fold CV

Subarea	Test RMSE	CV Accuracy	CV AccuracyCoV	AUC	Data Samples
Region 7 All					
riverine section 6	0.226	0.929	0.307	0.9791	10098
riverine section 8	0.225	0.942	0.674	0.9767	3099
upland section 6	0.102	0.988	0.200	0.9944	1230
upland section 8	0.231	0.929	0.977	0.9667	1098
Region 8 All					
riverine section 2	0.162	0.965	0.319	0.992	1926
riverine section 3	0.137	0.975	0.533	0.991	2232
Region 9/10 All					
riverine section 1	0.287	0.884	0.374	0.925	10190
riverine section 2	0.349	0.824	0.179	0.866	31624
riverine section 5	0.342	0.836	0.262	0.880	31332
riverine section 7	0.283	0.882	0.486	0.937	9893
riverine section 12	0.254	0.915	0.349	0.960	5492

Table 14 - RF Model Prediction Errors and Accuracy from test set and 10-fold CV

Subarea	Test RMSE	CV Accuracy	CV AccuracyCoV	AUC	Data Samples
Region 7 All					
riverine section 1	0.147	0.978	0.057	0.991	39482
riverine section 2	0.091	0.991	0.134	0.998	6269
riverine section 3	0.160	0.972	0.108	0.989	27405
riverine section 4	0.130	0.983	0.222	0.994	9923
riverine section 5	0.105	0.989	0.169	0.996	11281
riverine section 7	0.111	0.986	0.147	0.997	8660
riverine section 9	0.128	0.983	0.316	0.995	5977
upland section 2	0.032	0.998	0.061	1.000	3249
upland section 3	0.051	0.997	0.062	1.000	7510
upland section 4	0.040	0.998	0.081	1.000	1820
upland section 5	0.047	0.998	0.055	0.999	11018
upland section 7	0.055	0.996	0.161	1.000	2351
upland section 9	0.083	0.995	0.250	1.000	820
Region 8 All					
riverine section 1	0.125	0.983	0.128	0.993	28382
riverine section 4	0.138	0.978	0.168	0.994	20921
riverine section 5	0.157	0.973	0.120	0.992	19555
riverine section 6	0.163	0.969	0.118	0.989	22063
riverine section 7	0.156	0.968	0.166	0.994	11353
riverine section 8	0.187	0.954	0.102	0.987	23868
riverine section 9	0.157	0.969	0.051	0.992	34272
upland section 1	0.072	0.994	0.083	0.999	21629
upland section 2	0.081	0.994	0.054	0.998	25368
upland section 3	0.045	0.998	0.072	1.000	5007
upland section 4	0.073	0.995	0.068	0.999	18500
upland section 5	0.083	0.992	0.073	0.998	16780
upland section 6	0.097	0.990	0.059	0.999	18561
upland section 7	0.106	0.989	0.124	0.998	15123
upland section 8	0.115	0.986	0.055	0.994	62986
upland section 9	0.080	0.993	0.032	0.998	55394
Region 9/10 All					
riverine section 3	0.101	0.986	0.119	0.998	10967
riverine section 4	0.095	0.990	0.307	0.998	957
riverine section 6	0.149	0.971	0.116	0.990	16430
riverine section 8	0.088	0.992	0.102	0.999	5508
riverine section 9	0.079	0.994	0.352	0.999	2151

Subarea	Test RMSE	CV Accuracy	CV AccuracyCoV	AUC	Data Samples
riverine section 10	0.155	0.974	0.176	0.990	10153
riverine section 11	0.124	0.982	0.154	0.996	15429
riverine section 13	0.096	0.989	0.123	0.998	6013
riverine section 14	0.125	0.983	0.087	0.996	15436
riverine section 15	0.091	0.992	0.126	0.998	3473
upland section 1	0.077	0.993	0.050	0.998	14632
upland section 2	0.066	0.995	0.052	0.999	13336
upland section 3	0.071	0.994	0.099	0.999	4301
upland section 4	0.097	0.990	0.203	0.998	6664
upland section 5	0.081	0.993	0.069	0.999	19985
upland section 6	0.080	0.993	0.059	0.998	23566
upland section 7	0.068	0.995	0.090	0.999	9658
upland section 8	0.070	0.994	0.119	0.999	7399
upland section 9	0.042	1.000	0.000	1.000	440
upland section 10	0.091	0.992	0.138	0.998	7591
upland section 11	0.101	0.989	0.051	0.997	48491
upland section 12	0.096	0.990	0.070	0.997	9983
upland section 13	0.058	0.997	0.093	1.000	10485
upland section 14	0.099	0.989	0.078	0.998	61548
upland section 15	0.075	0.993	0.128	0.999	8910

This is the result of a lower prevalence of site-present locations and an often more restricted choice of site locations in reference to the predictor variables in the upland subareas. The RMSE statistic is very sensitive to large magnitude errors, of which there are more in the riverine areas. This is because there is a higher prevalence of sites and more area than is considered sensitive to archaeological sites. Therefore, there are more cells that are observed to be background (a value of zero) than are predicted to be likely site locations (a value close to one). There are more of these high magnitude differences in the riverine areas, which tend to raise the RMSE; the opposite effect is true for the uplands. However, even with bias derived from known site prevalence and the overall size of the subareas, the RMSE values are all quite low and show models with a high degree of discrimination and the ability to correctly predict known site-present cells from the hold-out samples.

The RMSE and accuracy CoV show the percent change in the error/accuracy within the 10 out-of-fold samples for each CV repetition. The largest RMSE CoV value, which shows a larger magnitude of variation between the error/accuracy rates, is 7.2%. While this shows notable swings in the RMSE of the out-of-fold samples, the fact that they are percentages of very small RMSE values leads to low error rates even at the upper end of the variation. The upland and riverine subareas have slightly different average RMSE/Accuracy CoV sample means (2.26 vs. 1.18 RMSE CoV for all LR out-of-fold samples; 0.39 vs. 0.43 Accuracy CoV for all MARS out-of-fold samples; and 0.09 vs. 0.15

Accuracy CoV for all RF out-of-fold samples). While not a significant trend, the difference in CoV between riverine and upland areas is derived from the same biases of prevalence and area noted above.

The tables and discussion above show the steps for variable selection, parameterization, and error rates based on a 25% hold-out sample and 10-fold CV. The error rates resulting from the 10-fold CV, expressed as average RMSE, Accuracy, and the CoV of each show that the LR, MARS, and RF algorithms are variably successful in identifying the pattern of predictor variables that define the location of known sites within all selected subareas. Additionally, the AUC values (a single number that is designed to show the quality of a model across all thresholds) show that the models are very accurate for each of the selected subareas. Based on these findings, all of the selected models appear to be capable of detecting the known sites as well as predicting the location of site-present cells that were held-out from the model building. There are no red-flags that would indicate that any one subarea has an inadequate or poorly performing model. The findings in the next chapter will demonstrate how these models are applied to each subarea and how the thresholds for sensitivity strata are determined.

SPECIAL NOTE ON REGION 9/10 UPLAND AND RIVERINE SECTION 9

The tables above give a variety of modeling metrics for the models selected to represent each subarea. These metrics are generated by statistical methods that seek to fit the given data while still being able to generalize to areas that have not yet been surveyed for sites. The assumptions of these methods, as discussed throughout this project, are based to a high degree on the correlation of archaeological site locations to natural features. These features need not be in the exact spatial arrangement as they were thousands of years ago when the site was occupied, but still require a systemic association to the areas where we now find sites. The use of tests such as K-S and MW help us to identify relevant environmental variables that differentiate site locations from the background, and the statistical tests have their own internal variable selection measures. However, in the case where the environment is so affected by historic land alteration as to no longer resemble the pre-contact landscape, this assumption is violated. This is very likely the case for upland and riverine section 9 of Region 9/10. These subareas contain the entirety of the city and county of Philadelphia.

The City of Philadelphia is located at the confluence of the Delaware River and its largest tributary, the Schuylkill River. The area was once a vast ecotone of upland resources overlooking miles of marsh and meandering tributaries. For the reasons that early Europeans were drawn to the area, it is likely that Native Americans valued the region's wealth of resources for thousands of years prior. Until relatively recently, the prehistoric archaeology of Philadelphia was assumed to be nearly nonexistent and only preserved in very special circumstances. The small number of PASS sites in the Philadelphia City limits attests to this perception. However, through a number of archaeological surveys over the past decade, and more recently because of surveys associated with PennDOT's reconstruction of I-95 within the city, many additional prehistoric sites have been located. These more recent finds have been located within the Philadelphia waterfront and within the developed core of the city. These sites have been found under nineteenth- and twentieth-century buildings, under rail yards, in backyards, in graveyards, within alley ways, and in a variety of previously unexpected settings. These sites range in size from a few square meters to many acres and often include features and intact soils. The understanding relative to preservation potential gained from these recent finds is that sites likely exist throughout the city and that special or unique preservation environments are not required. It appears that the placement of fill, building material, and previous building practices left larger than expected portions of the early and pre-historic landscape intact. While the prospection and detection of sites in these settings requires different techniques such as mechanical stripping and monitoring compared to comparable surveys in less developed areas, the sites are present when properly looked for.

While this finding is great for gaining a better understanding of prehistoric occupation of the City of Philadelphia, it greatly complicates the process of correlative inductive modeling as undertaken here. Essentially, the current environmental data based on elevation and hydrology combined with the small sample of PASS sites recorded for Philadelphia violates the assumption that the current environment is a proxy for the past environment. Although this assumption is tested in every developed location in the Commonwealth, it clearly cannot hold up to the massive resurfacing of the dense urban center. For these reasons, the models chosen to represent riverine and upland sections 9 of Region 9/10 should be taken as merely suggestive of what the current group of PASS sites indicate. A very different set of assumptions and methods would be required to model the sensitivity of Philadelphia including reconstruction of elements of the prehistoric environments, identifying factors in preservation potential, and mapping historic cut/fill and basement depth, all of which are beyond the scope of the current study.

6
THRESHOLD SELECTION AND FINALIZATION – REGIONS 7, 8, 9, AND 10

In the previous chapter, the subarea models for LR, MARS, and RF were validated using a hold-out sample, 10-fold CV to produce prediction error estimates (RMSE) and percent accuracy, prediction error stability across hold-out samples (CoV), and a measure of a model's ability to discriminate site-present and background cells across the range of predicted probabilities (AUC). From these values, the LR, MARS, and RF models selected for each subarea appear to accurately classify known site locations and do so with a relatively low variation in prediction accuracy. Whereas the previous chapter detailed the model building and validation process using random samples of sites and background from each subarea, the data presented in this chapter will show the results of the models applied to the full population of data for each subarea, as well as how choosing different thresholds affects the final evaluation of sensitivity.

COMPARING MODELS AT 0.5 PREDICTED PROBABILITY

The AUC statistic presented in the tables in Chapter 5, along with RMSE and accuracy, give impressions of the models' overall ability to predict site-present cells. However, as elaborated in the beginning of this report, models that seek to define presence and absence are best evaluated at a given threshold. There are many different methods and issues for finding optimal and useful thresholds, but the best method is specific to a single model problem or field of study. For these reasons, a model's applicability and usefulness for a certain purpose is directly related to the threshold that is selected to represent presence and absence. Further along in this chapter, each model will be evaluated at a selected threshold, but this creates an uneven field from which to compare models. In order to better compare the results of models on more level terms, it is best to pick a common threshold and calculate model metrics uniformly. Table 15, Table 16, and Table 17 compare each of the models at an arbitrary predicted probability threshold of $p = 0.5$. This threshold choice is essentially arbitrary, but choosing a threshold halfway between the extremes of the predicted probability distribution ($p = 0$ and $p = 1$) offers the most balanced point to compare results. The point of choosing this arbitrary threshold is to compare model results without the assumptions derived from implicitly selected thresholds as described in the section following this.

These tables present a series of metrics that allow the models to be directly compared with one another. As discussed in Chapter 4 of the Task 4 report, the Kappa statistic can be greatly affected by the balance of positive and negative observation; in the case of these models, that is effectively controlled by the prevalence of known archaeological sites. For these reasons, the tables below present a mean from a sample of Kappa statistics drawn from the site-present prediction compared to 1,000 bootstrapped background cell samples, at a ratio of three background cells to one site-present cell. Using the 3:1 ratio downsamples the background cell data set and removes the drastic imbalance created by modeling large areas with low known site prevalence. Further, the 1,000 bootstrapped samples of background cells guard against drawing an unrepresentative sample to represent the environmental background. Even with these safeguards in place, the prevalence of known sites still has some influence on the Kappa, as can be seen in the trend of higher Kappa statistics for upland subareas. Since the Kappa compares the model against an estimate of the chances of randomly finding a site, and known sites are generally dispersed in upland areas, the by-chance occurrence of sites is lower and therefore the Kappa will be a bit higher for a successful model. However, despite this small bias, the mean Kappa statistics presented in the tables below offer a way to compare the models outright and against each other. The 95% confidence intervals of Kappa sample are also listed. Finally, the tables below present the percent-sites, percent-background, and Kg at the 0.5 threshold.

Table 15 - Comparing Kg and Kappa at a Threshold of 0.5, Selected LR Models

Subarea	Back-ground %	Site-Present %	Kg @ 0.5	3:1 Balanced Mean Kappa	Upper 95%	Lower 95%
Region 7 All						
upland section 1	6.094	91.373	0.933	0.825	0.831	0.818

Table 16 - Comparing Kg and Kappa at a Threshold of 0.5, Selected MARS Models

Subarea	Back-ground %	Site-Present %	Kg @ 0.5	3:1 Balanced Mean Kappa	Upper 95%	Lower 95%
Region 7 All						
riverine section 6	10.186	91.899	0.889	0.762	0.770	0.755
riverine section 8	7.846	87.835	0.911	0.773	0.786	0.760
upland section 6	1.339	67.398	0.980	0.724	0.748	0.700
upland section 8	8.121	77.049	0.895	0.685	0.711	0.659
Region 8 All						
riverine section 2	4.942	50.104	0.901	0.494	0.519	0.469
riverine section 3	3.522	98.790	0.964	0.923	0.932	0.914
Region 9/10 All						
riverine section 1	19.533	70.000	0.721	0.450	0.460	0.440
riverine section 2	25.325	80.562	0.686	0.473	0.479	0.468
riverine section 5	32.560	81.083	0.598	0.391	0.397	0.386
riverine section 7	21.343	82.644	0.742	0.531	0.540	0.522
riverine section 12	10.984	69.082	0.841	0.578	0.591	0.565

Table 17 - Comparing Kg and Kappa at a Threshold of 0.5, Selected RF Models

Subarea	Back-ground %	Site-Present %	Kg @ 0.5	3:1 Balanced Mean Kappa	Upper 95%	Lower 95%
Region 7 All						
riverine section 1	2.873	99.909	0.971	0.945	0.947	0.943
riverine section 2	1.669	99.984	0.983	0.969	0.973	0.966
riverine section 3	3.727	99.759	0.963	0.928	0.931	0.926
riverine section 4	3.076	100.000	0.969	0.944	0.948	0.941
riverine section 5	1.562	99.973	0.984	0.970	0.973	0.967
riverine section 7	2.566	99.965	0.974	0.952	0.956	0.949
riverine section 9	2.973	99.950	0.970	0.946	0.951	0.941
upland section 2	0.301	99.877	0.997	0.994	0.996	0.992
upland section 3	0.583	99.920	0.994	0.989	0.991	0.987
upland section 4	0.371	100.000	0.996	0.993	0.996	0.990
upland section 5	0.372	99.982	0.996	0.993	0.994	0.992
upland section 7	1.081	100.000	0.989	0.981	0.985	0.976
upland section 9	2.194	100.000	0.978	0.962	0.973	0.952
Region 8 All						
riverine section 1	2.303	99.940	0.977	0.956	0.958	0.954
riverine section 4	3.046	99.785	0.969	0.943	0.946	0.941
riverine section 5	3.957	99.816	0.960	0.927	0.930	0.924
riverine section 6	3.940	99.615	0.960	0.925	0.927	0.922
riverine section 7	23.390	98.934	0.764	0.624	0.631	0.616
riverine section 8	30.621	99.556	0.692	0.532	0.537	0.526
riverine section 9	3.576	99.323	0.964	0.928	0.930	0.925
upland section 1	1.182	99.954	0.988	0.978	0.980	0.976
upland section 2	0.945	99.921	0.991	0.981	0.983	0.980
upland section 3	0.428	100.000	0.996	0.992	0.994	0.990
upland section 4	1.030	99.989	0.990	0.981	0.983	0.980
upland section 5	1.701	99.952	0.983	0.969	0.971	0.967
upland section 6	1.946	99.903	0.981	0.964	0.966	0.962
upland section 7	2.668	99.947	0.973	0.952	0.955	0.950
upland section 8	2.177	99.957	0.978	0.959	0.960	0.957
upland section 9	1.156	99.953	0.988	0.978	0.979	0.977
Region 9/10 All						
riverine section 3	22.669	99.927	0.773	0.631	0.639	0.623
riverine section 4	2.012	100.000	0.980	0.964	0.973	0.954
riverine section 6	28.245	96.153	0.706	0.545	0.552	0.538
riverine section 8	14.907	99.964	0.851	0.755	0.764	0.745
riverine section 9	0.993	100.000	0.990	0.982	0.986	0.977

Subarea	Back-ground %	Site-Present %	Kg @ 0.5	3:1 Balanced Mean Kappa	Upper 95%	Lower 95%
riverine section 10	3.355	99.823	0.966	0.936	0.940	0.932
riverine section 11	2.880	99.708	0.971	0.946	0.949	0.943
riverine section 13	1.631	99.767	0.984	0.968	0.972	0.964
riverine section 14	2.925	99.760	0.971	0.945	0.948	0.942
riverine section 15	1.742	99.971	0.983	0.968	0.973	0.963
upland section 1	0.995	99.993	0.990	0.981	0.983	0.979
upland section 2	1.113	99.955	0.989	0.979	0.981	0.977
upland section 3	0.934	99.907	0.991	0.982	0.985	0.978
upland section 4	2.535	99.970	0.975	0.955	0.959	0.951
upland section 5	1.167	99.920	0.988	0.978	0.979	0.976
upland section 6	1.204	99.958	0.988	0.977	0.979	0.976
upland section 7	1.148	100.000	0.989	0.979	0.981	0.976
upland section 8	1.191	99.932	0.988	0.978	0.981	0.975
upland section 9	1.145	100.000	0.989	0.980	0.991	0.970
upland section 10	1.485	99.947	0.985	0.972	0.975	0.969
upland section 11	2.183	99.988	0.978	0.959	0.961	0.958
upland section 12	1.576	99.820	0.984	0.969	0.972	0.966
upland section 13	0.761	99.990	0.992	0.986	0.988	0.984
upland section 14	1.943	99.961	0.981	0.964	0.965	0.963
upland section 15	1.313	99.944	0.987	0.976	0.978	0.973

The above tables show that the models as applied to the full subarea study area are generally very good at identifying site-present locations relative to a random chance of finding a site. Between the models, the Kappa results show a relatively consistent trend within the different model types. As illustrated in Figure 26, across all model types the mean Kappa statistics range from a low of $k = 0.39$ to a high of $k = 0.99$; most with relatively narrow 95% confidence intervals. Unsurprisingly, the average Kappa for all models (including those selected for the final raster layer and those not selected) of a particular model type are lowest with Model 2 ($k = 0.65$) and highest with the RF model ($k = 0.93$), with LR ($k = 0.49$) and MARS ($k = 0.59$) models in between. The most notable trend in Figure 26, is the majority of upland subareas scoring a higher Kappa (average $k = 0.96$) than the majority of riverine subareas (average $k = 0.80$). This trend is most likely attributable to the lower prevalence of known sites in the uplands and the lower chance of randomly findings a site there. The Kg statistic and site/background percentages show that the models are successful at capturing the known site pattern within a small portion of the model.

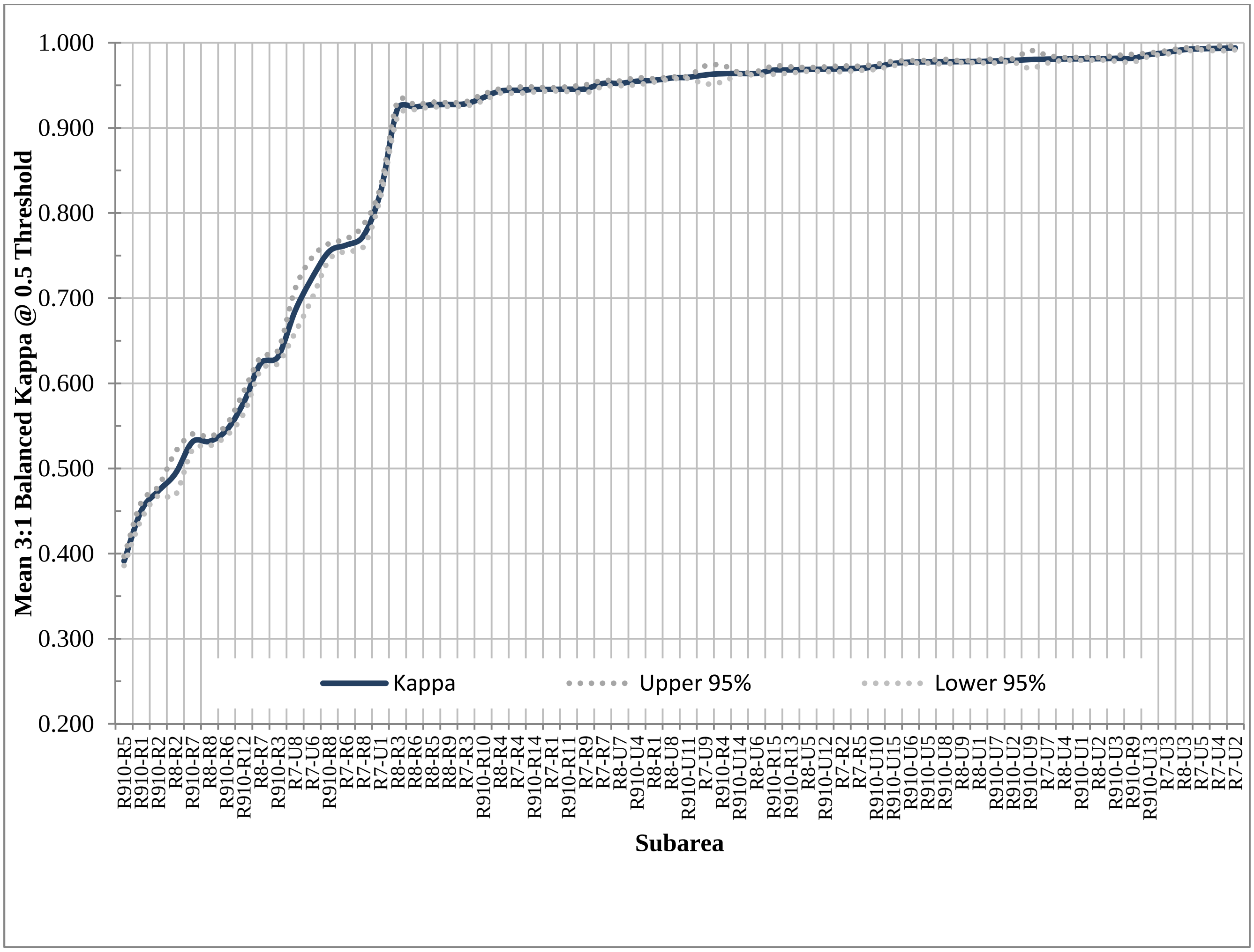


Figure 26 - 3:1 balance mean Kappa and 95% confidence intervals for all subarea models.

ESTABLISHING MODEL THRESHOLDS

As discussed in detail in the Task 4 report and repeated here for clarity, the discriminatory ability of the models created in this project is at a level not yet seen in APM and raises a new host of questions regarding the purpose and intention of these models. The low background percentages of these models relative to the site-present percentages are drastically smaller than in most previous APM, but in fact reflect the reality of a low prevalence phenomenon such as archaeological sites. While the models and methodology employed here have been adjusted to account for low prevalence and unequal weights between false-positives (low weight) and false-negatives (high weight) the reality that archaeological site occurrence only comprises a very finite portion of the total landscape is inescapable. The means of dealing with this reality has now been shifted from using the lower discriminant, less accurate, and obfuscated models of the past to using more thoughtful interpretation, problem-specific model applications, and a better understanding of the model's abilities and limitations. A large part of this reckoning is the better understanding and application of model thresholds.

Due to the ability of modern statistical models to identify patterns and discriminate site locations much more effectively than in the past, the onus of portioning site-present from site-absent areas has shifted. In the past, many model-building efforts had the simple goal of maximizing the site-present percent and minimizing the site-likely area. This was the primary challenge of the modeling effort, and the thresholds that determined site-likely areas were often an afterthought or predicted on the low performance of the model. With the MARS model, RF model, and other innovations in statistical modeling, achieving very well fit—and at times overfit—models is not as great a challenge. No longer is the goal of simply reducing the area within which a majority of the sites are contained sufficient. The models presented here are capable of minimizing that area to a small portion of the landscape that is closer to the true prevalence of known sites and more sensitive to previous survey bias. The new goal given these advances is to accurately model the site pattern with a low error rate and then select model thresholds that best achieve the goals of the project. If the project aims to minimize the site-likely area, then a higher threshold is useful. To generalize the site-likely area, a lower threshold is useful. As discussed in the Task 4 report, the selection of an appropriate threshold can be based on a number of factors, including arbitrary decisions, field or project-specific standards and goals, or optimization based on quantitative model metrics. To illustrate the points above, the Task 4 report provided a series of different thresholds appropriate for different model objectives. Although only two thresholds were chosen to partition the final models, the full variety of thresholds is also presented here. This is for the purpose of comparison between the models of Task 4, Task 5, and Task 6, but also to provide these thresholds in the event that these models are to be repartitioned for a different purpose.

On the other hand, the proportionally weighted models are much more akin to traditional models that sought to primarily maximize the correct site prediction while secondarily trying to limit the growth of the site-likely area. The use of discriminatory variables and proportional weighting definitely lift these models above the common judgmental APM, but not to the level of the statistical models. This is not a bad thing; it is, however, an inescapable reality of the method used in areas of low site counts. The proportional models suffer the same fate as the statistical models in being subject to the need for clearly defined and justified thresholds. For that reason, the proportionally weighted models were put through the same threshold creation routine as the statistical models and will be presented along with them for the remainder of the report. It may be helpful to repeat that the output sensitivity of the proportionally weighted models are on the same zero to 1 scale as the statistical models so the thresholds, Kg, and Kappa are also scaled appropriately.

Table 18, Table 19, and Table 20 present eight different potential thresholds based on optimized model metrics and previous research in APM. These values are graphically represented in a chart for each subarea, included as Appendix F. The thresholds presented here are termed as:

- MaxKappa: the threshold that maximizes the Kappa statistic
- Max Kg: the threshold that maximizes the Kg statistic
- Sens = Spec: the threshold at which sensitivity and specificity are equal
- X-Over: the threshold at which site-present and background lines cross in the cross-over graph
- Sens @ 0.85: the threshold that is optimized for a sensitivity of 0.85
- Spec @ 0.67: the threshold that is optimized for a specificity of 0.67
- Pred = Obs: the threshold at which the predicted site prevalence equals the observed or assigned site prevalence (calculated at two different assigned values)

Table 18 - Optimal Thresholds for Various Selection Methods; Selected LR Models

Threshold Type	Maximize		Balanced		Domain Specific		Prevalence Based	
Subarea	MaxKappa	MaxKG	Sens = Spec	X-Over	Sens @ 0.85	Spec @ 0.67	Pred = Obs @ 0.1	Pred = Obs @ 0.2
Region 7 All								
upland section 1	0.69	1.00	0.58	0.60	0.50	0.47	0.72	0.62

Table 19 - Optimal Thresholds for Various Selection Methods; Selected MARS Models

Threshold Type	Maximize		Balanced		Domain Specific		Prevalence Based	
Subarea	MaxKappa	MaxKG	Sens = Spec	X-Over	Sens @ 0.85	Spec @ 0.67	Pred = Obs @ 0.1	Pred = Obs @ 0.2
Region 7 All								
riverine section 6	0.96	1.00	0.52	0.54	0.62	0.10	0.52	0.19
riverine section 8	0.95	1.00	0.41	0.42	0.59	0.13	0.40	0.21
upland section 6	0.98	1.00	0.11	0.14	0.10	0.06	0.15	0.09
upland section 8	0.99	1.00	0.38	0.40	0.39	0.21	0.44	0.30
Region 8 All								
riverine section 2	0.77	0.80	0.17	0.18	0.16	0.07	0.29	0.13
riverine section 3	0.94	1.00	0.70	0.72	0.92	0.02	0.08	0.04
Region 9/10 All								
riverine section 1	0.64	0.94	0.44	0.46	0.35	0.35	0.62	0.48
riverine section 2	0.78	0.96	0.51	0.54	0.42	0.41	0.69	0.55
riverine section 5	0.69	0.74	0.55	0.56	0.44	0.48	0.83	0.66
riverine section 7	0.92	0.98	0.51	0.52	0.46	0.33	0.75	0.52
riverine section 12	0.79	1.00	0.32	0.34	0.26	0.21	0.53	0.33

Table 20 - Optimal Thresholds for Various Selection Methods; Selected RF Models

Threshold Type	Maximize		Balanced		Domain Specific		Prevalence Based	
Subarea	MaxKappa	MaxKG	Sens = Spec	X-Over	Sens @ 0.85	Spec @ 0.67	Pred = Obs @ 0.1	Pred = Obs @ 0.2
Region 7 All								
riverine section 1	0.81	1.00	0.73	0.74	0.89	0.07	0.20	0.10
riverine section 2	0.92	1.00	0.73	0.76	0.94	0.07	0.18	0.12
riverine section 3	0.80	1.00	0.73	0.74	0.87	0.01	0.20	0.08
riverine section 4	0.90	1.00	0.77	0.80	0.93	0.07	0.23	0.10
riverine section 5	0.86	1.00	0.75	0.76	0.94	0.01	0.14	0.08
riverine section 7	0.92	1.00	0.76	0.78	0.92	0.01	0.16	0.08
riverine section 9	0.95	1.00	0.76	0.78	0.94	0.07	0.24	0.14
upland section 2	0.99	1.00	0.61	0.64	0.99	0.01	0.08	0.04
upland section 3	0.95	1.00	0.64	0.66	0.96	0.01	0.10	0.08
upland section 4	0.97	1.00	0.65	0.66	0.97	0.01	0.10	0.04
upland section 5	0.94	1.00	0.75	0.78	0.98	0.01	0.12	0.08
upland section 7	0.98	1.00	0.71	0.72	0.97	0.01	0.14	0.08
upland section 9	0.96	1.00	0.70	0.72	0.93	0.13	0.27	0.18
Region 8 All								
riverine section 1	0.84	1.00	0.74	0.76	0.92	0.07	0.16	0.10
riverine section 4	0.84	1.00	0.66	0.68	0.87	0.15	0.29	0.20
riverine section 5	0.85	1.00	0.71	0.72	0.87	0.15	0.31	0.21
riverine section 6	0.80	1.00	0.68	0.70	0.86	0.13	0.30	0.18
riverine section 7	0.99	1.00	0.77	0.80	0.78	0.30	0.87	0.60
riverine section 8	0.98	1.00	0.82	0.84	0.78	0.36	0.92	0.83
riverine section 9	0.75	1.00	0.64	0.66	0.81	0.09	0.22	0.14
upland section 1	0.95	1.00	0.71	0.74	0.95	0.07	0.16	0.10
upland section 2	0.84	1.00	0.70	0.72	0.95	0.01	0.12	0.08
upland section 3	0.98	1.00	0.82	0.84	0.99	0.01	0.08	0.04
upland section 4	0.94	1.00	0.75	0.76	0.97	0.07	0.16	0.10
upland section 5	0.93	1.00	0.74	0.76	0.94	0.07	0.20	0.12
upland section 6	0.88	1.00	0.65	0.68	0.89	0.09	0.21	0.14
upland section 7	0.94	1.00	0.70	0.72	0.92	0.09	0.26	0.16
upland section 8	0.84	1.00	0.79	0.80	0.94	0.07	0.18	0.10
upland section 9	0.88	1.00	0.73	0.74	0.95	0.07	0.14	0.08
Region 9/10 All								
riverine section 3	0.99	1.00	0.94	0.96	0.94	0.11	0.98	0.89
riverine section 4	0.89	1.00	0.68	0.70	0.90	0.11	0.23	0.16
riverine section 6	0.99	1.00	0.78	0.80	0.76	0.38	0.96	0.73

Threshold Type	Maximize		Balanced		Domain Specific		Prevalence Based	
Subarea	MaxKappa	MaxKG	Sens = Spec	X-Over	Sens @ 0.85	Spec @ 0.67	Pred = Obs @ 0.1	Pred = Obs @ 0.2
riverine section 8	0.99	1.00	0.84	0.86	0.94	0.21	0.61	0.39
riverine section 9	0.98	1.00	0.73	0.76	0.98	0.09	0.18	0.14
riverine section 10	0.80	1.00	0.73	0.76	0.88	0.01	0.17	0.08
riverine section 11	0.86	1.00	0.67	0.68	0.87	0.13	0.27	0.18
riverine section 13	0.92	1.00	0.67	0.70	0.91	0.01	0.15	0.08
riverine section 14	0.85	1.00	0.67	0.70	0.86	0.11	0.26	0.17
riverine section 15	0.94	1.00	0.78	0.80	0.95	0.07	0.18	0.10
upland section 1	0.92	1.00	0.77	0.78	0.97	0.07	0.14	0.10
upland section 2	0.95	1.00	0.75	0.78	0.97	0.01	0.12	0.08
upland section 3	0.97	1.00	0.62	0.64	0.94	0.07	0.14	0.08
upland section 4	0.96	1.00	0.76	0.78	0.95	0.09	0.24	0.16
upland section 5	0.91	1.00	0.70	0.72	0.93	0.01	0.12	0.08
upland section 6	0.89	1.00	0.72	0.74	0.95	0.07	0.15	0.10
upland section 7	0.95	1.00	0.77	0.80	0.97	0.07	0.15	0.08
upland section 8	0.95	1.00	0.73	0.76	0.97	0.07	0.15	0.10
upland section 9	0.99	1.00	0.88	0.90	0.99	0.01	0.17	0.08
upland section 10	0.95	1.00	0.77	0.78	0.96	0.01	0.14	0.08
upland section 11	0.90	1.00	0.78	0.80	0.94	0.07	0.18	0.12
upland section 12	0.90	1.00	0.70	0.72	0.93	0.01	0.10	0.08
upland section 13	0.96	1.00	0.77	0.78	0.98	0.07	0.14	0.08
upland section 14	0.91	1.00	0.75	0.76	0.94	0.11	0.21	0.15
upland section 15	0.96	1.00	0.73	0.74	0.96	0.07	0.18	0.12

The full description and technical details of each of these thresholds is presented in the Task 4 report; a summary of each is provided here. The first two thresholds, MaxKappa and MaxKg, are means of maximizing a particular metric to find a threshold. In this case it is maximizing Kappa (maximizing the proportion of correctly classified sites while accounting for random agreement) and maximizing Kg (maximizing the proportion of correctly classified sites while accounting for the area of the classification). The second two threshold metrics, Sens = Spec and X-Over, are ways to find where the model balances false-positive and false-negative errors. This is the point where the model's prediction is just as likely to be right about correctly predicting a site as it is correctly predicting a background cell. The metric of Sens = Spec is calculated from the ROC curve to find the threshold at which those type measures are about equal. The X-Over is included here because it has been traditionally cited in APM literature as the optimal location to define a threshold (Kvamme 1988). The third group of threshold selection methods presented here, Sens @ 0.85 and Spec @ 0.67, are labeled as "Domain Specific" thresholds because these allow for the specification of sensitivity or specificity based on an arbitrary value established for a specific purpose. In this case a specificity of 0.67 assures that no more than 33% of the true-negative observations (background cells) are classified as site-likely; the threshold for required sensitivity is set to 0.85. This assures that the site-likely area misclassifies no more than 15% of the known site-present cells. The final two thresholds, Pred = Obs @ 0.1 and Pred = Obs @ 0.2, are labeled as “Prevalence Based” because they account for the prevalence of positive observations (sites) to adjust the threshold values. The low prevalence of archaeological sites across the landscape poses an obstacle to the modeling effort. This is because the data being modeled are heavily imbalanced toward the negative observation (site not-present cells), and most models will favor predictions for the larger of the two classes.

Throughout Regions 7, 8, 9, and 10, the overall prevalence of known archaeological sites with a prehistoric component is 0.0031. Riverine subareas have an average prevalence of 0.0128 and upland subareas have an average prevalence of 0.0024. Figure 27 shows the prevalence of all subareas within Regions 7, 8, 9, and 10. The lowest prevalence is within Region 7 Upland Section 9 at 0.00009 and the highest is within Region 9 Riverine Section 5 at 0.0369. By setting the threshold for the site-likely area at 0.1, the threshold is compensating for survey and detection bias. Clearly, the density of archaeological sites varies widely throughout the state, but it is also clear that this is to some degree a function of survey bias. Establishing a baseline prevalence for site-likely predictions creates a basis for interpretation and consistency, much like Sens @ 0.85 and Spec @ 0.67.

The choice of appropriate thresholds for model prediction is driven by project needs and management goals. The threshold selection methods and thresholds discussed above are all appropriate for these models, depending on how they are to be used: ranging maximized thresholds are the most conservative, the cross-over thresholds are the most balanced, and the prevalence thresholds are the most liberal. Any one of these approaches could be effective given the problem at hand, but approaches such as the requirements of sensitivity or specificity and prevalence-based thresholds are likely the most applicable to APM. Freeman and Moisen (2008:57) came to the same conclusion based on studies in ecological modeling, which shares many of the same obstacles and goals as APM. Additionally, Freeman and Moisen concluded that no one set of thresholds or the resulting map can fulfill all of the objectives for which a model could be used, and that essentially the model should be viewed as a tool that needs to be adapted to a specific task through the use of thresholds. They state that, “[u]ltimately, maps will typically have multiple and sometimes conflicting management applications and thus providing users with a continuous probability surface may be the most versatile method … allowing threshold choice to be matched up with map use” Freeman and Moisen (2008:57).

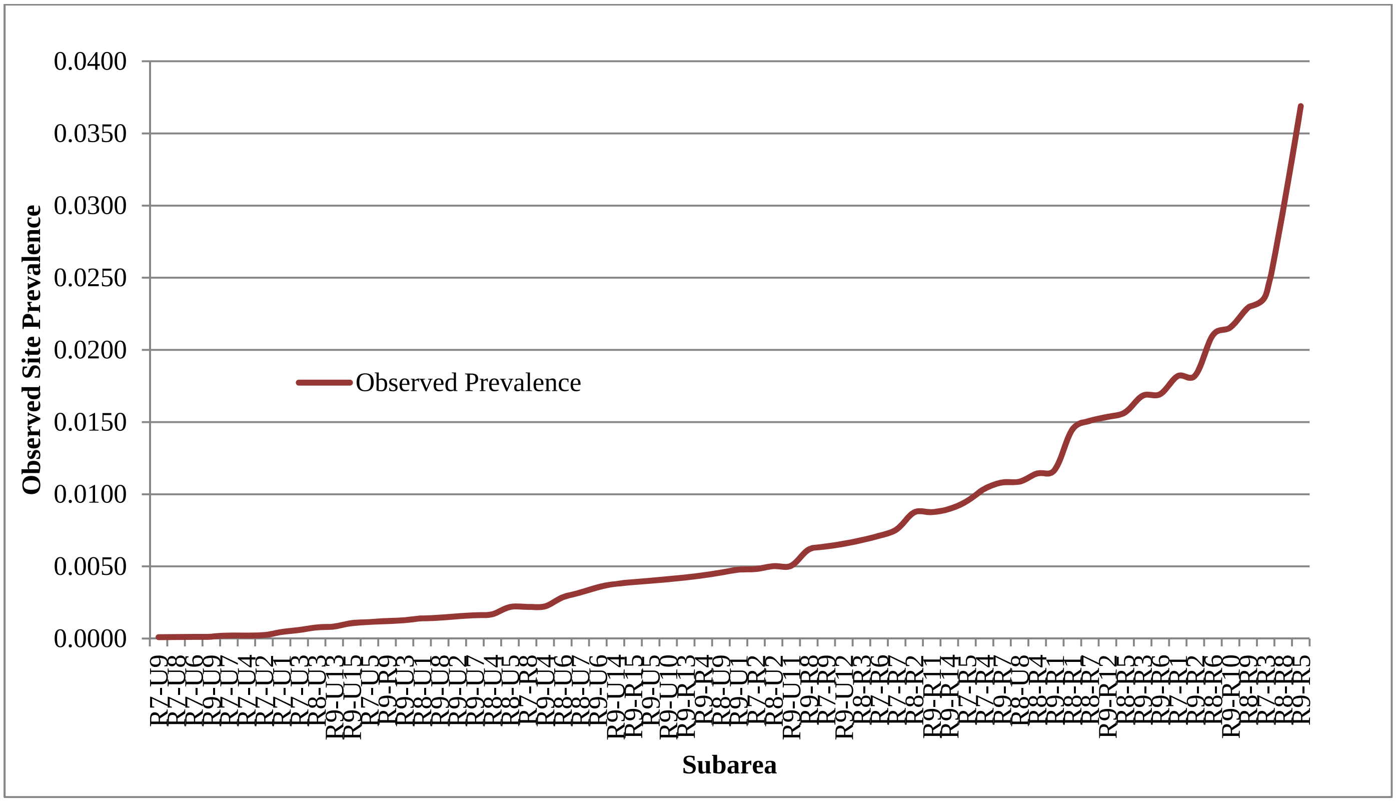


Figure 27 - Average prevalence of prehistoric sites by subarea.

SELECTED MODEL THRESHOLDS

This project supports Freeman and Moisen's conclusion and will provide the continuous probability distribution maps as a part of the final deliverable. However, this project also recognizes that with the insight gained through this analysis, a recommended set of thresholds should be provided and maps based on these thresholds should be created.

The thresholds selected for this project are based on both the required specificity and prevalence methods. The threshold for high sensitivity sets the predicted site-likely prevalence to 0.1. This threshold assumes that there is a large portion of the archaeological record that has not yet been discovered in each subarea. The true prevalence of archaeological sites in a region would be very difficult to estimate, especially in a region where very few sites are easily detected from surface survey (as opposed to arid desert regions with many sites on the surface). However, a prevalence target of 0.1 is well higher than the highest observed prevalence and incorporates approximately 9–11% of the subarea for each model.

The threshold for the low end of moderate probability, and therefore the low end of the site-likely area, is set at a specificity target of 0.67. This assures that no more than 33% of the true-negative observations (background cells) are classified as site-likely. In essence, this sets the site-likely area at close to 33% of the total subarea. This threshold is used in response to the Mn model goal of maximizing site-present locations within 33% of the study area (Mn/Model n.d.). As discussed earlier, the recommendation by Oehlert and Shea (2007) of requiring a sensitivity of 0.85 and minimizing specificity is not very useful here because it does not set a lower bound on specificity. The implementation of the specificity at a 0.67 threshold used here establishes a lower bound (at 0.67) and takes a more conservative approach than suggested by Oehlert and Shea.

On balance, the use of these two threshold measures creates a standardized set of high, moderate, and low classifications across the three regions. As evident in Table 21, Table 22, and Table 23, the combined site-likely area of high and moderate probability includes from 81% to 100% of the known site-present cells in a site-likely area from 13% to 34% of the study area, for Kg statistics ranging from 0.58 to 0.87: an average Kg of 0.705. The boxplots in Figure 28 show the variation on Kg statistics for the 66 selected models across the four model types. As anticipated, the mean Kg increases and the variation in Kg decreases as the models become more powerful. The confusion matrices for each of the models, classified as site-likely (high and moderate sensitivity) and site-unlikely (low sensitivity), are presented in Appendix G. The overall confusion matrix representing the site-likely classification for the entirety of Regions 7, 8, 9, and 10 is presented in Table 24. Figure 29 depicts an overview of high, moderate, and low sensitivity for the entirety of Regions 7, 8, 9, and 10. These data will be provided as ESRI raster grids and GeoTiff formats for detailed viewing and analysis.

Table 21 - Kg and Cell Percentages at Suggested Final Thresholds, Selected LR Models

	Pred = Obs @ 0.1, High Sensitivity				Specificity @ 0.67, Moderate Sensitivity			
Subarea	Threshold	% Background	% Sites	Kg	Threshold	% Background	% Sites	Kg
Region 7 All								
upland section 1	0.72	10%	95%	0.89	0.47	30%	97%	0.69

Table 22 - Kg and Cell Percentages at Suggested Final Thresholds, Selected MARS Models

	Pred = Obs @ 0.1, High Sensitivity				Specificity @ 0.67, Moderate Sensitivity			
Subarea	Threshold	% Background	% Sites	Kg	Threshold	% Background	% Sites	Kg
Region 7 All								
riverine section 6	0.52	10%	84%	0.88	0.10	32%	99%	0.68
riverine section 8	0.40	10%	83%	0.88	0.13	32%	98%	0.67
upland section 6	0.15	10%	82%	0.88	0.06	32%	91%	0.65
upland section 8	0.44	10%	80%	0.87	0.21	32%	99%	0.68
Region 8 All								
riverine section 2	0.29	10%	69%	0.86	0.07	31%	93%	0.67
riverine section 3	0.08	10%	99%	0.89	0.02	30%	99%	0.70
Region 9/10 All								
riverine section 1	0.62	10%	41%	0.76	0.35	33%	86%	0.61
riverine section 2	0.69	10%	51%	0.80	0.41	34%	87%	0.61
riverine section 5	0.83	10%	26%	0.61	0.48	34%	81%	0.58
riverine section 7	0.75	10%	50%	0.80	0.33	33%	92%	0.64
riverine section 12	0.53	10%	65%	0.85	0.21	33%	89%	0.63

Table 23 - Kg and Cell Percentages at Suggested Final Thresholds, Selected RF Models

	Pred = Obs @ 0.1, High Sensitivity				Specificity @ 0.67, Moderate Sensitivity			
Subarea	Threshold	% Background	% Sites	Kg	Threshold	% Background	% Sites	Kg
Region 7 All								
riverine section 1	0.20	10%	100%	0.90	0.07	27%	100%	0.73
riverine section 2	0.18	10%	100%	0.90	0.07	28%	100%	0.72
riverine section 3	0.20	10%	100%	0.90	0.01	24%	100%	0.76
riverine section 4	0.23	10%	100%	0.90	0.07	25%	100%	0.75
riverine section 5	0.14	9%	100%	0.91	0.01	32%	100%	0.68
riverine section 7	0.16	10%	100%	0.90	0.01	31%	100%	0.69
riverine section 9	0.24	10%	100%	0.90	0.07	31%	100%	0.69
upland section 2	0.08	7%	100%	0.93	0.01	13%	100%	0.87
upland section 3	0.10	9%	100%	0.91	0.01	27%	100%	0.73
upland section 4	0.10	10%	100%	0.90	0.01	24%	100%	0.76
upland section 5	0.12	9%	100%	0.91	0.01	30%	100%	0.70
upland section 7	0.14	10%	100%	0.90	0.01	30%	100%	0.70
upland section 9	0.27	10%	100%	0.90	0.13	30%	100%	0.70
Region 8 All								
riverine section 1	0.16	10%	100%	0.90	0.07	31%	100%	0.69
riverine section 4	0.29	10%	100%	0.90	0.15	30%	100%	0.70
riverine section 5	0.31	10%	100%	0.90	0.15	34%	100%	0.66
riverine section 6	0.30	10%	100%	0.90	0.13	32%	100%	0.68
riverine section 7	0.87	10%	72%	0.86	0.30	34%	100%	0.66
riverine section 8	0.92	10%	53%	0.82	0.36	35%	100%	0.65
riverine section 9	0.22	10%	100%	0.90	0.09	29%	100%	0.71
upland section 1	0.16	9%	100%	0.91	0.07	24%	100%	0.76
upland section 2	0.12	8%	100%	0.92	0.01	32%	100%	0.68
upland section 3	0.08	11%	100%	0.89	0.01	21%	100%	0.79
upland section 4	0.16	9%	100%	0.91	0.07	25%	100%	0.75
upland section 5	0.20	9%	100%	0.91	0.07	31%	100%	0.69
upland section 6	0.21	10%	100%	0.90	0.09	30%	100%	0.70
upland section 7	0.26	10%	100%	0.90	0.09	32%	100%	0.68
upland section 8	0.18	8%	100%	0.92	0.07	22%	100%	0.78
upland section 9	0.14	8%	100%	0.92	0.07	20%	100%	0.80
Region 9/10 All								
riverine section 3	0.98	9%	75%	0.87	0.11	32%	100%	0.68
riverine section 4	0.23	10%	100%	0.90	0.11	30%	100%	0.70
riverine section 6	0.96	10%	50%	0.80	0.38	34%	97%	0.65
riverine section 8	0.61	10%	100%	0.90	0.21	33%	100%	0.67

	Pred = Obs @ 0.1, High Sensitivity				Specificity @ 0.67, Moderate Sensitivity			
Subarea	Threshold	% Background	% Sites	Kg	Threshold	% Background	% Sites	Kg
riverine section 9	0.18	10%	100%	0.90	0.09	28%	100%	0.72
riverine section 10	0.17	10%	100%	0.90	0.01	33%	100%	0.67
riverine section 11	0.27	10%	100%	0.90	0.13	32%	100%	0.68
riverine section 13	0.15	10%	100%	0.90	0.01	30%	100%	0.70
riverine section 14	0.26	10%	100%	0.90	0.11	32%	100%	0.68
riverine section 15	0.18	10%	100%	0.90	0.07	28%	100%	0.72
upland section 1	0.14	9%	100%	0.91	0.07	26%	100%	0.74
upland section 2	0.12	9%	100%	0.91	0.01	28%	100%	0.72
upland section 3	0.14	10%	100%	0.90	0.07	24%	100%	0.76
upland section 4	0.24	10%	100%	0.90	0.09	33%	100%	0.67
upland section 5	0.12	9%	100%	0.91	0.01	25%	100%	0.75
upland section 6	0.15	10%	100%	0.90	0.07	24%	100%	0.76
upland section 7	0.15	9%	100%	0.91	0.07	23%	100%	0.77
upland section 8	0.15	10%	100%	0.90	0.07	24%	100%	0.76
upland section 9	0.17	10%	100%	0.90	0.01	28%	100%	0.72
upland section 10	0.14	10%	100%	0.90	0.01	32%	100%	0.68
upland section 11	0.18	9%	100%	0.91	0.07	29%	100%	0.71
upland section 12	0.10	11%	100%	0.89	0.01	25%	100%	0.75
upland section 13	0.14	10%	100%	0.90	0.07	20%	100%	0.80
upland section 14	0.21	9%	100%	0.91	0.11	27%	100%	0.73
upland section 15	0.18	10%	100%	0.90	0.07	31%	100%	0.69

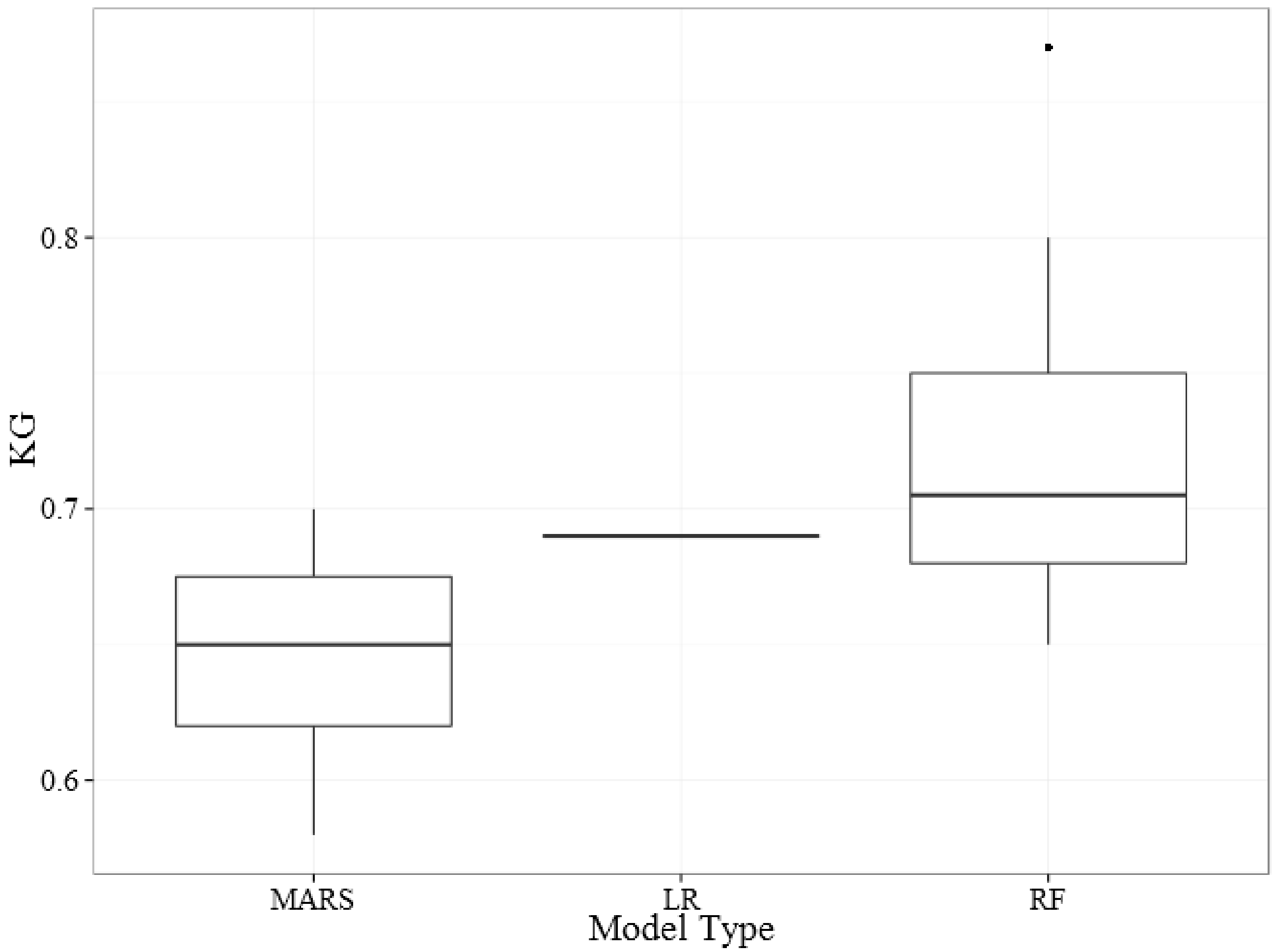


Figure 28 - Distribution of Kg statistics for each of the three model types.

Table 24 - Confusion Matrix for Site-Likely Area of Complete Regions 7, 8, 9, and 10 Selected Models

		Known Sites		
		Present	Absent	
Model Prediction	Present	970843	84046598	85017441
	Absent	14323	229961782	229976105
		985166	314008380	314993546

Sensitivity / TPR =	0.985
Specificity / TNR =	0.732
Prevalence =	0.0031
Kvamme Gain (Kg) =	0.726
Accuracy =	0.733
Positive Prediction Value (PPV) =	0.011
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.003
Positive Prediction Gain (PPG) =	3.651
Negative Prediction Gain (NPG) =	0.020
False Negative Rate (FNR) =	0.015
Detection Prevalence =	0.270

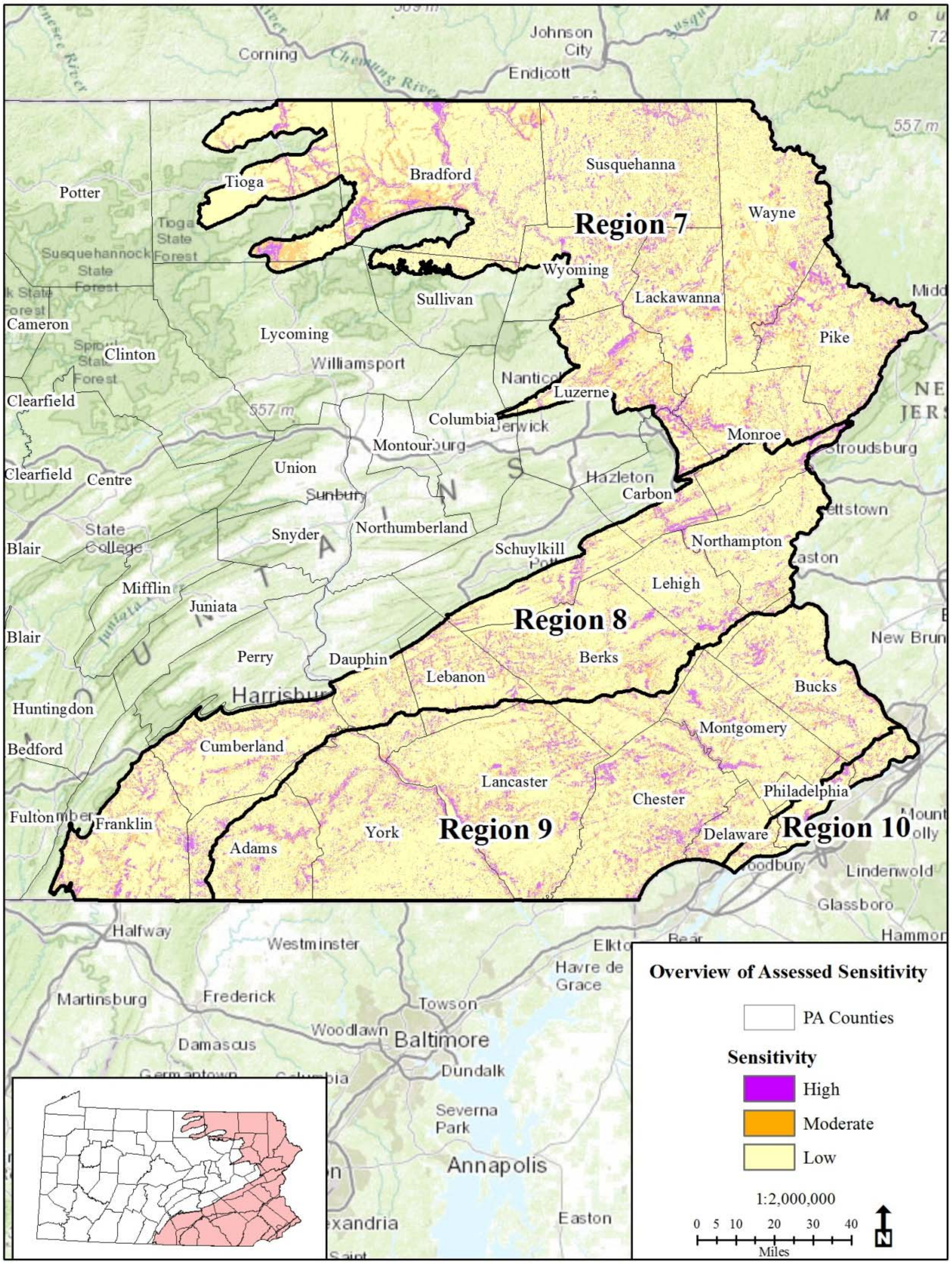


Figure 29 - Overview of assessed prehistoric sensitivity for Regions 4, 5, and 6.

7
CONCLUSIONS AND RECOMMENDATIONS

Over the course of modeling archaeological sensitivity in Regions7, 8, 9, and 10, 264 individual models were created for the 66 subareas. These included LR, MARS, RF, and proportionally weighted (Model 2) models for non-rock shelter sites and, in some subareas, for rock shelter sites as well. The total area covered by these models is 13,701 square miles, constituting all of eastern Pennsylvania. The methodology used to create these models involved the preparation of PASS site data, the development of 93 individual environmental variables, and the division of the regions into 66 separate subareas. Through the testing of each of the variables against the environmental background of each subarea, the parameterization and validation of statistical models, creation of additional models where there are few known sites or high proportions of rock shelters, and the final model selection based on error estimate results, Kg, and other metrics, a total of 66 models was selected from the candidates. The establishment of numerous potential thresholds based on variable criteria, and, finally, the application of selected thresholds and mosaicking of 66 separate subarea models into the final model for each of the regions completed the task. The end result is a model of all four regions that correctly classifies 98.5% of known site-present cells within 26.8% of the study area, for a Kg of 0.726. In actuality, the model is capable of correctly predicting the location of all archaeological sites and minimizing the site-likely area to a much smaller percent of the study area, but the selection of a low-end threshold for the site-likely area was intentionally set to approximately 33% of the study area. Compared to a random survey, the chances of finding a site in the combined high and moderate sensitivity area are 3.651 times greater.

The final 66 subarea models created for Regions 7, 8, 9, and 10 are derived from a variety of model types, including LR, MARS, and RF statistical models. Each of these models has their own strengths, weaknesses, and assumptions, as well as ability to address the bias-variance tradeoff that is amplified when using correlated environmental variables and often sparse site location data. However, each model type has been shown to be effective at identifying the patterns within known site locations and extrapolating that pattern to landforms that share similar characteristics. Further, each type of model has different abilities in addressing variations in data quality and sample size issues. Each of the statistical models is capable of providing internal metrics that offer information on the model's prediction errors and qualities of fit.

The results of the internal prediction error rate tests on the 10-fold CV samples (average RMSE = 0.176 for the LR models and an accuracy of 97.5% for MARS and RF models) and an average RMSE of 0.122 for all models on the held-out sample demonstrate that these models are capable of accurately predicting site-present cells that were not part of the model-building sample. This adds confidence that these models are not only able to identify landforms that the test sites are found on, but can also extrapolate this pattern to site locations outside of the test set. The suite of validation and testing statistics presented in the previous chapters all agree that these models are a good representation of the site sample from previously identified prehistoric archaeological sites. Further, these models better approximate a more realistic prevalence of prehistoric sites than previous and more generalized models. With the choice of classification thresholds that are appropriate for the particular management or research objective, these models should be valid and accurate tools to assist in project planning and sensitivity analyses.

All of the recommendations made in previous reports were addressed in this study, but none of them directly impacted the results presented above. The first recommendation in the Task 5 report concerned the incorporation of class weights and thresholds within the RF model to attempt to reduce model variance and increase generalizability of the results. This recommendation has carried over since Task 4 of this study. The concept was approached again in Task 6, but a solution to implementing class weights effectively and consistently from the model fitting to raster prediction stages was not found. This issue is larger than this particular study as the author of the statistical implementation of the RF algorithm is currently working on a solution. The testing done thus far on this statistical feature shows some promise for better addressing the severe class imbalance issues of our data. Further developments in the implementation of the RF algorithm may make this feature more efficient, in which case it should be tested in future modeling efforts. The second recommendation of the Task 5 report was to create proportionally weighted models for each of the subareas. This recommendation was followed in Task 6. Although none of these models was ultimately used, they will be part of the final deliverable.

The final recommendation of the Task 5 report was to experiment with additional statistical model types to compare to the current results. As stated throughout, the statistical models of Logistic Regression (LR), Multivariate Adaptive Recursive Splines (MARS), and Random Forest (RF) were selected for this project. These models were selected for a number of reasons including LR’s many previous uses in APM studies, the ability of MARS to handle nonparametric data and feature selection while still being understandable in the context of linear regression, and RF for its ability to classify noisy data, internal feature selection, and boosting with little parameterization. In the many fields that use modern computational techniques and mathematical statistics, however, there are a number of additional model types that have these features and additional capabilities not represented here. While it is not advisable to blindly search for a model technique that fits a particular dataset, it would be beneficial to test additional methods to see if their strengths could benefit the character of archaeological locational data. The design of the modeling framework developed during this project allows for additional model types to be plugged in and run on the data without requiring much additional effort, aside from the nontrivial effort required to understand and parameterize a new model type. During the processing of data for the Task 6 models, we began the effort to build a test bed to compare additional model types to the current types of LR, MARS, and RF. This effort will continue into the final Task 7 report where the results of the model comparison will be published.

This report concludes the effort to develop and apply a series of statistical models to the archaeological location data (i.e., PASS file data) across the Commonwealth of Pennsylvania. This substantial effort required thousands of computer hours to clean, process, tabulate, model, and predict many terabytes of data. A total of 4 models (Model 2, LR, MARS, and RF) were created for each of the 132 subareas by which we divided the state. Each model type covers an area of nearly 45,000 miles squared, totaling nearly 180,000 square miles of model coverage. At a grid cell resolution of ~10 × 10 m, this equates to approximately 1 billion raster cells covering the state for each model types or nearly 4 billion raster cells considering all four model types. This is in addition to the nearly 90 billion raster cells for the environmental attribute layers used to fit the models. Along with these massive data sets, this project required the calculation of many millions of statistical tests. These included the K-S and MW tests to differentiate site-present and absent distributions, estimation of error rates for many thousands of cross-validation steps, fitting of models to the data, evaluation of model fit results, prediction of raster cells, calculation of appropriate thresholds, and creation of confusion matrices to display results. Some of these individual model fits required upwards of 16 hours per model per subarea to complete; multiplied by the 132 subareas across the state. In order to complete the thousands of computer hours required to produce these analytical results, techniques such as parallel processing and resources such as using remote high powered servers and storage (i.e., “Cloud” computing) were required. The details of this modeling process, statistical methods, and analytical results are contained within this and the previous five reports. The final report of this project (Task 7) will summarize the process detailed above, expand on the character of the archaeological location data, compare results to additional model types, provide a roadmap for the software framework that made this effort possible, illuminate the understanding of APM and site data gained through this process, and provide recommendations for next generation modeling efforts.

8
REFERENCES CITED

Affleck, Richard, Bradford Botwick, and Ingrid Wuebber

1994 Cultural Resources Investigation, Beltzville Lake, LehightonVicinity, Carbon and Monroe Counties, Pennsylvania. Reported by The Cultural Resources Group, Louis Berger & Associates, East Orange, New Jersey, to U.S. Army Corps of Engineers, Philadelphia District, Philadelphia, Pennsylvania.

Akaike, Hirotugu

1974 A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19(6):716–723.

Bergman, Christopher, John Doershuk, Roger Moeller, Philip LaPorta, and Joseph Schulderein

1998 An Introduction to the Early and Middle Archaic Occupations at Sandts Eddy. In *The Archaic Period in Pennsylvania: Hunter-Gatherers of the Early and Middle Holocene Period*, edited by Paul A. Raber, Patricia E. Miller, and Sarah M. Neusius, pp. 45–76. Pennsylvania Historical and Museum Commission, Harrisburg.

Binford, Lewis R.

1980 Willow Smoke and Dogs' Tails: Hunter-Gatherer Settlement Systems and Archaeological Site Formation. *American Antiquity* 45(1):5–20.

Botwick, Bradford, and Robert D. Wall

1994 Prehistoric Settlement in the Uplands of the Upper Delaware Valley: Recent Surveys and Testing in the Delaware Water Gap National Recreation Area. *Journal of Middle Atlantic Archaeology* 10:63–73.

Boyd, Varna G., Gary F. Coppock, Kathleen A. Ferguson, Benjamin R. Fischler, Bernard K. Means, and Frank Vento

2000 Prehistoric Archaeological Synthesis: The U.S. 219 Meyersdale Bypass Project. U.S. 219 Meyersdale Bypass Project, S.R. 6219, Section B08, Somerset County, Pennsylvania. Reported by Greenhorne & O'Mara, Inc., Mechanicsburg, Pennsylvania, to Pennsylvania Department of Transportation, Engineering District 9-0, Hollidaysburg.

Breiman, Leo

2001 Random Forests. *Machine Learning* 45(1):5–32.

Carr, Kurt W.
1998 Archaeological Site Distribution and Patterns of Lithic Utilization During the Middle Archaic in Pennsylvania. In *The Archaic Period in Pennsylvania: Hunter-Gatherers of the Early and Middle Holocene Period*, edited by Paul A. Raber, Patricia E. Miller, and Sarah M. Neusius, pp. 77–90. Pennsylvania Historical and Museum Commission, Harrisburg.

Carr, Kurt W., and James M. Adovasio
2002 Paleoindians in Pennsylvania. In *Ice Age Peoples of Pennsylvania*, edited by Kurt Carr and James Adovasio, pp. 1–50. Pennsylvania Historical and Museum Commission, Recent Research in Pennsylvania Archaeology No. 2. Harrisburg.

Conover, W, J.
1999 *Practical Nonparametric Statistics*. 3rd ed. Wiley, New York, NY.

Custer, Jay F.
1996 *Prehistoric Cultures of Eastern Pennsylvania*. Pennsylvania Historical and Museum Commission, Anthropological Series No. 7. Harrisburg.

Dekin, Albert A.
1980 A Survey of Archaeology, History, and Cultural Resources in the Upper Delaware National Scenic and Recreational River, Pennsylvania and New York States. Reported by Public Archaeology Facility, Department of Anthropology, State University of New York, Binghamton, to National Park Service, Mid-Atlantic Region, Philadelphia, Pennsylvania.

Dent, Richard J.
2002 Paleoindian Occupation in the Upper Delaware Valley: Revisiting Shawnee-Minisink and Nearby Sites. In *Ice Age Peoples of Pennsylvania*, edited by Kurt Carr and James Adovasio, pp. 51–78. Pennsylvania Historical and Museum Commission, Recent Research in Pennsylvania Archaeology No. 2. Harrisburg.

Dragoo, Don W.
1976 Some Aspects of Eastern North American Prehistory: A Review, 1975. *American Antiquity* 41(1):3–27.

East, Thomas C., Frank J. Vento, Christopher T. Espenshade, Margaret G. Sams, and Brian C. Henderson
2002 Phase I/II/III Archaeological Investigations, Northumberland and Union Counties, S.R. 0080, Section 52D, Bridge Expansion and Highway Improvement Project. E.R. # 1999-8000-042. Report prepared by Skelly and Loy, Inc., Monroeville, Pennsylvania and submitted to the Pennsylvania Department of Transportation, Engineering District 3-0, Montoursville, Pennsylvania.

Efron, B., and R. Tibshirani

1997 Improvements on Cross-Validation: The .632 + Bootstrap Method. *Journal of the American Statistical Association* 92(438):548–560.

Fawcett, Tom

2004 ROC Graphs: Notes and Practical Considerations for Researchers. *Pattern Recognition Letters* 27(8):882–891.

2006 An Introduction to ROC Analysis. *Pattern Recognition Letters* 27(2006):861–874.

Freeman, Elizabeth A., and Gretchen G. Moisen

2008 A Comparison of the Performance of Threshold Criteria for Binary Classification in Terms of Predicted Prevalence and Kappa. *Ecological Modeling* 217:48-58.

Friedman, J. H.

1991 Multivariate Adaptive Regression Splines. *The Annals of Statistics* 19:1.

Gingerich, Joseph A.M.,

2007 Shawnee-Minisink Revisited: Re-evaluating the Paleoindian Occupation. Unpublished M.A. Thesis, Department of Anthropology, University of Wyoming, Laramie.

Griffin, James B.

1967 Eastern North American Archaeology: A Summary. *Science* 156(3772):175–191.

Harris, Matthew D.

2013a Pennsylvania Department of Transportation Archaeological Predictive Model Set, Task 1: Literature Review. Prepared for Pennsylvania Department of Transportation, Bureau of Planning and Research, Harrisburg. URS Corporation, Burlington, New Jersey.

2013b Pennsylvania Department of Transportation Archaeological Predictive Model Set, Task 2: Designating Modeling Regions. Prepared for Pennsylvania Department of Transportation, Bureau of Planning and Research, Harrisburg. URS Corporation, Burlington, New Jersey.

2014 Pennsylvania Department of Transportation Archaeological Predictive Model Set, Task 3: Pilot Model Study. Prepared for Pennsylvania Department of Transportation, Bureau of Planning and Research, Harrisburg. URS Corporation, Burlington, New Jersey.

Harris, Matthew D., Tod L. Benedict, and Douglas C. MacVarish

2010 Phase I and II Archaeological Survey, SCI Graterford Prison Expansion, and Phase III Analysis of an Early Woodland Procurement Site at 36MG0443, Graterford, Montgomery

County, Pennsylvania. ER # 95-2780-091-I Reported by John Milner Associates, Inc., Philadelphia, Pennsylvania, to Hill International, Inc., Philadelphia, Pennsylvania.

Harris, Matthew D., George C. Cress, Kimberly Morrel, Patricia Miller, and Jennifer Rankin
2014c Phase II Archaeological Excavation at 36PH14, South Meadow Area of Historic Bartram's Garden, Philadelphia, Pennsylvania. Reported by URS Corporation, Burlington, New Jersey, to City of Philadelphia, Division of Aviation.

Harris, Matthew D., Susan Landis, and Andrew R. Sewell
2014a Pennsylvania Department of Transportation Archaeological Predictive Model Set, Task 4: Study Regions 1, 2, and 3. Prepared for Pennsylvania Department of Transportation, Bureau of Planning and Research, Harrisburg. URS Corporation, Burlington, New Jersey.

2014b Pennsylvania Department of Transportation Archaeological Predictive Model Set, Task 5: Study Regions 4, 5, and 6. Prepared for Pennsylvania Department of Transportation, Bureau of Planning and Research, Harrisburg. URS Corporation, Burlington, New Jersey.

Hart, John P. (editor)
1995 Archaeological Investigations at the Memorial Park Site (36CN164), Clinton County, Pennsylvania. E.R. # 1981-0405-035. Report prepared GAI Consultants, Inc., Pittsburgh for the Baltimore District of the U.S. Army Corps of Engineers.

Hatch, James A.
1994 The Structure and Antiquity of Prehistoric Jasper Quarries in the Reading Prong, Pennsylvania. *Journal of Middle Atlantic Archaeology* 10:21–42.

Hay, Conrad A.
1993 Predictive Model for Archaeological Resources, U.S. Route 202, Section 700, Bucks and Montgomery Counties, Pennsylvania. Reported by Archaeological and Historical Consultants, Inc., Centre Hall, Pennsylvania, to Pennsylvania Department of Transportation, Engineering District 6-0, King of Prussia, Pennsylvania.

Hunter, Richard W., and Ian C. G. Burrows
1990 Cultural Resources Investigations, The Francis E. Walter Dam and Reservoir Modification Project, Kidder Township, Carbon Township, Bear Creek and Buck Townships, Luzerne County, and Tobyhanna Township, Monroe County, Pennsylvania. Reported by Hunter Research, Inc., Trenton, New Jersey, to U.S. Army Corps of Engineers, Philadelphia District, Philadelphia, Pennsylvania.

Jefferies, Richard W.

1990 Archaic Period. In *The Archaeology of Kentucky: Past Accomplishments and Future Directions*, edited by D. Pollack, pp. 143–246. Kentucky Heritage Council, Frankfort.

1996 Hunters and Gatherers after the Ice Age. In *Kentucky Archaeology*, edited by R. Barry Lewis, pp. 39–78. University of Kentucky Press, Lexington.

Justice, Noel D.

1987 *Stone Age Spear and Arrow Points of the Midcontinental and Eastern United States: A Modern Survey and Reference*. Bloomington, Indiana: Indiana University Press.

1995 *Stone Age Spear and Projectile Points of the Midcontinental and Eastern United States*. Indiana University Press, Bloomington.

Kingsley, R.G., with J.A. Robertson, and D.G. Roberts

1990 The Archeology of the Lower Schuylkill River Valley in Southeastern Pennsylvania. Philadelphia Electric Company, Philadelphia, Pennsylvania.

Kingsley, Robert G., Joseph Schulderein, James A. Robertson, and Daniel R. Hayes

1991 Archaeology of the Lower Black's Eddy Site, Bucks County, Pennsylvania: Final Report. Reported by John Milner Associates, West Chester, Pennsylvania, to County of Bucks, Office of the Commissioners, Doylestown, Pennsylvania.

Kratzner, Judson, Richard White, and Paul D. Schopp

2008 Phase Ib/II Archaeological Investigation, Sugarhouse Casino Site (36Ph137), 941–1025 North Delaware Avenue, City of Philadelphia, Philadelphia County, Pennsylvania. Reported by A. D. Marble & Company, Conshohocken, Pennsylvania, to HSP Gaming, L. P., c/o Keating Consulting, LLC, Philadelphia, Pennsylvania.

Kvamme, Kenneth L.

1988 Development and Testing of Quantitative Models. In *Quantifying the Present and Predicting the Past*, edited by W. Judge and L. Sebastian, pp. 325-428. U.S. Government Printing Office, Washington, D.C.

Lawrence, John. W., and Brain M. Albright

2012 Data Recovery Archaeological Investigation Report, River Road Site (36BU379), I-95/Scudder FallsBridge Improvement Project, Lower Makefield Township, Bucks County, Pennsylvania: Volume I. Reported by AECOM Transportation, Trenton, New Jersey, to Delaware River Joint Toll Bridge Commission, New Hope, Pennsylvania.

Lehmann, Erich L.

1975 Nonparametrics: Statistical Methods Based on Ranks. Holden-Day, Oakland.

1986 *Testing Statistical Hypothesis*. 2nd ed. Wiley, New York, NY.

Liaw, Andy, and Matthew Wiener

2002 Classification and Regression by randomForest. *R News* 2(3):18–22.

MacDonald, Douglas H.

2003 Pennsylvania Archaeological Data Synthesis: The Upper Juniata River Sub-Basin (Watersheds A-D). E.R. # 00-2888-013. Walter Industrial Park: Mitigation of Adverse Effects, U.S. Department of Commerce, Economic Development Administration, Greenfield Township, Blair County, Pennsylvania. Reported by GAI Consultants, Inc., Monroeville, Pennsylvania, to Keller Engineers, Inc., Hollidaysburg, Pennsylvania.

McCracken, Richard J.

1989 The Trojan Site (36BR149): A Preliminary Report on a Paleo Indian Manifestation in Bradford County, Pennsylvania. *Pennsylvania Archaeologist* 59 (2): 1–21.

Milanich, Jerald T.

1994 *Archaeology of Precolumbian Florida*. University Press of Florida, Gainesville.

Milborrow, Stephen

2014 Notes on the Earth Package. Electronic document: http://cran.r-project.org/web/packages/earth/vignettes/earth-notes.pdf.

Miller, Patricia E., and Marcia M. Kodlick

2006 Archaeological Predictive Model Test Results, PA 23 EIS Project, SR 0023 Section EIS, Lancaster County, Pennsylvania. E.R.# 03-8015-071. Reported by KCI Technologies, Inc., Mechanicsburg, Pennsylvania, for Pennsylvania Department of Transportation, Engineering District 8-0, Harrisburg.

Miller, Patricia E., Frank J. Vento, and James T. Marine

2007 Archaeological Investigations, Susquehanna River Bridge, Dauphin and York Counties, Pennsylvania. Volume I: Research Design and Prehistoric Context. ER No. 1999-1130-042. Reported by KCI Technologies, Inc., Mechanicsburg, Pennsylvania, to Pennyslvania Turnpike Commission, Harrisburg, Pennsylvania.

Mn/Model

n.d. Project Background.Electronic document: http://www.dot.state.mn.us/mnmodel/about/history.html, Accessed May 8, 2014.

Mooney, Douglas B., Rose L. Moore, Philip A. Perazio, Niles R. Rinehart, and James P. Davis

2003 Phase I Cultural Resource Investigation of the Planned Bushkill Road Schools Complex Project Area, Lehman Township, Pike County, Pennsylvania. E.R. # 95-3070-103-H. Reported by Kittattiny Archaeological Research, Inc., Stroudsbourg, Pennsylvania, for F. X. Browne, Inc., Landsdale, Pennsylvania.

Oehlert, Gary W., and Brian Shea

2007 *Statistical Methods for Mn/Model Phase 4*. Research Services Section of Minnesota Department of Transportation, St. Paul.

Pampel, F. C. (editor)

2000 *Logistic Regression: A Primer.* Vol. 132. Sage, Thousand Oaks, CA.

Pan Cultural Associates, Inc.

2005 Salem Township ACT 537 Sewer Facilities Plan Update Revision, Salem Township, Luzerne County, Pennsylvania: Phase I/II Archaeological Investigation of Sites 38LU0191 & 38LU0270. Reported by Pan Cultural Associates, Inc., Pittston, Pennsylvania, to Salem Township, Luzerne County, Pennsylvania

Petyk, Richard, Dane Snyder, and Nick Avery

2010 Final Phase I and II Investigations, Results of Cultural Resource Survey for the Proposed 300 Line Project in Pennsylvania. E.R. #08-2473-042. Reported by Gray & Pape, Inc., Providence, Rhode Island, to AECOM Environment, Sagamore Beach, Massachusetts.

Prufer, Olaf H., and Dana A. Long

1986 *The Archaic of Northeastern Ohio*. Kent State University Press, Kent, Ohio.

Quinn, Allen G., with contributions by Judith E. Thomas and David C. Hyland

1994 Phase I Archaeological reconnaissance of the Shades Beach Park Study Area: A Report of the Pennsylvania Department of Environmental Resources to the National Oceanic and Atmospheric Administration Pursuant to NOAA Award No. NA370Z0351. Report to Harborcreek Township, Harborcreek, Pennsylvania, from Mercyhurst Archaeological Institute, Erie, PA. U.S. Government Printing Office <http://www.gpo.gov/fdsys/pkg/CZIC-qh76-5-h3-q56-1994/html/CZIC-qh76-5-h3-q56-1994.htm>. Accessed 3 January 2014.

Raber, Paul A.,

2003 Problems and Prospects in the Study of the Early and Middle Woodland Periods. In *Foragers and Farmers of Early and Middle Woodland Periods in Pennsylvania*, edited by Paul A. Raber and Verna L. Cowin, pp. v–vii. Pennsylvania Historical and Museum Commission, Recent Research in Pennsylvania Archaeology No. 3. Harrisburg.

2010 Chert Use and Local Settlement in Central Pennsylvania: Investigations at 36Ce523. *Journal of Middle Atlantic Archaeology* 26:115–140.

Raber, Paul A., Patricia E. Miller, and Sarah M. Neusius
1998 The Archaic Period in Pennsylvania: Current Models and Future Directions. In *The Archaic Period in Pennsylvania: Hunter-Gatherers of the Early and Middle Holocene Period*, edited by Paul A. Raber, Patricia E. Miller, and Sarah M. Neusius, pp. 121–137, Pennsylvania Historical and Museum Commission, Harrisburg.

Robertson, James A., and Robert G. Kingsley
1994 The Lower Black's Eddy Site and Prehistoric Settlement Systems in the Middle Delaware Valley. *Journal of Middle Atlantic Archaeology* 10:75–90.

Salkind, Neil J. (editor)
2007 *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, CA.

Sassaman, Kenneth E.
1993 *Early Pottery in the Southeast: Tradition and Innovation in Cooking Technology*. The University of Alabama Press, Tuscaloosa.

Siegel, Peter E., Tod L. Benedict, and Robert G. Kingsley
1999 Archaeological Data Recovery at the Fahs II and Oberly Island sites: Structure, Function, and Context in the Lower Lehigh Valley. Northampton County, Pennsylvania, S.R. 0033, Section 0001. E.R. # 88-0224-095. Reported by John Milner Associates, West Chester, Pennsylvania, to URS Grenier Woodward Clyde, King of Prussia, Pennsylvania.

Siegel, Peter E., Douglas C. Kellogg, and Robert G. Kingsley
2001 Prehistoric Settlement Patterns in Upland Settings: An Analysis of Site Data in a Sample of Exempted Watersheds, Brandywine Creek Watershed (Watershed H), Chester, Lancaster, and Delaware Counties, Pennsylvania. Prepared for the Pennsylvania Historical and Museum Commission under a Historic Preservation Grant awarded to the Pennsylvania Archaeological Council. http://home.earthlink.net/~pacweb/spframeset.html. Accessed 10 January 2014.

Stafford, C. Russell
1994 Structural Changes in Archaic Landscape Use in the Dissected Uplands of Southwestern Indiana. *American Antiquity* 59(2):219–237.

Stewart, R. Michael, and John Cavallo
1991 Delaware Valley Middle Archaic. *Journal of Middle Atlantic Archaeology* 7: 19-42.

Stewart, Michael and Judson Kratzer
1989 Prehistoric Site Locations on the Unglaciated Appalachian Plateau. *Pennsylvania Archaeologist* 59 (1):19¬36.

Stewart, Michael, and William Schindler
2008 Analysis of Artifacts from the Kings Quarry Site (36LH12). Reported by The Center for Experimental Archaeology, Church Hill, Maryland, to the Pennsylvania Historical and Museum Commission, Harrisburg, and the Pennsylvania Heritage Society, Harrisburg.

Viera, Anthony J., and Joanne M. Garrett
2005 Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine* 37(5):360-363.

Walker, Jesse O., Michael Tomkins, and Robert J. Lore
2012 Phase I/II Archaeology Survey, Vera Cruz Sewer Project, Upper Milford Township, Lehigh County, Pennsylvania. Reported by Richard Grubb & Associates, Inc., Allentown, Pennsylvania, to SSM Group, Inc., Bethlehem, Pennsylvania.

Wallace, Paul A.W.
1965 Indian Paths of Pennsylvania. Pennsylvania Historical and Museum Commission, Harrisburg.

Wheatley, David, and Mark Gillings
2002 *Spatial Technology and Archaeology. The Archaeological Application of GIS*. Taylor & Francis, London, UK.

Wilson, Edward W., and W. Fred Kinsey
1982 Archaeological Site Survey of East-Central Pennsylvania: July 1, 1981 – June 30, 1982. Reported by Pennsylvania Historical and Museum Commission, Harrisburg, for same.

Wyatt, Andrew
2003 Early and Middle Woodland Settlement Data for the Susquehanna Basin. In *Foragers and Farmers of Early and Middle Woodland Periods in Pennsylvania*, edited by Paul A. Raber and Verna L. Cowin, pp. 35–48. Pennsylvania Historical and Museum Commission, Recent Research in Pennsylvania Archaeology No. 3. Harrisburg.

Wyatt, Andrew, Robert H. Eiswert, Richard C. Petyk, and Richard T. Baublitz
2005 Phase III Archaeological Investigations at the Raker I Site (36Nb58), Route 147 Climbing Lane Project, Upper Augusta Township, Northumberland County, Pennsylvania. E.R. # 2000-6173-097. Prepared by McCormick Taylor, Inc., Harrisburg for the Pennsylvania Department of Transportation, Engineering District 3-0, Montoursville.

APPENDIX A

ACRONYMS AND GLOSSARY OF TERMS

ACRONYMS

AIC	Akaike Information Criterion
APM	Archaeological Predictive Modeling
AUC	Area Under Curve
CoV	Coefficient of Variation
CRGIS	Cultural Resources Geographic Information System
CV	Cross-Validation
GCV	Generalized Cross-Validation
GIS	Geographic Information Systems
Kg	Kvamme Gain
K-S	Kolmogorov–Smirnov
LR	Logistic Regression
MARS	Multivariate Adaptive Regression Splines
MW	Mann-Whitney
NPG	Negative Prediction Gain
NPV	Negative Prediction Value
PASS	Pennsylvania Archaeological Site Survey
PPG	Positive Predictive Gain
PPV	Positive Prediction Value
RF	Random Forests/randomForest
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristics
TNR	True-Negative Rate
TPR	True-Positive Rate
UDR	Unexpected Discovery Rate

TERMS

page in report text
(first used)

Accuracy (in error estimates for MARS and RF models) 65

The measurement of accuracy is used in many classification methods. This measure is simply the percent of observations (site-present or site-absent) that are correctly classified by the algorithm. As used in this report, the accuracy is the percentage of observations from the out-of-bag sample that were correctly classified by the model. This is an internal metric that assess the model's ability to correctly predict data that were not used in the fitting of the model.

Adaptive Regression Splines (see Multivariate Adaptive Regression Splines) 1

Akaike Information Criterion (AIC)..62

A measure of relative model quality that balances goodness of fit and model complexity. This measure is used in model selection to choose the model that has the best fit relative to complexity for a given data set. Within a series of nested candidate models, the one with the lowest AIC will likely represent the model with the best goodness of fit without being over-fit or over-parameterized (see Akaike 1974).

Archaeological Predictive Modeling (APM)...1

The field of study concerning the use of existing archaeological data or theory to predict the sensitivity of locations for the presence of archaeological material.

Area Under Curve (AUC) (see also Receiver Operating Characteristics) 59

Also referred to as Area Under Receiver Operating Characteristics Curve (AUROC), AUC is a measure of the balance between a model's Sensitivity and Specificity across the full range of cut-off points. The AUC is a single measure that captures a model's ability to balance True Positive Rate and False Positive Rate across the full range of the model's output. The higher the AUC, the higher the Sensitivity and Specificity across the full range of the model, and the more likely the model is to correctly classify a randomly chosen positive instance. AUC is used in model selection to assess a model's ability to correctly classify observations (see Fawcett 2006).

Confusion Matrix 82

A classification table in the form of a 2-cell × 2-cell contingency table that shows how many sites were correctly predicted as sites and how much of the non-site area was correctly predicted as such. This method is frequently used as a means to assess the ability of a model to classify observations (see Fawcett 2006).

Coefficient of Variation (CoV) 65

The CoV is a statistic that measures the normalized dispersion within a frequency distribution. The acronym CoV is used in this study to avoid confusion with the acronym used for Cross-Validation (CV). The CoV is calculated as the ratio of the standard deviation to the mean and is also referred to as Relative Standard Deviation (RSD). The CoV represents the percentage of standard deviation from the sample mean (see Lehmann 1986).

Cohen's Kappa Coefficient (see Kappa)...71

Cross-Validation (CV) (see Generalized Cross Validation and K-folds Cross-Validation)..........62

Cultural Resources Geographic Information System (CRGIS) 46

Computerized database and mapping tool for the visualization and analysis of cultural resources data within the Commonwealth of Pennsylvania. This tool is developed and administered through a join agreement between the Pennsylvania Historical and Museum Commission and the Pennsylvania Department of Transportation. (This tool is available at: www.portal.state.pa.us/portal/server.pt/community/crgis/3802.)

False Negative Rate (FNR)...87

The fraction of the positive observation (site locations) that are incorrectly classified as a negative observation (site not-likely). The FNR is derived from the Confusion Matrix and calculated by dividing the number of false negatives by total number of observed positive observations. This number is also interpreted as the Type-II error rate, or beta (β).

Generalized Cross-Validation (GCV)..62

GCV is a statistical method that estimates performance or prediction error from within a model based on weight assigned to model complexity. GCV approximates the measure of performance that would be derived through leave-one-out Cross-Validation. In this project, the GCV relates to the internal performance measure derived from the Multivariate Adaptive Regression Splines model (see Milborrow 2014).

Geographic Information Systems (GIS) ....n/a

A GIS is a computer application that stores, manages, displays, and manipulates information with a spatial component (see Wheatley and Gillings 2002).

K-folds Cross-Validation 62

Cross-Validation is the method by which a sample of observations is split into a number of different but equal-sized classes. The number of classes is referred to as K and the classes themselves are referred to as folds, hence "K-folds Cross-Validation." This is a method by which models can be validated on test sets that were not part of the training set, while at the same time, using the entire data set for modeling (see Efron and Tibshirani 1997).

Kappa coefficient 71

The Kappa coefficient, or Cohen's Kappa coefficient, is a statistical measure of a predictions agreement with real observations after accounting for chance agreement. In this project, the Kappa is used in a similar fashion as the Kvamme Gain statistic. However, the Kappa's calculation of by-chance observation is more inclusive that the Kvamme Gain. The Kappa statistic is derived from the confusion matrix and is used to compare model results of similar prevalence (see Viera and Garrett 2005).

Kolmogorov–Smirnov (K-S) Test 61

A non-parametric statistical test that measures the equality of continuous unpaired probability distributions to each other (two-sample test) or a reference distribution (one-sample test). In this study, the K-S test is used to test whether the distribution of an environmental variable is significantly different between known site locations and the overall environmental background (see Conover 1999).

Kvamme Gain (Kg)..1

The Kg is a metric used to assess the ability of a model to correctly classify positive observations (site present) given the area in which positive observations are predicted to occur (site-likely area). The higher the gain, the greater the ratio of percent sites present to percent of the modeled area considered site-likely. This measure does not take into account model precision or True Positive Rate (Sensitivity), meaning that an equivalent Kg statistic can be reached by correctly predicting 16% of known sites in 5% of the area or 95% of known sites in 30% of the area (see Kvamme 1988).

Logistic Regression (LR)...1

Logistic Regression is a statistical model used to predict for a binary response (0 or 1) or to classify a categorical response ("dead" or "alive") based on one or more predictors. This method uses a S-shaped logistic transformation to model the binary response probability as the log odds of the linear function of the predictor variables. Simply, the model fits the linear model to the S-shaped curve so that the prediction is kept between 0 and 1 (see Pampel 2000).

Mann-Whitney (MW) U Test 61

The Mann-Whitney U Test is a non-parametric statistical test that evaluates the dissimilarity of unpaired distributions by ranking the observations and comparing the mean ranks. This test is similar in concept to the Kolmogorov–Smirnov Test, but uses a ranked approach as opposed to a distance approach. The MW U Test is more sensitive to changes in the median of two distributions (see Lehman 1975).

mtry 62

This is the name of a key parameter in the RF model. One of the key features of RF is the random selection of a subset of the predictor variables to test at each node in the tree building process. The number of randomly selected variables to try is called "*mtry*.. By default, *mtry* is set to $\sqrt{p}$ for classification problems and $p/3$ in regression problems. In this project, *mtry* is optimized through cross-validation to the lowest error rate of the out-of-fold sample.

Multivariate Adaptive Regression Splines (MARS)..1

A statistical model that is an extension of the Generalized Linear Model. This method approximates a non-linear model by fitting piecewise linear segments that are connected at nodes referred to as hinge functions. The hinge functions provide the point at which the two straight lines join. A sequence of lines and hinges approximates a non-linear Spline. The MARS model uses a forward pass to find the best fit that minimizes the Sum of Squared Error. This first pass is referred to as "greedy" because it seeks the best fit regardless of how many terms, or line and hinge segments, it creates. To avoid over-fitting, the MARS method has a second pass that prunes the terms created in the first path to assess which can be removed without having large negative effects on the model's performance; this lowers the model's complexity and variance. The MARS method uses Generalized Cross-Validation to assess how pruning affects performance. This method was introduced by Friedman (1991).

Negative Prediction Gain (NPG) 87

The NPG is a statistic that is derived from the confusion matrix to assess a model's ability to correctly classify site-unlikely areas. The NPG quantifies how much less likely a site discovery is at a location labeled site-unlikely using the model than if surveying at random. Ideally, a model would have a low NPG and a high Positive Predictive Gain (see Oehlert and Shea 2007).

Negative Prediction Value (NPV) 87

The NPV is a measure that is derived from the confusion matrix. This measures the probability that a non-site cell is correctly labeled as a background cell (see Oehlert and Shea 2007).

nprune 62

This is the name of a key parameter in the MARS model. This algorithm includes a backwards pass that prunes the model down to reduce variance and eliminate unneeded model terms. The *nprune* parameter is used to set the *maximum* number of terms that are allowed to remain in the model; the fewer terms, the more simple the model. Through this parameter, models can be trimmed for the purpose of model size, complexity, or generality of the fit. By default, *nprune* is set to NULL so that the model is unrestrained in the number of terms. For this project, the *nprune* parameter is set through cross-validation to the lowest error rate of the out-of-fold sample.

Pennsylvania Archaeological Site Survey (PASS)...1

The PASS files are a collection of paper forms, maps, reports, and photographs that document the location and attributes of known archaeological sites within the Commonwealth of Pennsylvania. These files have been digitized and can be accessed through the Cultural Resources Geographic Information System.

Positive Predictive Gain (PPG)...87

The PPG is a statistic that is derived from the Confusion Matrix to assess a model's ability to correctly classify site-likely areas. The PPG quantifies how much more likely a site discovery is at a location labeled site-likely using the model than if surveying at random. Ideally, a model would have a high PPG and a low Negative Prediction Value (see Oehlert and Shea 2007).

Positive Prediction Value (PPV)..87

The PPV is a measure that is derived from the Confusion Matrix. This measures the probability that a site cell is correctly labeled as a site-likely cell (see Oehlert and Shea 2007).

Prevalence 77

Prevalence is the proportion of a population found to have a particular condition. In this case, the population is the total number of ~10 × 10-m raster cells that make up each subarea and the condition is that a cell be within a known archaeological site. Determining prevalence is important in these models because the low number of cells within known archaeological sites is very small compared to the overall area being predicted, leading to highly imbalanced data in terms of site-presence versus site-absence.

Random Forests ....n/a

Random Forests is trademarked statistical classification algorithm created by Leo Breiman and Adele Cutler. Random Forests is a tree based ensemble method that builds off the ideas of Classification and Regression Trees and Bagging. The primary features of Random Forests include internal testing through Bootstrap Aggregating and variable importance via random subset selection (see Breiman 2001).

randomForest (RF) (see also Random Forests) 1

RF is an implementation of the Random Forests classification algorithm written in the R Statistical Language (see Liaw and Wiener 2002).

Receiver Operating Characteristics (ROC)..79

The ROC is a graphical representation of statistical classification model results. The ROC graph typically takes on a curved shape and is therefore often referred to as the ROC curve. The x-axis of the ROC graph is a model's False Positive Rate and the y-axis is the True Positive Rate; both are scaled from 0 to 1. The quantities on the x- and y-axes are also referred to as 1 – Specificity and Sensitivity, respectively. The actual curve in the graphic is generated by calculating the True Positive Rate and False Positive Rate for each cut-point of the model's prediction. The graphic also contains a line (often dashed) that originates at point 0,0 and goes at a 45-degree angle to point 1,1. This line represents a model that has no predictive power. The closer the ROC curve is to the upper left corner of the graph (which is point $x = 0$, $y = 1$), the greater the predictive power. Put another way, the best classification has the largest area under the curve. A line of this description will have a high True Positive Rate for the entire range of False Positive Rates. The ROC curve can be used to estimate the total predictive power of the model, often enumerated as the Area Under Curve, to compare similar models across all cut-points, or select an optimal cut-point to use for classification, resulting in a Confusion Matrix (see Fawcett 2004).

The RMSE is a statistic, or loss function, used to quantify the difference between an estimate and a true value. The RMSE is calculated as the square root of the Mean Squared Error. When calculated on Out-of-Sample predictions, such as in this project, the RMSE represents the sample standard deviation of the prediction errors. The formula below is how RMSE is calculated, where n = the number of data values, y_j is the observed j^{th} value and $\hat{y}_j$ is the predicted j^{th} value for all j values from 1 to n. Therefore the RMSE is the square root of the average of all squared errors.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}$$

A benefit of RMSE over Mean Squared Error is that it is scaled to the dependent variable and is therefore directly interpretable. With a binary dependent variable (0 to 1), the RMSE is taken as the distance on average between the predicted probability and the true value (see Salkind 2007).

Sensitivity is a term used for a classification's True Positive Rate; this value is also referred to as Recall. Sensitivity is the total fraction of sites that are classified by the model to be in the site-likely area. This measure is akin to the concept of precision and Type II errors. Sensitivity is calculated for a cut-point within a classification model as the number of correctly predicted positive observations (correctly classified sites) divided by the total number of actual positive observations (known sites) (see Oehlert and Shea 2007).

Specificity is a termed used for a classification's True Negative Rate. Specificity is the fraction of background that is classified as site-unlikely by the model. This measure is akin to the concept of accuracy and Type I errors. Specificity is calculated for a cut-point within a classification model as the number of correctly predicted negative observations (correctly classified non-sites) divided by the total number of actual negative observations (background cells) (see Oehlert and Shea 2007).

The TNR is a measure of a model's classification at a given cut-point. Often referred to as a model's Specificity, the TNR is calculated as the percent of negative observations correctly classified as such. In this project, this would be the rate at which background cells are correctly classified as site un-likely cells (see Oehlert and Shea 2007).

The TPR is a measure of a model's classification at a given cut-point. Often referred to as a models Sensitivity, the TPR is calculated as the percent of positive observations correctly classified as such. In this project, this would be the rate at which known site-present cells are correctly classified as site-likely cells (see Oehlert and Shea 2007).

The UDR is a measurement of a model's classification ability at a given cut-point. The UDR is defined as the probability of a cell containing a site given that the model predicted it as site-unlikely. That can be thought of as the rate of unintentional discovery, or "oops" rate (see Oehlert and Shea 2007).

APPENDIX B

SITE TYPES AND LANDFORMS RECORDED IN THE PASS DATABASE, BY TIME PERIOD

Region 7 Site Types by Landform, Paleoindian Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Isolated Find	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2
Open Habitation, Prehistoric	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3
Part of Multi-component Site	0	7	1	0	0	0	0	0	0	0	1	0	0	0	1	1	11
Total	0	10	1	0	0	1	0	0	0	0	1	1	0	0	1	1	16

Region 7 Site Types by Landform, Early Archaic Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Open Habitation, Prehistoric	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Rockshelter/cave	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Part of Multi-component Site	0	12	1	0	2	0	0	0	0	0	0	0	0	0	0	1	16
Total	0	12	1	0	3	0	0	1	0	0	0	0	0	0	0	1	18

Region 7 Site Types by Landform, Middle Archaic Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Open Habitation, Prehistoric	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Open Prehistoric Site, Unknown Function	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3
Rockshelter/cave	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Part of Multi-component Site	0	31	6	0	8	19	0	0	0	0	0	1	0	1	0	3	69
Total	0	34	7*	0	8	19*	0	0	0	1	1	1	0	1	0	3	75

*Note: One single component village likely miscatergorized in PASS data

Region 7 Site Types by Landform, Late Archaic Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Lithic Reduction	0	1	0	0	0	0	0	0	0	1	0	0	0		0	0	2
Open Habitation, Prehistoric	0	12	0	0	6	8	1	0	0	0	0	0	0	3	0	0	30
Open Prehistoric Site, Unknown Function	0	2	0	0	1	2	0	0	0	1	0	0	0	2	0	0	8
Rockshelter/cave	0	0	0	0	0	0	0	1	0	0	0	0	0		0	0	1
Unknown Function Open Site Greater than 20M Radius	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
(blank)	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
Part of Multi-component Site	2	95	10	0	15	32	3	7	1	2	2	2	0	6	0	7	184
Total	3	110	10	0	22	44	4	8	1	4	2	2	0	11	0	7	228

Region 7 Site Types by Landform, Terminal Archaic Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Lithic Reduction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2
Open Habitation, Prehistoric	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	16
Open Prehistoric Site, Unknown Function	0	4	0	0	1	1	0	1	0	0	0	0	0	0	0	0	7
Unknown Function Surface Scatter Less than 20M Radius	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
(blank)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
Part of Multi-component Site	1	89	4	0	10	23	4	5	2	2	1	0	0	4	0	2	147
Total	1	108	4	0	11	24	4	6	2	2	1	0	0	4	1	7	175

Region 7 Site Types by Landform, Early Woodland Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Lithic Reduction	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Rockshelter/cave	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Part of Multi-component Site	0	55	8	0	10	20	4	1	1	0	1	1	0	0	0	2	103
Total	0	55	8	0	10	21	4	1	1	0	2	1	0	0	0	2	105

Region 7 Site Types by Landform, Middle Woodland Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Open Habitation, Prehistoric	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Open Prehistoric Site, Unknown Function	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Rockshelter/cave	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	2
Part of Multi-component Site	0	47	4	0	3	10	0	2	0	0	1	1	0	1	0	1	70
Total	0	50	4	0	3	10	0	3	0	0	2	1	0	1	0	1	75

Region 7 Site Types by Landform, Late Woodland Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Cemetery	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	3
Lithic Reduction	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	2
Open Habitation, Prehistoric	0	39	2	0	2	16	1	1	0	1	0	1	0	0	0	0	63
Open Prehistoric Site, Unknown Function	0	4	0	0	1	3	0	2	0	1	0	0	0	0	0	2	13
Other Specialized Aboriginal Site	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Rockshelter/cave	0	0	0	0	3	0	0	8	0	2	1	0	0	0	0	0	14
Unknown Function Open Site Greater than 20M Radius	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Unknown Function Surface Scatter Less than 20M Radius	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Village	0	13	3	0	0	4	0	0	0	0	0	0	0	0	0	0	20
(blank)	0	3	0	0	3	0	0	0	0	0	0	0	0	0	0	5	11
Part of Multi-component Site	2	109	9	0	11	23	4	7	1	2	3	2	0	4	0	0	177
Total	2	171	15	0	20	49	5	18	1	6	4	3	0	5	0	8	307

*NOTE: Villages that have Archaic or Paleoindian material but no Early Woodland are counted as single component Late Woodland sites in this table.

Region 8 Site Types by Landform, Paleoindian Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Isolated Find	0	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0	4
Open Habitation, Prehistoric	0	3	0	0	0	1	0	1	0	0	0	0	0	1	0	0	6
Open Prehistoric Site, Unknown Function	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Part of Multi-component Site	0	5	0	0	5	7	0	0	0	0	1	1	0	1	0	0	20
Total	0	9	0	0	6	9	0	1	0	0	1	1	0	3	0	1	31

Region 8 Site Types by Landform, Early Archaic Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Open Habitation, Prehistoric	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Open Prehistoric Site, Unknown Function	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Quarry	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Rockshelter/cave	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Unknown Function Surface Scatter Less than 20M Radius	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
(blank)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Part of Multi-component Site	0	4	0	0	10	18	2	0	1	2	0	1	1	1	1	4	45
Total	0	5	0	0	10	18	2	0	1	2	2	2	1	2	1	5	51

Region 8 Site Types by Landform, Middle Archaic Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Isolated Find	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Lithic Reduction	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Open Habitation, Prehistoric	0	2	0	0	2	1	4	1	1	0	0	0	0	0	0	0	11
Open Prehistoric Site, Unknown Function	0	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	4
Unknown Function Open Site Greater than 20M Radius	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
(blank)	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	4	6
Part of Multi-component Site	0	26	1	1	29	41	7	1	5	7	1	3	0	9	2	5	138
Total	0	31	1	1	31	44	11	2	6	10	1	3	0	10	2	9	162

Region 8 Site Types by Landform, Late Archaic Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Lithic Reduction	0	1	0	0	0	2	0	1	0	2	0	0	0	0	0	1	7
Open Habitation, Prehistoric	0	25	0	0	72	45	13	9	4	0	0	0	3	12	0	1	184
Open Prehistoric Site, Unknown Function	0	3	1	0	6	8	1	2	0	0	0	4	0	5	1	3	34
Quarry	0	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0	4
Unknown Function Open Site Greater than 20M Radius	0	1	0	0	0	0	0		0	0	0	0	0	1	0	0	2
(blank)	0	9	1	0	4	7	0	1	0	0	0	1	0	0	0	3	26
Part of Multi-component Site	0	87	2	5	97	104	45	11	9	12	3	4	4	28	1	14	426
Total	0	127	4	5	179	166	59	25	13	14	3	10	7	47	2	22	683

Region 8 Site Types by Landform, Terminal Archaic Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Open Habitation, Prehistoric	0	3	0	0	10	0	9	3	2	1	0	0	0	4	0	0	32
Open Prehistoric Site, Unknown Function	0	2	2	0	0	7	1	0	0	0	0	0	0	0	0	1	13
(blank)	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	3
Part of Multi-component Site	1	62	1	4	63	73	35	9	7	10	4	2	2	17	0	8	298
Total	1	67	3	4	74	81	45	12	9	11	4	2	2	21	0	10	346

Region 8 Site Types by Landform, Early Woodland Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Open Habitation, Prehistoric	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	4
Open Prehistoric Site, Unknown Function	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
(blank)	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Part of Multi-component Site	0	27	0	2	23	27	8	5	0	5	0	0	1	13	0	6	117
Total	0	28	1	2	24	28	8	6	0	6	0	0	1	13	0	6	123

Region 8 Site Types by Landform, Middle Woodland Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Open Habitation, Prehistoric	0	2	0	0	0	3	0	0	0	0	1	0	0	1	0	0	7
Open Prehistoric Site, Unknown Function	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
(blank)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Part of Multi-component Site	0	21	0	2	17	26	6	6	2	3	0	0	0	6	0	5	94
Total	0	23	0	2	17	29	6	6	2	3	1	0	0	8	0	6	103

Region 8 Site Types by Landform, Late Woodland Period.

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Cemetery	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Lithic Reduction	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3	4
Open Habitation, Prehistoric	0	16	0	0	18	16	2	4	2	0	1	0	1	3	0	2	65
Open Prehistoric Site, Unknown Function	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3
Other Specialized Aboriginal Site	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	2
Quarry	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Rockshelter/cave	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
(blank)	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	2
Part of Multi-component Site	1	49	3	5	79	59	29	10	9	7	0	0	3	15	0	11	280
Total	1	67	3	5	100	76	31	15	11	8	1	0	4	20	0	17	359

Region 9 Site Types by Landform, Paleoindian Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Isolated Find	0	2	0	0	1		0	0	0	0	0	0	0	0	0	0	3
Open Habitation, Prehistoric	0	1	0	0	1	2	0	0	0	0	0	0	0	1	0	0	5
(blank)	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	3
Part of Multi-component Site	0	10	1	0	1	7	0	1	0	1	0	0	0	5	0	1	27
Total	0	13	1	0	3	11*	0	1	0	1	0	0	0	7	0	1	38

36LA007 Schultz noted as single component village (actually Susquehannock); 36LA0092 Reitz noted as SC Cemetery.

Region 9 Site Types by Landform, Early Archaic Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Lithic Reduction	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2
Open Habitation, Prehistoric	0	1	0	0	1	0	1	1	0	1	0	1	0	1	0	1	8
Open Prehistoric Site, Unknown Function	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	3
Unknown Function Open Site Greater than 20M Radius	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
(blank)	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Part of Multi-component Site	0	19	0	4	17	18	1	3	1	5	3	1	1	4	0	6	83
Total	0	22	1	4	18	19	2	4	1	8	4	2	1	5	0	7	98

Region 9 Site Types by Landform, Middle Archaic Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Isolated Find	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Lithic Reduction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Open Habitation, Prehistoric	0	5	0	0	1	4	1	0	2	0	1	0	0	0	0	0	14
Open Prehistoric Site, Unknown Function	0	0	0	0	2	2	0	0	0	1	1	0	0	0	0	1	7
Part of Multi-component Site	0	41	2	1	47	68	4	14	7	13	10	1	2	20	4	6	240
Total	0	46	2	1	50	74	5	14	9	14	13	1	2	20	4	8	263

Region 9 Site Types by Landform, Late Archaic Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Cemetery	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Isolated Find	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Lithic Reduction	0	2	0	0	0	1	0	0	0	2	1	0	0	3	0	7	16
Open Habitation, Prehistoric	0	28	2	1	66	50	8	2	1	2	1	0	3	10	1	7	182
Open Prehistoric Site, Unknown Function	0	8	0	0	17	16	1	15	2	5	2	1	2	6	0	5	80
Other Specialized Aboriginal Site	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	2
Quarry	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	3
Unknown Function Open Site Greater than 20M Radius	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	3
(blank)	0	3	1	1	4	4	2	3	2	0	1	1	0	2	0	2	26
Part of Multi-component Site	0	107	4	12	150	151	15	24	12	24	12	5	3	41	7	22	589
Total	0	148	8	14	239	222	28	44	19	33	18	7	8	62	10	43	903

Region 9 Site Types by Landform, Terminal Archaic Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Lithic Reduction	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Open Habitation, Prehistoric	0	12	0	0	12	18	2	1	0	0	0	0	1	1	0	0	47
Open Prehistoric Site, Unknown Function	0	0	0	0	1	0	0	0	0	2	0	0	0	1	0	0	4
Quarry	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2
(blank)	0	1	0	0	2	1	0	0	1	1	0	0	0	0	0	0	6
Part of Multi-component Site	0	94	1	12	90	91	10	15	6	8	8	2	2	22	5	7	373
Total	0	107	1	12	108	110	12	16	7	11	8	2	3	24	5	7	433

Region 9 Site Types by Landform, Early Woodland Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Lithic Reduction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Open Habitation, Prehistoric	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Open Prehistoric Site, Unknown Function	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	3
Rockshelter/cave	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	3
Part of Multi-component Site	0	41	1	8	31	44	2	10	5	6	6	1	1	9	2	10	177
Total	0	41	2	8	32	47	2	11	5	6	7	1	1	9	2	11	185

Region 9 Site Types by Landform, Middle Woodland Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Lithic Reduction	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
Open Habitation, Prehistoric	0	1	0	0	1	1	0	0	0	0	0	0	0	2	0	0	5
Open Prehistoric Site, Unknown Function	0	2	0	0	2	1	0	2	0	0	0	0	0	0	0	0	7
Other Specialized Aboriginal Site	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Rockshelter/cave	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	2
(blank)	1	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	4
Part of Multi-component Site	0	37	1	3	41	47	1	12	2	7	4	2		8	0	10	175
Total	1	40	1	3	45	52	2	15	2	8	4	2	0	10	0	10	195

Region 9 Site Types by Landform, Late Woodland Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Cemetery	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	3
Lithic Reduction	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	1	4
Open Habitation, Prehistoric	0	10	0	0	5	16	7	1	1	0	0	0	0	4	0	2	46
Open Prehistoric Site, Unknown Function	0	4	2	1	8	3	0	0	1	5	0	0	1	2	0	1	28
Other Specialized Aboriginal Site	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Rockshelter/cave	0	1	0	0	0	0	0	5	0	0	0	0	0	0	0	0	6
Unknown Function Surface Scatter Less than 20M Radius	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	2
Village	0	3	0	0	3	3	4	0	0	0	0	0	0	0	0	0	13
(blank)	0	2	0	1	2	3	0	2	0	0	0	0	0	1	0	1	12
Part of Multi-component Site	0	101	3	13	88	108	9	19	7	16	8	3	2	27	1	18	423
Total	0	121	5	15	108	137	20	27	9	22	9	3	3	34	2	23	538

Region 10 Site Types by Landform, Middle Archaic Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Part of Multi-component Site	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Total	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1

Region 10 Site Types by Landform, Late Archaic Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Other Specialized Aboriginal Site	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
(blank)	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2
Part of Multi-component Site	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	6
Total	0	1	0	0	0	8	0	0	0	0	0	0	0	0	0	0	9

Region 10 Site Types by Landform, Terminal Archaic Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Open Habitation, Prehistoric	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
Part of Multi-component Site	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
Total	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	4

Region 10 Site Types by Landform, Middle Woodland Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Part of Multi-component Site	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Total	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1

Region 10 Site Types by Landform, Late Woodland Period

Site Type	Beach	Flood Plain	Rise in Flood Plain	Island	Stream Bench	Terrace	Hill Ridge/Toe	Hillslope	Hilltop	Lower Slope	Middle Slope	Ridgetop	Saddle	Upland Flat	Upper Slope	(Blank)	Total
Open Prehistoric Site, Unknown Function	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Part of Multi-component Site	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	4
Total	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	5

APPENDIX C

VARIABLES CONSIDERED

WITHIN REGIONS 7, 8, AND 9/10

Predictor	Family	Measure	Neighborhood Sizes	Description
aspect	Topography	bearing	n/a	Orientation of slope relative to north
aws050	Soils - aggregate	water storage - integer	n/a	Water that is available to plants in the top 50 cm of soil. AWS is expressed as centimeters of water, reported as the average of all components in the map unit.
c_hyd_min	Hydrology	cost-distance	n/a	Minimum distance to stream or water body
c_hyd_min_wt	Hydrology	cost-distance	n/a	Minimum distance to stream, water body, or wetland
c_trail_dist	Topography - Cultural	cost-distance	n/a	Cost-distance to historically documented Native American trails (Wallace 1965).
cd_conf	Hydrology	cost-distance	n/a	Cost-Distance to stream confluence (NHD flow lines)
cd_drnh	Hydrology	cost-distance	n/a	Cost-Distance to stream heads (NHD flow lines)
cd_h1	Hydrology	cost-distance	n/a	Cost-distance to historic streams
cd_h2	Hydrology	cost-distance	n/a	Cost-distance to NHD flow lines
cd_h3	Hydrology	cost-distance	n/a	Cost-distance to NHD water bodies
cd_h4	Hydrology	cost-distance	n/a	Cost-distance to NWI wetlands
cd_h5	Hydrology	cost-distance	n/a	Cost-distance to NWI water bodies
cd_h6	Hydrology	cost-distance	n/a	Cost-distance to 4th order and higher streams
cd_h7	Hydrology	cost-distance	n/a	Cost-distance to 3rd order and higher streams
dem_fll	Topography	elevation, meters (float)	n/a	1/3rd Arc-second digital elevation model as float, with sinks filled
drcdry	Soils - aggregate	classification, nominal	n/a	Drainage class (dominant condition) - the NRCS describes natural soil drainage classes that represent the moisture condition of the soil in its natural condition throughout the year
drcwet	Soils - aggregate	classification, nominal	n/a	Drainage class (wet conditions) - the NRCS describes natural soil drainage classes that represent the moisture condition of the wettest soil component in its natural condition throughout the year

Predictor	Family	Measure	Neighborhood Sizes	Description
e_hyd_min	Hydrology	Euclidian-distance, meters	n/a	Minimum distance to stream or water body
e_hyd_min_wt	Hydrology	Euclidian-distance, meters	n/a	Minimum distance to stream, water body, or wetland
e_trail_dist	Topography - Cultural	Euclidian-distance, meters	n/a	Euclidian distance to historically documented Native American trails (Wallace 1965).
ed_conflu	Hydrology	Euclidian-distance, meters	n/a	Euclidian distance to stream confluence (NHD flow lines)
ed_drnh	Hydrology	Euclidian-distance, meters	n/a	Euclidian distance to stream heads (NHD flow lines)
ed_h1	Hydrology	Euclidian-distance, meters	n/a	Euclidian distance to historic streams
ed_h2	Hydrology	Euclidian-distance, meters	n/a	Euclidian distance to NHD flow lines
ed_h3	Hydrology	Euclidian-distance, meters	n/a	Euclidian distance to NHD water bodies
ed_h4	Hydrology	Euclidian-distance, meters	n/a	Euclidian distance to NWI wetlands
ed_h5	Hydrology	Euclidian-distance, meters	n/a	Euclidian distance to NWI water bodies
ed_h6	Hydrology	Euclidian-distance, meters	n/a	Euclidian distance to 4th order and higher streams
ed_h7	Hydrology	Euclidian-distance, meters	n/a	Euclidian distance to 3rd order and higher streams
eldrop#c	Topography	elevation, meters	1,8,10,16,32 cells	Drop in elevation over # cell neighborhood
elev_2_conf	Topography - Hydrology	vertical-distance, meters	na	Elevation to stream confluence (NHD flow lines)
elev_2_drainh	Topography - Hydrology	vertical-distance, meters	na	Elevation to stream head (NHD flow lines)
elev_2_strm	Topography - Hydrology	vertical-distance, meters	na	Elevation to stream (NHD flow lines)
flowdir	Hydrology	direction, bearing	na	Flow direction based on DEM
flw_acum	Hydrology	accumulation, cells	na	Flow accumulation based on DEM

Predictor	Family	Measure	Neighborhood Sizes	Description
niccdcd	Soils - aggregate	classification, nominal	n/a	The broadest category in the land capability classification system for soils; the dominant capability class, under nonirrigated conditions, for the map unit based on composition percentage of all components in the map unit.
random	Random	random float (0 to 1)	na	Randomly selected number between 1 and 0
rel_#c	Topography	index, 0 to 1	1,8,10,16,32 cells	Relative topographic position
rng_#c	Topography	elevation range, integer	1,8,10,16,32 cells	Range of elevation in # cell neighborhood
slope_deg	Topography	slope, degrees	n/a	Topographic slope measured in degrees
slope_pct	Topography	slope, percent	n/a	Topographic slope measured in percent rise over run
slpvr_#c	Topography	slope range, integer	1,8,10,16,32 cells	Slope variability within # cell neighborhood
std_#c	Topography	standard deviation	1,8,10,16,32 cells	Standard deviation of elevation range within # cell neighborhood
tpi_#c	Topography	index, integer	5,10,50,100,250 cells	Topographic Position Index. Position of cell relative to surrounding landscape within # cell neighborhood
tpi_cls#c	Topography	classification, nominal	5,10,50,100,250 cells	TPI standardized and classified into 1 standard deviation groups within # cell neighborhood
tpi_sd#c	Topography	standard deviation	5,10,50,100,250 cells	Standard deviation of TPI within # cell neighborhood
tri_#c	Topography	index, integer	1,8,10,16,32 cells	Topographic Ruggedness Index. Measure of terrain roughness within # cell neighborhood
twi#c	Topography - Hydrology	index, integer	1,8,10,16,32 cells	Topographic Wetness Index. Measure of upslope accumulation within # cell neighborhood
vrf_#c	Topography	index, integer	1,8,10,16,32 cells	Vector Roughness Factor. Measure of three-dimensional variation in slope within # cell neighborhood

APPENDIX D

VARIABLES SELECTED

FOR EACH OF 66 MODELS

WITHIN REGIONS 7, 8, AND 9/10

Region 7 All - Riverine Section 1				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_trail_dist	0.329718771	p < 0.001	1111945839	p < 0.001
cd_drnh	0.260696162	p < 0.001	1508953884	p < 0.001
cd_h4	0.229346142	p < 0.001	1433183465	p < 0.001
cd_h5	0.236488341	p < 0.001	1480679089	p < 0.001
e_hyd_min	0.401009151	p < 0.001	2873297548	p < 0.001
ed_h2	0.394006607	p < 0.001	2847384725	p < 0.001
ed_h6	0.500259607	p < 0.001	1207775303	p < 0.001
elev_2_drainh	0.257844684	p < 0.001	2312143653	p < 0.001
elev_2_strm	0.35286079	p < 0.001	1378858428	p < 0.001
niccdcd	0.261968889	p < 0.001	1456900238	p < 0.001
rng_32c	0.353563788	p < 0.001	1163026987	p < 0.001
slpvr_10c	0.245065423	p < 0.001	1354810553	p < 0.001
std_32c	0.349734845	p < 0.001	1177997398	p < 0.001
tpi_10c	0.420885732	p < 0.001	2913323374	p < 0.001
tpi_cls10c	0.372457382	p < 0.001	2710319237	p < 0.001
tpi_sd10c	0.420809699	p < 0.001	2913102133	p < 0.001
tri_10c	0.251494522	p < 0.001	1343293470	p < 0.001
vrf_32c	0.252811096	p < 0.001	1326991199	p < 0.001
random	0.008006543	p = 0.103	1940520582	p = 0.107

Region 7 All - Riverine Section 2				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
cd_conf	0.374218546	p < 0.001	408448629.9	p < 0.001
cd_drnh	0.305747891	p < 0.001	190783409.2	p < 0.001
cd_h4	0.386039612	p < 0.001	423188570.6	p < 0.001
e_trail_dist	0.567997915	p < 0.001	456858663.5	p < 0.001
ed_h2	0.371014891	p < 0.001	434860668.7	p < 0.001
ed_h5	0.513740195	p < 0.001	138202505.8	p < 0.001
ed_h7	0.475734951	p < 0.001	410907350.8	p < 0.001
elev_2_conf	0.382115395	p < 0.001	438271895.3	p < 0.001
elev_2_drainh	0.359014625	p < 0.001	389290803.2	p < 0.001
elev_2_strm	0.418572616	p < 0.001	384002557.8	p < 0.001
niccdcd	0.246491507	p < 0.001	213619275.9	p < 0.001
rng_32c	0.324478349	p < 0.001	218457511.1	p < 0.001
tpi_250c	0.434478716	p < 0.001	386358646.7	p < 0.001
tpi_cls250c	0.377881416	p < 0.001	387529981	p < 0.001
tpi_sd250c	0.434321756	p < 0.001	386250654.9	p < 0.001
vrf_32c	0.370200993	p < 0.001	170310715.1	p < 0.001
random	0.007833265	p = 0.594	305454672.2	p = 0.725

Region 7 All - Riverine Section 3				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.271389806	p < 0.001	1586528241	p < 0.001
c_trail_dist	0.53741048	p < 0.001	694471224.6	p < 0.001
cd_drnh	0.331322234	p < 0.001	975272262.6	p < 0.001
drcwet	0.474279374	p < 0.001	2160391407	p < 0.001
e_hyd_min	0.477013618	p < 0.001	2155585662	p < 0.001
ed_h2	0.533237013	p < 0.001	2201256420	p < 0.001
ed_h5	0.39688242	p < 0.001	1011500992	p < 0.001
ed_h6	0.605871913	p < 0.001	661899610.9	p < 0.001
eldrop32c	0.272743015	p < 0.001	1745841822	p < 0.001
elev_2_conf	0.272938135	p < 0.001	1652271383	p < 0.001
elev_2_drainh	0.28950202	p < 0.001	1631332835	p < 0.001
elev_2_strm	0.370145441	p < 0.001	1057324258	p < 0.001
niccdcd	0.291636845	p < 0.001	1011974431	p < 0.001
rel_10c	0.358504733	p < 0.001	1989738812	p < 0.001
slpvr_32c	0.304066112	p < 0.001	1891512815	p < 0.001
tpi_250c	0.487543258	p < 0.001	598329021.9	p < 0.001
tpi_cls250c	0.461885166	p < 0.001	656638768.3	p < 0.001
tpi_sd250c	0.487738216	p < 0.001	598066043.9	p < 0.001
tri_32c	0.300483387	p < 0.001	1886322670	p < 0.001
random	0.00531362	p = 0.502	1349260947	p = 0.620

Region 7 All - Riverine Section 4				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_trail_dist	0.418237841	p < 0.001	263629187	p < 0.001
cd_drnh	0.405869173	p < 0.001	216573464.4	p < 0.001
e_hyd_min_wt	0.337810674	p < 0.001	452133567	p < 0.001
ed_h2	0.539278684	p < 0.001	544604729.2	p < 0.001
ed_h5	0.275480891	p < 0.001	250679220.7	p < 0.001
ed_h7	0.365214275	p < 0.001	289877070.6	p < 0.001
elev_2_strm	0.235602292	p < 0.001	313477324.9	p < 0.001
niccdcd	0.247453469	p < 0.001	246931300.4	p < 0.001
rel_10c	0.242885624	p < 0.001	404168316.7	p < 0.001
rng_32c	0.313989183	p < 0.001	239046728.9	p < 0.001
slpvr_16c	0.275059498	p < 0.001	239505026.6	p < 0.001
std_32c	0.263301082	p < 0.001	236806561.4	p < 0.001
tpi_10c	0.35058641	p < 0.001	491239484.7	p < 0.001
tpi_cls10c	0.351010844	p < 0.001	465471272	p < 0.001
tpi_sd10c	0.351259679	p < 0.001	491488166.5	p < 0.001
tri_16c	0.279579009	p < 0.001	237855421.3	p < 0.001
random	0.007042856	p = 0.668	335995767.1	p = 0.720

Region 7 All - Riverine Section 5				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.198717521	p < 0.001	518990444.3	p < 0.001
c_trail_dist	0.242184858	p < 0.001	646686603.3	p = 0.135
cd_conf	0.349000712	p < 0.001	838760919.4	p < 0.001
cd_drnh	0.317727629	p < 0.001	433474704.9	p < 0.001
cd_h4	0.380544195	p < 0.001	897732278	p < 0.001
e_hyd_min_wt	0.323403353	p < 0.001	882049142.6	p < 0.001
ed_h2	0.373755307	p < 0.001	945626493.8	p < 0.001
ed_h5	0.257767533	p < 0.001	500651933.2	p < 0.001
ed_h6	0.251876987	p < 0.001	529120833.5	p < 0.001
elev_2_conf	0.269344022	p < 0.001	816568760.8	p < 0.001
elev_2_drainh	0.356737676	p < 0.001	836388398.3	p < 0.001
niccdcd	0.322534463	p < 0.001	428931033.5	p < 0.001
rel_10c	0.230930622	p < 0.001	794297785.8	p < 0.001
rng_32c	0.289472863	p < 0.001	444534620	p < 0.001
slpvr_8c	0.264197185	p < 0.001	451455903.2	p < 0.001
std_32c	0.213961814	p < 0.001	471178330.4	p < 0.001
tpi_10c	0.32561108	p < 0.001	932925030.3	p < 0.001
tpi_cls10c	0.325425911	p < 0.001	884823640	p < 0.001
tpi_sd10c	0.325685901	p < 0.001	932884867.1	p < 0.001
tri_8c	0.264201011	p < 0.001	455812113	p < 0.001
random	0.00772836	p = 0.328	643677076.9	p = 0.469

Region 7 All - Riverine Section 6				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_hyd_min	0.346814058	p < 0.001	790286744.8	p < 0.001
c_trail_dist	0.67460311	p < 0.001	240880742.8	p < 0.001
cd_h6	0.553849283	p < 0.001	345370959.4	p < 0.001
drcwet	0.409184943	p < 0.001	803542911.7	p < 0.001
ed_conf	0.488602604	p < 0.001	969891100	p < 0.001
ed_drnh	0.344040093	p < 0.001	879609253.3	p < 0.001
ed_h2	0.427040821	p < 0.001	899967168.7	p < 0.001
ed_h4	0.275482624	p < 0.001	795984418.7	p < 0.001
ed_h5	0.324577707	p < 0.001	453018002.8	p < 0.001
eldrop32c	0.285370228	p < 0.001	736806314.9	p < 0.001
elev_2_strm	0.352473552	p < 0.001	493865423.4	p < 0.001
niccdcd	0.283484305	p < 0.001	450368457.6	p < 0.001
rng_32c	0.387454109	p < 0.001	896020345.1	p < 0.001
slpvr_32c	0.409983533	p < 0.001	924704104	p < 0.001
std_32c	0.368392018	p < 0.001	886198267.2	p < 0.001
tpi_250c	0.551914567	p < 0.001	214578364.5	p < 0.001
tpi_cls250c	0.488827716	p < 0.001	270953796.9	p < 0.001
tpi_sd250c	0.551914887	p < 0.001	214674729.5	p < 0.001
tri_32c	0.409924431	p < 0.001	924281739.1	p < 0.001
random	0.008718759	p = 0.230	592275167.5	p = 0.345

Region 7 All - Riverine Section 7				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
cd_h2	0.430433699	p < 0.001	676423181.9	p < 0.001
cd_h4	0.398333511	p < 0.001	618366311.1	p < 0.001
drcwet	0.425497381	p < 0.001	654456824.7	p < 0.001
e_hyd_min_wt	0.503510122	p < 0.001	714802797.5	p < 0.001
e_trail_dist	0.54386926	p < 0.001	170491532.2	p < 0.001
ed_drnh	0.398199423	p < 0.001	646101686.7	p < 0.001
ed_h5	0.351026373	p < 0.001	360434163.2	p < 0.001
ed_h6	0.656228722	p < 0.001	175174943.2	p < 0.001
eldrop32c	0.386709277	p < 0.001	656524896.9	p < 0.001
elev_2_drainh	0.433832268	p < 0.001	227028858.9	p < 0.001
niccdcd	0.355961839	p < 0.001	343209548.3	p < 0.001
rel_16c	0.443971995	p < 0.001	696608616.2	p < 0.001
slpvr_32c	0.323001781	p < 0.001	585504309.1	p < 0.001
tpi_250c	0.519074291	p < 0.001	174287414.6	p < 0.001
tpi_cls250c	0.477724238	p < 0.001	213080994.2	p < 0.001
tpi_sd250c	0.518906423	p < 0.001	174444588	p < 0.001
tri_32c	0.321627104	p < 0.001	584885719.1	p < 0.001
twi32c	0.309143458	p < 0.001	260579953.3	p < 0.001
random	0.009813073	p = 0.212	442573107.8	p = 0.630

Region 7 All - Riverine Section 8				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aspect	0.428061769	p < 0.001	68917563.02	p < 0.001
cd_drnh	0.555771597	p < 0.001	216840928.9	p < 0.001
drcwet	0.430495775	p < 0.001	187833096.4	p < 0.001
e_trail_dist	0.591332889	p < 0.001	44537707.04	p < 0.001
ed_h1	0.706442699	p < 0.001	39150186.21	p < 0.001
ed_h4	0.576710896	p < 0.001	234357009.2	p < 0.001
ed_h6	0.797221918	p < 0.001	23434556.9	p < 0.001
eldrop32c	0.452258927	p < 0.001	108228620.2	p < 0.001
elev_2_drainh	0.406853626	p < 0.001	77259529.53	p < 0.001
niccdcd	0.393543744	p < 0.001	95283716.93	p < 0.001
rel_8c	0.485658458	p < 0.001	87823588.28	p < 0.001
rng_32c	0.526975928	p < 0.001	203792395.8	p < 0.001
slope_pct	0.428346842	p < 0.001	112141707.1	p < 0.001
slpvr_32c	0.617031642	p < 0.001	217751989.7	p < 0.001
std_32c	0.527960362	p < 0.001	207160487.6	p < 0.001
tpi_250c	0.660553032	p < 0.001	39654808.97	p < 0.001
tpi_cls250c	0.646202181	p < 0.001	52700187.36	p < 0.001
tpi_sd250c	0.660516068	p < 0.001	39680576.49	p < 0.001
tri_32c	0.617280263	p < 0.001	217689593.7	p < 0.001
twi32c	0.441301522	p < 0.001	185675277.7	p < 0.001
random	0.015182246	p = 0.208	143271643.8	p = 0.194

Region 7 All - Riverine Section 9				
Predictor	Mean D	Mean KS p	Mean U	Mean MW p
aws050	0.312482541	p < 0.001	217121278.7	p < 0.001
c_trail_dist	0.425797852	p < 0.001	103186505.2	p < 0.001
cd_conf	0.350537012	p < 0.001	120829863.9	p < 0.001
cd_drnh	0.291249928	p < 0.001	141669546.7	p < 0.001
cd_h1	0.298138668	p < 0.001	132237381.6	p < 0.001
cd_h4	0.316736722	p < 0.001	132906926.7	p < 0.001
drcwet	0.333579722	p < 0.001	210183746.2	p < 0.001
ed_h5	0.323017784	p < 0.001	137890194.3	p < 0.001
ed_h6	0.359284724	p < 0.001	114524460	p < 0.001
elev_2_drainh	0.388618055	p < 0.001	117894178.9	p < 0.001
niccdcd	0.289001097	p < 0.001	193376298.3	p < 0.001
rel_32c	0.398120936	p < 0.001	243061154.2	p < 0.001
rng_32c	0.418820027	p < 0.001	100679362.5	p < 0.001
slpvr_16c	0.391080957	p < 0.001	104542695	p < 0.001
std_32c	0.392392877	p < 0.001	97373932.44	p < 0.001
tpi_10c	0.417085853	p < 0.001	234186209.6	p < 0.001
tpi_sd10c	0.416904061	p < 0.001	234150571.7	p < 0.001
tri_16c	0.392045836	p < 0.001	104674051.4	p < 0.001
vrf_32c	0.39321656	p < 0.001	93170829.38	p < 0.001
random	0.010150812	p = 0.592	164177832.7	p = 0.769

Region 7 All - Upland Section 1				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_hyd_min	0.637755292	p < 0.001	356430250.3	p < 0.001
c_trail_dist	0.594358129	p < 0.001	361818749.3	p < 0.001
cd_conf	0.797843472	p < 0.001	128412998.7	p < 0.001
cd_h2	0.664504211	p < 0.001	330419953.6	p < 0.001
cd_h4	0.703218179	p < 0.001	267122028.5	p < 0.001
cd_h5	0.660005959	p < 0.001	299354667.6	p < 0.001
cd_h7	0.895681608	p < 0.001	52253918.44	p < 0.001
eldrop32c	0.635820948	p < 0.001	362741885.2	p < 0.001
elev_2_conf	0.744331795	p < 0.001	259183691.7	p < 0.001
elev_2_strm	0.878280429	p < 0.001	75729135.23	p < 0.001
niccdcd	0.525911959	p < 0.001	748635506.4	p < 0.001
rel_32c	0.559036371	p < 0.001	447288476.4	p < 0.001
rng_16c	0.606366502	p < 0.001	398257141.2	p < 0.001
slope_deg	0.529143412	p < 0.001	578435603.1	p < 0.001
std_16c	0.569539927	p < 0.001	433776031	p < 0.001
tpi_250c	0.830649098	p < 0.001	111499890.8	p < 0.001
tpi_cls250c	0.78877386	p < 0.001	165969082.7	p < 0.001
tpi_sd250c	0.830630121	p < 0.001	111582344.1	p < 0.001
random	0.009725551	p = 0.043	1532889443	p = 0.138

Region 7 All - Upland Section 2				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.405737351	p < 0.001	267567979.3	p < 0.001
c_hyd_min	0.636415657	p < 0.001	71347866.62	p < 0.001
cd_conf	0.607813948	p < 0.001	126687411	p < 0.001
cd_h2	0.388275476	p < 0.001	157231929.7	p < 0.001
cd_h4	0.415822213	p < 0.001	165141158.4	p < 0.001
cd_h5	0.77847204	p < 0.001	32304838.58	p < 0.001
cd_h6	0.427690211	p < 0.001	251173350.7	p < 0.001
e_trail_dist	0.695715163	p < 0.001	536507543.8	p < 0.001
eldrop32c	0.572599856	p < 0.001	87791530.93	p < 0.001
elev_2_conf	0.574468607	p < 0.001	146663546	p < 0.001
elev_2_drainh	0.405520118	p < 0.001	162894284	p < 0.001
elev_2_strm	0.559431205	p < 0.001	129292966	p < 0.001
rel_16c	0.5669609	p < 0.001	106408012.3	p < 0.001
rng_16c	0.482179985	p < 0.001	117286190.7	p < 0.001
slope_deg	0.473629319	p < 0.001	137834790.9	p < 0.001
std_8c	0.468391832	p < 0.001	131337444.7	p < 0.001
tpi_100c	0.709383422	p < 0.001	93987205.4	p < 0.001
tpi_cls100c	0.684888287	p < 0.001	91438011.65	p < 0.001
tpi_sd100c	0.709445724	p < 0.001	93895146.01	p < 0.001
vrf_32c	0.527097001	p < 0.001	110114719.9	p < 0.001
random	0.007670972	p = 0.575	334368147	p = 0.510

Region 7 All - Upland Section 3				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.441308491	p < 0.001	997460892.7	p < 0.001
c_hyd_min	0.526470652	p < 0.001	353047041.9	p < 0.001
c_trail_dist	0.693427964	p < 0.001	244353055	p < 0.001
cd_conf	0.642359796	p < 0.001	240254388.9	p < 0.001
cd_drnh	0.427248345	p < 0.001	397549620.6	p < 0.001
cd_h2	0.532701915	p < 0.001	340414012.7	p < 0.001
cd_h4	0.51311661	p < 0.001	353082824.7	p < 0.001
cd_h5	0.649792508	p < 0.001	224558652.3	p < 0.001
cd_h6	0.775742092	p < 0.001	125431703.5	p < 0.001
eldrop32c	0.51651077	p < 0.001	334154206.8	p < 0.001
elev_2_conf	0.597593506	p < 0.001	295342716.9	p < 0.001
elev_2_strm	0.748063668	p < 0.001	177096348.7	p < 0.001
niccdcd	0.526002931	p < 0.001	301780934.5	p < 0.001
rel_32c	0.626836588	p < 0.001	230997782.4	p < 0.001
rng_8c	0.492050819	p < 0.001	315251410.5	p < 0.001
slope_pct	0.394552134	p < 0.001	403936978.3	p < 0.001
slpvr_32c	0.405104691	p < 0.001	1237025783	p < 0.001
std_10c	0.490169566	p < 0.001	322177394.6	p < 0.001
tpi_250c	0.769393346	p < 0.001	184720672	p < 0.001
tpi_cls250c	0.764547498	p < 0.001	181447005.3	p < 0.001
tpi_sd250c	0.769355005	p < 0.001	184623121.7	p < 0.001
tri_32c	0.397473805	p < 0.001	1228225722	p < 0.001
random	0.006878025	p = 0.374	833134949.6	p = 0.425

Region 7 All - Upland Section 4				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.658459146	p < 0.001	103664417.5	p < 0.001
c_hyd_min_wt	0.579812147	p < 0.001	64596301.61	p < 0.001
c_trail_dist	0.604401145	p < 0.001	149016160	p < 0.001
cd_conf	0.788413187	p < 0.001	37515097.62	p < 0.001
cd_drnh	0.565744734	p < 0.001	99555415.14	p < 0.001
cd_h2	0.456532888	p < 0.001	125852593	p < 0.001
cd_h4	0.744508942	p < 0.001	34450650.2	p < 0.001
cd_h5	0.716044442	p < 0.001	33977322.84	p < 0.001
cd_h7	0.671453177	p < 0.001	59478240.89	p < 0.001
eldrop16c	0.590823568	p < 0.001	60836425.37	p < 0.001
elev_2_conf	0.683263575	p < 0.001	67830404.23	p < 0.001
elev_2_drainh	0.509863887	p < 0.001	89816804.15	p < 0.001
elev_2_strm	0.740562438	p < 0.001	47558590.6	p < 0.001
rel_8c	0.509376315	p < 0.001	120234308.1	p < 0.001
rng_16c	0.70453829	p < 0.001	42712253.02	p < 0.001
slope_pct	0.554183902	p < 0.001	73285139.22	p < 0.001
std_16c	0.667610045	p < 0.001	46932312.59	p < 0.001
tpi_100c	0.577686636	p < 0.001	95707736.8	p < 0.001
tpi_sd100c	0.578247186	p < 0.001	95606531.07	p < 0.001
vrf_32c	0.461081158	p < 0.001	122741042.6	p < 0.001
random	0.013655754	p = 0.131	246658777.7	p = 0.358

Region 7 All - Upland Section 5				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_hyd_min	0.484245676	p < 0.001	228435776.5	p < 0.001
c_trail_dist	0.351458351	p < 0.001	509461474.3	p < 0.001
cd_conf	0.361082001	p < 0.001	384928878.7	p < 0.001
cd_drnh	0.559186748	p < 0.001	208257679.5	p < 0.001
cd_h2	0.53276699	p < 0.001	196343715.5	p < 0.001
cd_h4	0.468729221	p < 0.001	352727840	p < 0.001
cd_h5	0.541176505	p < 0.001	193470884.1	p < 0.001
cd_h6	0.298413886	p < 0.001	444885010.9	p < 0.001
drcdry	0.299176531	p < 0.001	408678580.2	p < 0.001
eldrop32c	0.51663474	p < 0.001	217648518.2	p < 0.001
elev_2_conf	0.399394102	p < 0.001	352065138.6	p < 0.001
elev_2_drainh	0.312708817	p < 0.001	515670817	p < 0.001
elev_2_strm	0.298281019	p < 0.001	383804290.2	p < 0.001
niccdcd	0.39270764	p < 0.001	308131880.5	p < 0.001
rel_32c	0.405973718	p < 0.001	341167939.1	p < 0.001
rng_32c	0.51585407	p < 0.001	236301388.2	p < 0.001
slope_deg	0.319667241	p < 0.001	339352242.8	p < 0.001
std_16c	0.456192988	p < 0.001	263853777.8	p < 0.001
tpi_50c	0.333481931	p < 0.001	345433896.7	p < 0.001
tpi_cls50c	0.289977157	p < 0.001	365596035.2	p < 0.001
tpi_sd50c	0.333246375	p < 0.001	345737288.4	p < 0.001
tri_10c	0.321911749	p < 0.001	379669549.2	p < 0.001
random	0.009149886	p = 0.182	637156084.6	p = 0.299

Region 7 All - Upland Section 6				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_hyd_min	0.538307413	p < 0.001	231359990.1	p < 0.001
c_trail_dist	0.881534581	p < 0.001	54664203.3	p < 0.001
cd_conf	0.748720477	p < 0.001	108254226.4	p < 0.001
cd_h2	0.571864755	p < 0.001	222335649.8	p < 0.001
cd_h4	0.435854102	p < 0.001	198897618.3	p < 0.001
cd_h6	0.905155617	p < 0.001	42637286.98	p < 0.001
ed_drnh	0.603093082	p < 0.001	788799013.2	p < 0.001
ed_h5	0.656651984	p < 0.001	108254079.9	p < 0.001
eldrop32c	0.539762807	p < 0.001	211428800.9	p < 0.001
elev_2_conf	0.764577172	p < 0.001	119205728	p < 0.001
elev_2_drainh	0.499983447	p < 0.001	237883400.9	p < 0.001
elev_2_strm	0.888031836	p < 0.001	65848967.75	p < 0.001
niccdcd	0.55823611	p < 0.001	149115978.5	p < 0.001
rel_32c	0.786599928	p < 0.001	73749771.28	p < 0.001
rng_32c	0.434916414	p < 0.001	685023591.8	p < 0.001
slpvr_32c	0.636127076	p < 0.001	789573576.4	p < 0.001
tpi_250c	0.820125063	p < 0.001	45106113.07	p < 0.001
tpi_cls250c	0.779330145	p < 0.001	51242973.95	p < 0.001
tpi_sd250c	0.820227984	p < 0.001	45089654.47	p < 0.001
tri_32c	0.634458719	p < 0.001	788919494.1	p < 0.001
random	0.013594286	p = 0.036	442552866.6	p = 0.195

Region 7 All - Upland Section 7				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_trail_dist	0.432703535	p < 0.001	147816284.4	p < 0.001
cd_drnh	0.444602389	p < 0.001	284353522.3	p < 0.001
cd_h7	0.783431306	p < 0.001	38900479.97	p < 0.001
e_hyd_min	0.327379994	p < 0.001	122240800.8	p < 0.001
ed_conf	0.444008322	p < 0.001	100586702	p < 0.001
ed_h2	0.447055713	p < 0.001	97574635.27	p < 0.001
ed_h5	0.590346563	p < 0.001	66381121.79	p < 0.001
elev_2_conf	0.393350722	p < 0.001	108959672.9	p < 0.001
elev_2_drainh	0.657016064	p < 0.001	47570487.8	p < 0.001
elev_2_strm	0.728573478	p < 0.001	39934923.87	p < 0.001
niccdcd	0.38179497	p < 0.001	150690925.2	p < 0.001
rel_32c	0.464395814	p < 0.001	84108379.8	p < 0.001
slpvr_32c	0.557885805	p < 0.001	306654523.6	p < 0.001
tpi_250c	0.702319359	p < 0.001	42783549.65	p < 0.001
tpi_cls100c	0.666919886	p < 0.001	56993788.49	p < 0.001
tpi_sd250c	0.701959431	p < 0.001	42825390.94	p < 0.001
tri_32c	0.556024361	p < 0.001	306327539.6	p < 0.001
random	0.01147536	p = 0.348	203499239.1	p = 0.427

Region 7 All - Upland Section 8				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
cd_drnh	0.366066149	p < 0.001	83423096.99	p < 0.001
e_hyd_min	0.414440095	p < 0.001	34558506.29	p < 0.001
e_trail_dist	0.499089136	p < 0.001	22268056.35	p < 0.001
ed_conf	0.567230305	p < 0.001	24835383.21	p < 0.001
ed_h2	0.530790864	p < 0.001	25152822.44	p < 0.001
ed_h6	0.701981223	p < 0.001	17386440.6	p < 0.001
elev_2_strm	0.461445771	p < 0.001	25979116.37	p < 0.001
rng_16c	0.488344883	p < 0.001	90119855.42	p < 0.001
slpvr_32c	0.632200271	p < 0.001	99944417.8	p < 0.001
std_16c	0.47614774	p < 0.001	91492233.3	p < 0.001
tpi_250c	0.670967757	p < 0.001	19212196.21	p < 0.001
tpi_cls250c	0.659458405	p < 0.001	21841904.39	p < 0.001
tpi_sd250c	0.67096666	p < 0.001	19210241.46	p < 0.001
tri_32c	0.629924871	p < 0.001	99737000.19	p < 0.001
random	0.029982974	p = 0.0354	60080002.15	p = 0.230

Region 7 All - Upland Section 9				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_trail_dist	0.560207103	p < 0.001	16675431.1	p < 0.001
cd_conf	0.33271451	p < 0.001	19704496.66	p < 0.001
cd_h7	0.510203547	p < 0.001	16654096.45	p < 0.001
e_hyd_min	0.453058462	p < 0.001	17006831.89	p < 0.001
ed_drnh	0.287319	p < 0.001	24278360.83	p < 0.001
ed_h2	0.455497898	p < 0.001	15652186.74	p < 0.001
elev_2_conf	0.474301231	p < 0.001	16592332.52	p < 0.001
elev_2_strm	0.456791282	p < 0.001	18183585.28	p < 0.001
rel_32c	0.268711814	p < 0.001	23120692.66	p < 0.001
slpvr_32c	0.313208568	p < 0.001	39327833.68	p < 0.001
tpi_250c	0.425265525	p < 0.001	20663943.3	p < 0.001
tpi_cls250c	0.380766818	p < 0.001	22477696.44	p < 0.001
tpi_sd250c	0.425113899	p < 0.001	20659159.86	p < 0.001
tri_32c	0.314111186	p < 0.001	39097520.5	p < 0.001
random	0.020056961	p = 0.650	34994985.44	p = 0.473

Region 8 All - Riverine Section 1				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
drcwet	0.380848	p < 0.001	1785696493	p < 0.001
e_hyd_min_wt	0.233841	p < 0.001	1669095392	p < 0.001
e_trail_dist	0.404208	p < 0.001	908084965	p < 0.001
ed_conf	0.182318	p < 0.001	1607663678	p < 0.001
ed_h2	0.418585	p < 0.001	2015698019	p < 0.001
ed_h4	0.192255	p < 0.001	1576019889	p < 0.001
ed_h5	0.332759	p < 0.001	865371127	p < 0.001
ed_h6	0.409613	p < 0.001	752095862	p < 0.001
eldrop32c	0.199784	p < 0.001	1457921780	p < 0.001
niccdcd	0.177699	p < 0.001	1117548712	p < 0.001
rel_10c	0.246362	p < 0.001	1634529753	p < 0.001
rng_16c	0.202075	p < 0.001	1112125391	p < 0.001
slpvr_16c	0.185413	p < 0.001	1169045858	p < 0.001
tpi_100c	0.30689	p < 0.001	898588188	p < 0.001
tpi_cls10c	0.215925	p < 0.001	1673945125	p < 0.001
tpi_sd100c	0.307078	p < 0.001	898400257	p < 0.001
tri_16c	0.189365	p < 0.001	1166677267	p < 0.001
vrf_32c	0.215534	p < 0.001	963817158	p < 0.001
random	0.006558	p = 0.320	1321306612	p = 0.314

Region 8 All - Riverine Section 2				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.391358	p < 0.001	700325182	p < 0.001
cd_h4	0.307948	p < 0.001	1615681604	p < 0.001
cd_h7	0.584913	p < 0.001	628859211	p < 0.001
drcwet	0.351279	p < 0.001	1733725929	p < 0.001
e_hyd_min	0.628943	p < 0.001	2176201834	p < 0.001
e_trail_dist	0.410092	p < 0.001	779336173	p < 0.001
ed_drnh	0.525911	p < 0.001	1929167328	p < 0.001
ed_h2	0.639942	p < 0.001	2164962374	p < 0.001
elev_2_conf	0.364287	p < 0.001	1687192639	p < 0.001
elev_2_drainh	0.359815	p < 0.001	1776770924	p < 0.001
elev_2_strm	0.309057	p < 0.001	1036515945	p < 0.001
niccdcd	0.294528	p < 0.001	828939780	p < 0.001
rel_32c	0.501593	p < 0.001	2060097072	p < 0.001
rng_32c	0.364819	p < 0.001	731501730	p < 0.001
slpvr_10c	0.3191	p < 0.001	857712577	p < 0.001
std_32c	0.325219	p < 0.001	764670147	p < 0.001
tpi_10c	0.512723	p < 0.001	2090160111	p < 0.001
tpi_cls10c	0.48883	p < 0.001	1976650691	p < 0.001
tpi_sd10c	0.512658	p < 0.001	2089667207	p < 0.001
tri_16c	0.320183	p < 0.001	879172951	p < 0.001
vrf_32c	0.323601	p < 0.001	816889347	p < 0.001
random	0.008571	p = 0.087	1270386476	p = 0.178

Region 8 All - Riverine Section 3				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.363621	p < 0.001	182234796	p < 0.001
c_trail_dist	0.633129	p < 0.001	53119669	p < 0.001
cd_drnh	0.492758	p < 0.001	210642732	p < 0.001
cd_h4	0.244583	p < 0.001	131977545	p < 0.001
ed_conf	0.358089	p < 0.001	98089348	p < 0.001
ed_h1	0.882373	p < 0.001	16236983	p < 0.001
ed_h5	0.285638	p < 0.001	175512403	p < 0.001
ed_h6	0.802533	p < 0.001	27846833	p < 0.001
elev_2_drainh	0.463539	p < 0.001	84688358	p < 0.001
elev_2_strm	0.540317	p < 0.001	78126756	p < 0.001
flowdir	0.261436	p < 0.001	184440875	p < 0.001
rng_32c	0.281612	p < 0.001	168996642	p < 0.001
slpvr_32c	0.418306	p < 0.001	199618800	p < 0.001
std_32c	0.272901	p < 0.001	170860056	p < 0.001
tpi_250c	0.403054	p < 0.001	68550147	p < 0.001
tpi_cls250c	0.256947	p < 0.001	91669870	p < 0.001
tpi_sd250c	0.402966	p < 0.001	68560206	p < 0.001
tri_32c	0.406051	p < 0.001	197962011	p < 0.001
random	0.016183	p = 0.141	148257148	p = 0.740

Region 8 All - Riverine Section 4				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.157663	p < 0.001	1100622612	p < 0.001
cd_conf	0.154407	p < 0.001	1143863274	p < 0.001
drcwet	0.212642	p < 0.001	1196676625	p < 0.001
e_hyd_min	0.303798	p < 0.001	1226924656	p < 0.001
e_trail_dist	0.146203	p < 0.001	823754707	p < 0.001
ed_h2	0.226325	p < 0.001	1110107043	p < 0.001
ed_h4	0.210176	p < 0.001	780304480	p < 0.001
ed_h5	0.155542	p < 0.001	937962778	p < 0.001
ed_h6	0.161372	p < 0.001	1040364886	p < 0.001
eldrop16c	0.20355	p < 0.001	1240789826	p < 0.001
elev_2_conf	0.15653	p < 0.001	1171217618	p < 0.001
niccdcd	0.152405	p < 0.001	1020205543	p < 0.001
rel_10c	0.253673	p < 0.001	1313622667	p < 0.001
tpi_10c	0.266257	p < 0.001	1334702320	p < 0.001
tpi_cls10c	0.192992	p < 0.001	1230563975	p < 0.001
tpi_sd10c	0.266087	p < 0.001	1334516995	p < 0.001
twi16c	0.184688	p < 0.001	704635397	p < 0.001
random	0.008725	p = 0.108	956435404	p = 0.515

Region 8 All - Riverine Section 5				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.127622	p < 0.001	959825565	p < 0.001
drcwet	0.183857	p < 0.001	1010077261	p < 0.001
e_hyd_min_wt	0.216916	p < 0.001	1014962773	p < 0.001
e_trail_dist	0.161685	p < 0.001	930747253	p < 0.001
ed_conf	0.15943	p < 0.001	721931637	p < 0.001
ed_h2	0.182404	p < 0.001	962048797	p < 0.001
ed_h7	0.114573	p < 0.001	937360221	p < 0.001
eldrop16c	0.131731	p < 0.001	996273098	p < 0.001
elev_2_drainh	0.157109	p < 0.001	1024104096	p < 0.001
niccdcd	0.116756	p < 0.001	740764886	p < 0.001
rel_10c	0.19327	p < 0.001	1060994869	p < 0.001
rng_32c	0.151213	p < 0.001	745929350	p < 0.001
slpvr_32c	0.13011	p < 0.001	887792579	p < 0.001
std_32c	0.164532	p < 0.001	729851687	p < 0.001
tpi_5c	0.193678	p < 0.001	1072173603	p < 0.001
tpi_cls10c	0.153309	p < 0.001	1036174057	p < 0.001
tpi_sd5c	0.193547	p < 0.001	1071981371	p < 0.001
tri_32c	0.130124	p < 0.001	887810814	p < 0.001
twi16c	0.110718	p < 0.001	714578427	p < 0.001
random	0.00551	p = 0.594	850767438	p = 0.566

Region 8 All - Riverine Section 6				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_trail_dist	0.183585	p < 0.001	1002803927	p < 0.001
drcdry	0.199806	p < 0.001	1508692562	p < 0.001
e_hyd_min_wt	0.178755	p < 0.001	1409295782	p < 0.001
ed_h2	0.169029	p < 0.001	1366409864	p < 0.001
ed_h5	0.122752	p < 0.001	1076699837	p < 0.001
ed_h6	0.233674	p < 0.001	946834639	p < 0.001
eldrop16c	0.215585	p < 0.001	1519859206	p < 0.001
elev_2_conf	0.171359	p < 0.001	1474520593	p < 0.001
elev_2_drainh	0.139801	p < 0.001	1342504015	p < 0.001
niccdcd	0.135831	p < 0.001	1034628715	p < 0.001
rel_8c	0.23306	p < 0.001	1563742764	p < 0.001
rng_32c	0.125181	p < 0.001	1395251335	p < 0.001
slpvr_32c	0.164805	p < 0.001	1411969838	p < 0.001
std_32c	0.155347	p < 0.001	1418678614	p < 0.001
tpi_10c	0.193849	p < 0.001	1557047935	p < 0.001
tpi_cls5c	0.159568	p < 0.001	1486707256	p < 0.001
tpi_sd10c	0.193488	p < 0.001	1556488620	p < 0.001
tri_32c	0.164201	p < 0.001	1411859818	p < 0.001
twi16c	0.143435	p < 0.001	965944041	p < 0.001
random	0.005605	p = 0.463	1208375790	p = 0.593

Region 8 All - Riverine Section 7				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aspect	0.162902	p < 0.001	830286186	p < 0.001
aws050	0.157677	p < 0.001	811369853	p < 0.001
cd_drnh	0.156383	p < 0.001	841121229	p < 0.001
cd_h5	0.197028	p < 0.001	913099917	p < 0.001
drcdry	0.253318	p < 0.001	936549468	p < 0.001
e_hyd_min	0.287157	p < 0.001	958929165	p < 0.001
ed_h2	0.232835	p < 0.001	890636236	p < 0.001
ed_h7	0.182582	p < 0.001	824402437	p < 0.001
eldrop16c	0.188941	p < 0.001	912705109	p < 0.001
elev_2_conf	0.172887	p < 0.001	918914458	p < 0.001
elev_2_strm	0.157905	p < 0.001	858321421	p < 0.001
niccdcd	0.177317	p < 0.001	618848186	p < 0.001
rel_8c	0.244702	p < 0.001	967690953	p < 0.001
slope_deg	0.15451	p < 0.001	805302722	p < 0.001
slpvr_8c	0.206084	p < 0.001	575949503	p < 0.001
tpi_100c	0.21404	p < 0.001	933096481	p < 0.001
tpi_cls5c	0.182142	p < 0.001	899611404	p < 0.001
tpi_sd100c	0.214391	p < 0.001	933349314	p < 0.001
tri_8c	0.199077	p < 0.001	589125044	p < 0.001
twi16c	0.154273	p < 0.001	603254205	p < 0.001
random	0.007209	p = 0.328	739323810	p = 0.650

Region 8 All - Riverine Section 8				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
cd_conf	0.175601	p < 0.001	2181817969	p < 0.001
cd_h4	0.177566	p < 0.001	2181117309	p < 0.001
cd_h5	0.197812	p < 0.001	2261940542	p < 0.001
cd_h6	0.244814	p < 0.001	1703810101	p < 0.001
drcwet	0.194653	p < 0.001	2165873429	p < 0.001
e_hyd_min_wt	0.316817	p < 0.001	2459471604	p < 0.001
e_trail_dist	0.398589	p < 0.001	1022107774	p < 0.001
ed_h2	0.256914	p < 0.001	2266894347	p < 0.001
eldrop32c	0.203273	p < 0.001	2303706841	p < 0.001
elev_2_conf	0.176702	p < 0.001	2250054500	p < 0.001
rel_10c	0.210687	p < 0.001	2269877895	p < 0.001
slpvr_8c	0.143411	p < 0.001	1656719053	p < 0.001
tpi_250c	0.219429	p < 0.001	2277722796	p < 0.001
tpi_cls10c	0.166971	p < 0.001	2175558765	p < 0.001
tpi_sd250c	0.21905	p < 0.001	2276805930	p < 0.001
random	0.005089	p = 0.503	1773063764	p = 0.505

Region 8 All - Riverine Section 9				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.181545	p < 0.001	2887344158	p < 0.001
cd_conf	0.317357	p < 0.001	4014945285	p < 0.001
cd_drnh	0.176183	p < 0.001	3602362911	p < 0.001
cd_h4	0.216152	p < 0.001	3620964671	p < 0.001
cd_h6	0.3245	p < 0.001	3748055677	p < 0.001
drcwet	0.214142	p < 0.001	3680212653	p < 0.001
e_hyd_min	0.326692	p < 0.001	4145224044	p < 0.001
ed_h2	0.32051	p < 0.001	4092589724	p < 0.001
eldrop32c	0.299529	p < 0.001	4046143883	p < 0.001
elev_2_conf	0.340669	p < 0.001	4151995919	p < 0.001
elev_2_strm	0.333805	p < 0.001	4172318497	p < 0.001
rel_16c	0.372669	p < 0.001	4365381547	p < 0.001
slpvr_8c	0.191971	p < 0.001	2266784665	p < 0.001
tpi_10c	0.254187	p < 0.001	4063006403	p < 0.001
tpi_cls10c	0.254261	p < 0.001	3878575354	p < 0.001
tpi_sd10c	0.254842	p < 0.001	4064936401	p < 0.001
tri_10c	0.180722	p < 0.001	2319013263	p < 0.001
vrf_32c	0.18371	p < 0.001	2355705304	p < 0.001
random	0.005526	p = 0.321	2960292274	p = 0.357

Region 8 All - Upland Section 1				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_hyd_min_wt	0.356995	p < 0.001	739585417	p < 0.001
c_trail_dist	0.34946	p < 0.001	723289499	p < 0.001
cd_conf	0.403053	p < 0.001	585964851	p < 0.001
cd_h2	0.376445	p < 0.001	701761162	p < 0.001
cd_h4	0.409776	p < 0.001	604673689	p < 0.001
ed_h5	0.430145	p < 0.001	520702360	p < 0.001
ed_h7	0.583594	p < 0.001	396324049	p < 0.001
eldrop32c	0.325323	p < 0.001	770668054	p < 0.001
elev_2_conf	0.404835	p < 0.001	588650233	p < 0.001
elev_2_drainh	0.424756	p < 0.001	608253408	p < 0.001
elev_2_strm	0.49292	p < 0.001	472996723	p < 0.001
rel_32c	0.402196	p < 0.001	625979484	p < 0.001
tpi_250c	0.546001	p < 0.001	433140083	p < 0.001
tpi_cls250c	0.488708	p < 0.001	584816910	p < 0.001
tpi_sd250c	0.54621	p < 0.001	433085039	p < 0.001
random	0.007857	p = 0.176	1222613278	p = 0.221

Region 8 All - Upland Section 2				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_hyd_min	0.47441	p < 0.001	927239255	p < 0.001
c_trail_dist	0.551429	p < 0.001	685239441	p < 0.001
cd_conf	0.505218	p < 0.001	830761879	p < 0.001
cd_h2	0.439674	p < 0.001	944728388	p < 0.001
cd_h4	0.492644	p < 0.001	961421326	p < 0.001
cd_h5	0.514871	p < 0.001	932874894	p < 0.001
cd_h7	0.617226	p < 0.001	723673537	p < 0.001
eldrop32c	0.368361	p < 0.001	1167562225	p < 0.001
elev_2_conf	0.490775	p < 0.001	974304518	p < 0.001
elev_2_strm	0.575886	p < 0.001	694223278	p < 0.001
rng_32c	0.520032	p < 0.001	809255362	p < 0.001
slope_deg	0.386275	p < 0.001	1137649089	p < 0.001
slpvr_16c	0.346317	p < 0.001	1420771390	p < 0.001
std_32c	0.519561	p < 0.001	841301965	p < 0.001
tpi_250c	0.524617	p < 0.001	853767559	p < 0.001
tpi_cls250c	0.496503	p < 0.001	1038364524	p < 0.001
tpi_sd250c	0.524554	p < 0.001	854088001	p < 0.001
tri_8c	0.385032	p < 0.001	1242269415	p < 0.001
vrf_32c	0.328804	p < 0.001	1390386220	p < 0.001
random	0.005709	p = 0.323	2267251984	p = 0.460

Region 8 All - Upland Section 3				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.38745	p < 0.001	338552751	p < 0.001
cd_drnh	0.356882	p < 0.001	320861224	p < 0.001
cd_h4	0.362272	p < 0.001	152716643	p < 0.001
drcdry	0.301973	p < 0.001	184090884	p < 0.001
e_hyd_min	0.411173	p < 0.001	133713405	p < 0.001
e_trail_dist	0.442462	p < 0.001	338218882	p < 0.001
ed_conf	0.610736	p < 0.001	71239100	p < 0.001
ed_h2	0.432876	p < 0.001	123346461	p < 0.001
ed_h7	0.651644	p < 0.001	118468128	p < 0.001
elev_2_conf	0.417205	p < 0.001	137249910	p < 0.001
elev_2_drainh	0.528907	p < 0.001	103444253	p < 0.001
elev_2_strm	0.720994	p < 0.001	64675505	p < 0.001
rel_32c	0.370351	p < 0.001	145287934	p < 0.001
slpvr_32c	0.47133	p < 0.001	363736752	p < 0.001
tpi_100c	0.589939	p < 0.001	97814877	p < 0.001
tpi_cls250c	0.522668	p < 0.001	96832200	p < 0.001
tpi_sd100c	0.590049	p < 0.001	97821901	p < 0.001
tri_32c	0.439281	p < 0.001	355664974	p < 0.001
random	0.01215	p = 0.212	248667169	p = 0.767

Region 8 All - Upland Section 4				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
cd_conf	0.445637	p < 0.001	376101908	p < 0.001
cd_h5	0.404837	p < 0.001	393366227	p < 0.001
e_hyd_min	0.434385	p < 0.001	432913562	p < 0.001
ed_h2	0.450848	p < 0.001	422454273	p < 0.001
ed_h4	0.401916	p < 0.001	523253968	p < 0.001
ed_h7	0.331083	p < 0.001	645518889	p < 0.001
elev_2_conf	0.255749	p < 0.001	590223354	p < 0.001
elev_2_drainh	0.210941	p < 0.001	575623306	p < 0.001
elev_2_strm	0.315453	p < 0.001	582896145	p < 0.001
slpvr_32c	0.282882	p < 0.001	1133772136	p < 0.001
tpi_100c	0.349501	p < 0.001	543055252	p < 0.001
tpi_sd100c	0.349579	p < 0.001	542923133	p < 0.001
tri_32c	0.281966	p < 0.001	1132142818	p < 0.001
random	0.007072	p = 0.310	858432243	p = 0.622

Region 8 All - Upland Section 5				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.21533	p < 0.001	1075993669	p < 0.001
c_trail_dist	0.286276	p < 0.001	636418922	p < 0.001
cd_conf	0.501951	p < 0.001	304176113	p < 0.001
cd_h4	0.513924	p < 0.001	290050938	p < 0.001
cd_h5	0.38948	p < 0.001	414031461	p < 0.001
cd_h7	0.352836	p < 0.001	564753988	p < 0.001
e_hyd_min_wt	0.418113	p < 0.001	415321826	p < 0.001
ed_h2	0.400378	p < 0.001	432793612	p < 0.001
eldrop32c	0.193607	p < 0.001	620262251	p < 0.001
elev_2_conf	0.378805	p < 0.001	444360016	p < 0.001
elev_2_strm	0.329358	p < 0.001	565327441	p < 0.001
rel_32c	0.237706	p < 0.001	545437052	p < 0.001
slpvr_32c	0.211553	p < 0.001	1015510034	p < 0.001
tpi_100c	0.381056	p < 0.001	472514098	p < 0.001
tpi_sd100c	0.381595	p < 0.001	472321879	p < 0.001
tri_32c	0.210099	p < 0.001	1014168281	p < 0.001
random	0.006004	p = 0.500	837806984	p = 0.640

Region 8 All - Upland Section 6				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
cd_conf	0.418389	p < 0.001	645195974	p < 0.001
cd_h5	0.320458	p < 0.001	734401879	p < 0.001
cd_h7	0.480529	p < 0.001	651334217	p < 0.001
e_hyd_min_wt	0.479278	p < 0.001	540116000	p < 0.001
ed_h2	0.492517	p < 0.001	548418410	p < 0.001
ed_h4	0.388084	p < 0.001	715947567	p < 0.001
elev_2_conf	0.35427	p < 0.001	742348104	p < 0.001
elev_2_drainh	0.214395	p < 0.001	926419005	p < 0.001
elev_2_strm	0.481295	p < 0.001	581893623	p < 0.001
rel_32c	0.303141	p < 0.001	742794902	p < 0.001
slpvr_32c	0.266158	p < 0.001	1728100190	p < 0.001
tpi_100c	0.455844	p < 0.001	598284561	p < 0.001
tpi_cls250c	0.24482	p < 0.001	970849640	p < 0.001
tpi_sd100c	0.455832	p < 0.001	598104212	p < 0.001
tri_32c	0.264413	p < 0.001	1726113367	p < 0.001
random	0.005545	p = 0.457	1315546023	p = 0.560

Region 8 All - Upland Section 7				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.286221	p < 0.001	1018426519	p < 0.001
c_trail_dist	0.223202	p < 0.001	637904545	p < 0.001
cd_conf	0.385296	p < 0.001	389876573	p < 0.001
cd_drnh	0.230384	p < 0.001	561744889	p < 0.001
cd_h4	0.371807	p < 0.001	416927457	p < 0.001
cd_h5	0.300088	p < 0.001	499557453	p < 0.001
cd_h7	0.313922	p < 0.001	458003445	p < 0.001
e_hyd_min	0.572326	p < 0.001	270581000	p < 0.001
ed_h2	0.591546	p < 0.001	267140632	p < 0.001
eldrop32c	0.239697	p < 0.001	541313214	p < 0.001
elev_2_conf	0.339434	p < 0.001	447018049	p < 0.001
elev_2_drainh	0.355259	p < 0.001	445651152	p < 0.001
elev_2_strm	0.400831	p < 0.001	351964403	p < 0.001
niccdcd	0.226415	p < 0.001	571435093	p < 0.001
rel_32c	0.299864	p < 0.001	458528677	p < 0.001
rng_16c	0.186279	p < 0.001	655913085	p < 0.001
tpi_100c	0.402456	p < 0.001	428622821	p < 0.001
tpi_sd100c	0.402461	p < 0.001	428431203	p < 0.001
random	0.006746	p = 0.404	779071303	p = 0.636

Region 8 All - Upland Section 8				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.300034	p < 0.001	3822608522	p < 0.001
c_hyd_min	0.426992	p < 0.001	1375948512	p < 0.001
cd_conf	0.399833	p < 0.001	1482907500	p < 0.001
cd_drnh	0.333594	p < 0.001	1813276471	p < 0.001
cd_h2	0.483061	p < 0.001	1254012723	p < 0.001
cd_h4	0.273316	p < 0.001	1966353580	p < 0.001
cd_h5	0.250352	p < 0.001	2074031399	p < 0.001
cd_h7	0.479948	p < 0.001	1433645441	p < 0.001
drcdry	0.251021	p < 0.001	2304429823	p < 0.001
e_trail_dist	0.515552	p < 0.001	1378985078	p < 0.001
eldrop32c	0.328843	p < 0.001	1776546468	p < 0.001
elev_2_conf	0.41629	p < 0.001	1509815194	p < 0.001
elev_2_drainh	0.310727	p < 0.001	2100237197	p < 0.001
elev_2_strm	0.469691	p < 0.001	1349412430	p < 0.001
rel_32c	0.318823	p < 0.001	1808139662	p < 0.001
rng_16c	0.305964	p < 0.001	2029646596	p < 0.001
slope_deg	0.247293	p < 0.001	2133739360	p < 0.001
std_16c	0.276983	p < 0.001	2102022748	p < 0.001
tpi_50c	0.314165	p < 0.001	1907774325	p < 0.001
tpi_sd50c	0.31478	p < 0.001	1905483773	p < 0.001
random	0.005201	p = 0.368	3076300210	p = 0.443

Region 8 All - Upland Section 9				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.270703	p < 0.001	4707392435	p < 0.001
c_trail_dist	0.257461	p < 0.001	2969573968	p < 0.001
cd_conf	0.227789	p < 0.001	2880472043	p < 0.001
cd_h2	0.310991	p < 0.001	2484304702	p < 0.001
cd_h4	0.257993	p < 0.001	2792416432	p < 0.001
cd_h5	0.231206	p < 0.001	2885996007	p < 0.001
e_hyd_min	0.262605	p < 0.001	2608157504	p < 0.001
ed_drnh	0.160138	p < 0.001	4909970276	p < 0.001
ed_h6	0.310916	p < 0.001	4976623000	p < 0.001
elev_2_conf	0.194416	p < 0.001	2908723672	p < 0.001
elev_2_drainh	0.272817	p < 0.001	2664017727	p < 0.001
elev_2_strm	0.16647	p < 0.001	3269544889	p < 0.001
flowdir	0.174845	p < 0.001	4674339390	p < 0.001
rel_32c	0.275896	p < 0.001	2588098538	p < 0.001
tpi_100c	0.419011	p < 0.001	1873448590	p < 0.001
tpi_cls100c	0.350141	p < 0.001	2440393205	p < 0.001
tpi_sd100c	0.418731	p < 0.001	1875209860	p < 0.001
vrf_32c	0.147511	p < 0.001	3356666806	p < 0.001
random	0.00453	p = 0.466	4036300814	p = 0.543

Region 9/10 - Riverine Section 1				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.411162	p < 0.001	501493354.9	p < 0.001
c_hyd_min_wt	0.406407	p < 0.001	1234201672	p < 0.001
c_trail_dist	0.29262	p < 0.001	805958417.9	p < 0.001
cd_conf	0.379827	p < 0.001	1184990830	p < 0.001
cd_h2	0.37278	p < 0.001	1218697258	p < 0.001
cd_h4	0.349469	p < 0.001	1155197435	p < 0.001
cd_h5	0.324243	p < 0.001	1155064437	p < 0.001
cd_h6	0.361292	p < 0.001	1073052356	p < 0.001
drcdry	0.32571	p < 0.001	1048741137	p < 0.001
eldrop32c	0.379042	p < 0.001	1181861326	p < 0.001
elev_2_conf	0.442387	p < 0.001	1250131527	p < 0.001
elev_2_strm	0.416986	p < 0.001	1170009208	p < 0.001
niccdcd	0.487029	p < 0.001	1163197774	p < 0.001
rng_32c	0.476807	p < 0.001	1150028025	p < 0.001
slope_pct	0.304671	p < 0.001	1112318524	p < 0.001
std_32c	0.425469	p < 0.001	1156133071	p < 0.001
tpi_100c	0.380896	p < 0.001	1173062218	p < 0.001
tpi_cls250c	0.280301	p < 0.001	1053068716	p < 0.001
tpi_sd100c	0.380704	p < 0.001	1172906495	p < 0.001
tri_8c	0.276325	p < 0.001	1045261546	p < 0.001
random	0.005823	p = 0.553	798135199.3	p = 0.545

Region 9/10 - Riverine Section 2				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.159541	p < 0.001	1400347228	p < 0.050
c_trail_dist	0.217442	p < 0.001	1561277247	p < 0.001
cd_h4	0.155047	p < 0.001	1664914283	p < 0.001
drcdry	0.153248	p < 0.001	1710503411	p < 0.001
e_hyd_min_wt	0.297029	p < 0.001	1863711268	p < 0.001
ed_drnh	0.152948	p < 0.001	1689551485	p < 0.001
ed_h2	0.282543	p < 0.001	1825761826	p < 0.001
ed_h6	0.43074	p < 0.001	830548308.9	p < 0.001
eldrop32c	0.241053	p < 0.001	1892339279	p < 0.001
elev_2_conf	0.183363	p < 0.001	1782501571	p < 0.001
niccdcd	0.194898	p < 0.001	1169657627	p < 0.001
rel_16c	0.31834	p < 0.001	2008963982	p < 0.001
tpi_10c	0.311486	p < 0.001	2024130572	p < 0.001
tpi_cls10c	0.255846	p < 0.001	1910556194	p < 0.001
tpi_sd10c	0.311562	p < 0.001	2024460979	p < 0.001
twi32c	0.171401	p < 0.001	1068040636	p < 0.001
random	0.005815	p = 0.404	1415343561	p = 0.598

Region 9/10 - Riverine Section 3				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aspect	0.289031	p < 0.001	634745163.1	p < 0.001
aws050	0.35425	p < 0.001	718830129.6	p < 0.001
c_hyd_min	0.281346	p < 0.001	649056403.9	p < 0.001
cd_conf	0.307523	p < 0.001	682440599.3	p < 0.001
cd_drnh	0.273725	p < 0.001	638799948.7	p < 0.001
cd_h2	0.324975	p < 0.001	700831616.6	p < 0.001
cd_h5	0.298006	p < 0.001	611356453.6	p < 0.001
cd_h7	0.326749	p < 0.001	672771877.2	p < 0.001
drcdry	0.314556	p < 0.001	675583553.3	p < 0.001
e_trail_dist	0.392155	p < 0.001	311350103.7	p < 0.001
ed_h4	0.34038	p < 0.001	299835215.2	p < 0.001
eldrop10c	0.327731	p < 0.001	680018524.4	p < 0.001
elev_2_drainh	0.277474	p < 0.001	390479451.2	p < 0.001
elev_2_strm	0.344372	p < 0.001	690426480.8	p < 0.001
flowdir	0.24372	p < 0.001	611405059	p < 0.001
niccdcd	0.341343	p < 0.001	693163562.6	p < 0.001
rel_32c	0.285061	p < 0.001	636828309.1	p < 0.001
rng_16c	0.319804	p < 0.001	693717586.7	p < 0.001
slope_pct	0.310562	p < 0.001	653668004.7	p < 0.001
slpvr_16c	0.286931	p < 0.001	613459964.6	p < 0.001
std_32c	0.31664	p < 0.001	695920871	p < 0.001
tpi_250c	0.261644	p < 0.001	580887362.2	p < 0.001
tpi_sd250c	0.261726	p < 0.001	580880107.7	p < 0.001
tri_16c	0.288785	p < 0.001	615813327.8	p < 0.001
twi32c	0.317989	p < 0.001	326501664.7	p < 0.001
vrf_16c	0.314613	p < 0.001	676069625.1	p < 0.001
random	0.012869	p = 0.029	493423968	p = 0.195

Region 9/10 - Riverine Section 4				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
cd_h2	0.348317	p < 0.001	137351768	p < 0.001
cd_h4	0.287397	p < 0.001	123251185.2	p < 0.001
cd_h7	0.283971	p < 0.001	83030343.71	p < 0.001
drcwet	0.356521	p < 0.001	136329318.9	p < 0.001
e_hyd_min_wt	0.441474	p < 0.001	153095749.4	p < 0.001
e_trail_dist	0.315916	p < 0.001	110803726.1	p < 0.001
ed_drnh	0.252816	p < 0.001	70711212.8	p < 0.001
eldrop10c	0.229334	p < 0.001	122769349.8	p < 0.001
rel_10c	0.374731	p < 0.001	138871657.6	p < 0.001
rng_32c	0.325834	p < 0.001	74372244.53	p < 0.001
slpvr_16c	0.28645	p < 0.001	79095918.93	p < 0.001
std_32c	0.270791	p < 0.001	78560650.3	p < 0.001
tpi_10c	0.339542	p < 0.001	141548190.6	p < 0.001
tpi_cls10c	0.339721	p < 0.001	135219674.9	p < 0.001
tpi_sd10c	0.340071	p < 0.001	141588534.7	p < 0.001
tri_10c	0.294686	p < 0.001	82365583.25	p < 0.001
vrf_32c	0.312072	p < 0.001	66376164.1	p < 0.001
random	0.017674	p = 0.229	96345818.43	p = 0.544

Region 9/10 - Riverine Section 5				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
cd_drnh	0.303052	p < 0.001	1204067204	p < 0.001
drcdry	0.267197	p < 0.001	2177026821	p < 0.001
e_hyd_min	0.338377	p < 0.001	2531755357	p < 0.001
e_trail_dist	0.249183	p < 0.001	2172913640	p < 0.001
ed_h2	0.339521	p < 0.001	2493090098	p < 0.001
ed_h6	0.324255	p < 0.001	1392631802	p < 0.001
elev_2_conf	0.235131	p < 0.001	2194892098	p < 0.001
elev_2_drainh	0.338944	p < 0.001	2441238595	p < 0.001
rel_16c	0.299201	p < 0.001	2417160177	p < 0.001
rng_10c	0.240463	p < 0.001	1277754523	p < 0.001
slpvr_8c	0.252194	p < 0.001	1205159456	p < 0.001
std_8c	0.253992	p < 0.001	1237927463	p < 0.001
tpi_10c	0.36842	p < 0.001	2517460166	p < 0.001
tpi_cls10c	0.339156	p < 0.001	2437712437	p < 0.001
tpi_sd10c	0.368117	p < 0.001	2517253880	p < 0.001
tri_8c	0.249275	p < 0.001	1212150875	p < 0.001
random	0.006686	p = 0.237	1790169236	p = 0.364

Region 9/10 - Riverine Section 6				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.158433	p < 0.001	1495565636	p < 0.001
c_hyd_min_wt	0.305902	p < 0.001	2319782288	p < 0.001
c_trail_dist	0.332272	p < 0.001	2208154113	p < 0.001
cd_conf	0.197768	p < 0.001	2076740002	p < 0.001
cd_drnh	0.184468	p < 0.001	1988027405	p < 0.001
cd_h2	0.281473	p < 0.001	2293582201	p < 0.001
cd_h4	0.252335	p < 0.001	2210400961	p < 0.001
cd_h5	0.185405	p < 0.001	1992022703	p < 0.001
cd_h7	0.298464	p < 0.001	2284126802	p < 0.001
drcdry	0.23843	p < 0.001	2119822950	p < 0.001
eldrop16c	0.267944	p < 0.001	2216314838	p < 0.001
elev_2_conf	0.189823	p < 0.001	2031315054	p < 0.001
elev_2_drainh	0.162367	p < 0.001	1838040126	p < 0.001
elev_2_strm	0.280294	p < 0.001	2151962506	p < 0.001
rel_16c	0.236682	p < 0.001	2170735157	p < 0.001
slope_pct	0.161566	p < 0.001	1959985830	p < 0.001
tpi_10c	0.207267	p < 0.001	2025440828	p < 0.001
tpi_cls10c	0.15932	p < 0.001	1949200317	p < 0.001
tpi_sd10c	0.207329	p < 0.001	2025877123	p < 0.001
random	0.01076	p = 0.021	1628759971	p = 0.215

Region 9/10 - Riverine Section 7				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.213903	p < 0.001	657992978.6	p < 0.001
c_trail_dist	0.365344	p < 0.001	371938214.1	p < 0.001
cd_conf	0.204053	p < 0.001	680671483.4	p < 0.001
cd_drnh	0.218986	p < 0.001	618371990.4	p < 0.001
cd_h2	0.385039	p < 0.001	828091304.3	p < 0.001
cd_h3	0.216781	p < 0.001	683663307.1	p < 0.001
cd_h4	0.219568	p < 0.001	698789276.9	p < 0.001
drcdry	0.365933	p < 0.001	775822685.5	p < 0.001
e_hyd_min	0.391273	p < 0.001	810645427.3	p < 0.001
ed_h7	0.380117	p < 0.001	367735700	p < 0.001
eldrop32c	0.33817	p < 0.001	792680846.3	p < 0.001
elev_2_conf	0.327775	p < 0.001	770515400.3	p < 0.001
elev_2_strm	0.21721	p < 0.001	477288101.3	p < 0.001
niccdcd	0.300413	p < 0.001	390806474.8	p < 0.001
rel_16c	0.370134	p < 0.001	827277432.3	p < 0.001
slpvr_32c	0.215463	p < 0.001	669657696.5	p < 0.001
tpi_10c	0.365473	p < 0.001	813554994.8	p < 0.001
tpi_cls10c	0.282047	p < 0.001	764113174.4	p < 0.001
tpi_sd10c	0.36533	p < 0.001	813547429	p < 0.001
tri_32c	0.21516	p < 0.001	669672158.3	p < 0.001
twi32c	0.226613	p < 0.001	358281759.8	p < 0.001
random	0.006647	p = 0.540	543817838.6	p = 0.479

Region 9/10 - Riverine Section 8				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.221518	p < 0.001	366784106	p < 0.001
cd_drnh	0.162436	p < 0.001	373324743.4	p < 0.001
cd_h2	0.208958	p < 0.001	394959706.6	p < 0.001
cd_h4	0.155303	p < 0.001	381817055.1	p < 0.001
cd_h6	0.310858	p < 0.001	422863523.4	p < 0.001
drcdry	0.17119	p < 0.001	378120173.5	p < 0.001
e_hyd_min	0.272208	p < 0.001	424549679.7	p < 0.001
e_trail_dist	0.299071	p < 0.001	220225711.2	p < 0.001
ed_h5	0.198434	p < 0.001	393372001.8	p < 0.001
eldrop16c	0.162383	p < 0.001	390163770.2	p < 0.001
elev_2_strm	0.191889	p < 0.001	368557402.8	p < 0.001
rel_16c	0.247823	p < 0.001	419334283.4	p < 0.001
rng_32c	0.196718	p < 0.001	288072289.4	p < 0.001
slpvr_32c	0.176643	p < 0.001	282257493.8	p < 0.001
std_32c	0.23168	p < 0.001	289187792.2	p < 0.001
tpi_50c	0.269069	p < 0.001	397577751.9	p < 0.001
tpi_cls50c	0.266997	p < 0.001	387792099.6	p < 0.001
tpi_sd50c	0.268734	p < 0.001	397450590.7	p < 0.001
tri_32c	0.176644	p < 0.001	282593326	p < 0.001
twi16c	0.161344	p < 0.001	261450484.2	p < 0.001
random	0.010359	p = 0.250	323084170.9	p = 0.712

Region 9/10 - Riverine Section 9				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aspect	0.259823	p < 0.001	66727289.52	p < 0.001
aws050	0.335844	p < 0.001	69182597.81	p < 0.001
c_hyd_min	0.334767	p < 0.001	72176087.14	p < 0.001
cd_h2	0.27249	p < 0.001	68570133.55	p < 0.001
cd_h5	0.230563	p < 0.001	62914924.33	p < 0.001
drcdry	0.288802	p < 0.001	70865503.34	p < 0.001
e_trail_dist	0.308193	p < 0.001	64518939.5	p < 0.001
ed_conf	0.313674	p < 0.001	65291745.16	p < 0.001
ed_drnh	0.330756	p < 0.001	66504506.11	p < 0.001
ed_h4	0.411325	p < 0.001	40852736.35	p < 0.001
ed_h6	0.380353	p < 0.001	39652437.79	p < 0.001
eldrop16c	0.319303	p < 0.001	72514555.5	p < 0.001
elev_2_drainh	0.338385	p < 0.001	73881824.91	p < 0.001
elev_2_strm	0.225035	p < 0.001	64372416.63	p < 0.001
niccdcd	0.277108	p < 0.001	44992488.28	p < 0.001
rel_32c	0.269316	p < 0.001	71233319.04	p < 0.001
rng_16c	0.325101	p < 0.001	65382284.37	p < 0.001
slpvr_16c	0.241256	p < 0.001	57771696.21	p < 0.050
std_32c	0.393681	p < 0.001	65628244.28	p < 0.001
tpi_100c	0.282149	p < 0.001	42460657.79	p < 0.001
tpi_sd100c	0.281998	p < 0.001	42465513.6	p < 0.001
tri_16c	0.240779	p < 0.001	57716823.01	p < 0.010
twi16c	0.214179	p < 0.001	41337983.77	p < 0.001
vrf_32c	0.239797	p < 0.001	66886608.65	p < 0.001
random	0.023587	p = 0.198	54297204.39	p = 0.096

Region 9/10 - Riverine Section 10				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.22606	p < 0.001	614440563.8	p < 0.001
cd_drnh	0.311158	p < 0.001	384397958.3	p < 0.001
drcdry	0.217191	p < 0.001	723730386.3	p < 0.001
e_hyd_min_wt	0.239216	p < 0.001	698552824.6	p < 0.001
e_trail_dist	0.191746	p < 0.001	653846225.3	p < 0.001
ed_h2	0.204264	p < 0.001	658160117.9	p < 0.001
ed_h4	0.179741	p < 0.001	470148072	p < 0.001
ed_h6	0.455641	p < 0.001	796054774	p < 0.001
elev_2_drainh	0.259187	p < 0.001	729319073.3	p < 0.001
niccdcd	0.187389	p < 0.001	467852488.5	p < 0.001
rel_32c	0.230452	p < 0.001	752582215.8	p < 0.001
rng_32c	0.28429	p < 0.001	405794317.3	p < 0.001
slpvr_32c	0.223373	p < 0.001	441269587.1	p < 0.001
std_32c	0.240722	p < 0.001	413713667.2	p < 0.001
tpi_250c	0.322792	p < 0.001	802448460.8	p < 0.001
tpi_cls100c	0.306012	p < 0.001	759398067.4	p < 0.001
tpi_sd250c	0.322476	p < 0.001	802161153.5	p < 0.001
tri_32c	0.224122	p < 0.001	440964290.7	p < 0.001
random	0.008809	p=0.211	573752847.4	p=0.408

Region 9/10 - Riverine Section 11				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
cd_h2	0.2475	p < 0.001	1613014441	p < 0.001
cd_h3	0.196605	p < 0.001	1515523795	p < 0.001
cd_h4	0.232439	p < 0.001	1546051691	p < 0.001
cd_h7	0.335438	p < 0.001	1633651169	p < 0.001
drcdry	0.194364	p < 0.001	1471562064	p < 0.001
e_hyd_min	0.265226	p < 0.001	1490821472	p < 0.001
e_trail_dist	0.227097	p < 0.001	999461873.6	p < 0.001
eldrop32c	0.287475	p < 0.001	1670047320	p < 0.001
elev_2_conf	0.208778	p < 0.001	1520811283	p < 0.001
elev_2_strm	0.39207	p < 0.001	1745312516	p < 0.001
rel_10c	0.239407	p < 0.001	1551592379	p < 0.001
rng_32c	0.274404	p < 0.001	1550530259	p < 0.001
slope_pct	0.205604	p < 0.001	1508146577	p < 0.001
std_32c	0.271365	p < 0.001	1558712776	p < 0.001
tpi_10c	0.237306	p < 0.001	1496273508	p < 0.001
tpi_cls10c	0.202133	p < 0.001	1456068875	p < 0.001
tpi_sd10c	0.237033	p < 0.001	1495990415	p < 0.001
tri_16c	0.189095	p < 0.001	1412596210	p < 0.001
twi32c	0.236404	p < 0.001	746951185.1	p < 0.001
random	0.007152	p=0.241	1163162301	p=0.457

Region 9/10 - Riverine Section 12				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.305092	p < 0.001	923356149.1	p < 0.001
c_trail_dist	0.417633	p < 0.001	431296611.6	p < 0.001
cd_conf	0.280339	p < 0.001	535481305.2	p < 0.001
cd_drnh	0.500047	p < 0.001	366982962.9	p < 0.001
cd_h5	0.303965	p < 0.001	520344683.5	p < 0.001
drcdry	0.320856	p < 0.001	1089243143	p < 0.001
e_hyd_min_wt	0.293633	p < 0.001	1077252588	p < 0.001
ed_h2	0.259623	p < 0.001	1032948579	p < 0.001
ed_h4	0.333756	p < 0.001	656351316.7	p < 0.001
ed_h7	0.572432	p < 0.001	1232959914	p < 0.001
elev_2_drainh	0.50079	p < 0.001	429824688.3	p < 0.001
elev_2_strm	0.478563	p < 0.001	1130730016	p < 0.001
niccdcd	0.311678	p < 0.001	630774922.5	p < 0.001
rng_32c	0.36286	p < 0.001	500208347.8	p < 0.001
slpvr_32c	0.275948	p < 0.001	678771816.4	p < 0.001
std_32c	0.292666	p < 0.001	596586714.9	p < 0.001
tpi_100c	0.254555	p < 0.001	770857596.1	p < 0.001
tpi_sd100c	0.254964	p < 0.001	770486277.9	p < 0.001
tri_32c	0.275868	p < 0.001	678365624.2	p < 0.001
random	0.014132	p=0.001	817801239.4	p=0.198

Region 9/10 - Riverine Section 13				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.317891	p < 0.001	518212824.6	p < 0.001
cd_h2	0.430805	p < 0.001	656406196.9	p < 0.001
e_hyd_min_wt	0.413255	p < 0.001	625837365	p < 0.001
e_trail_dist	0.413194	p < 0.001	244533572.5	p < 0.001
ed_conf	0.355615	p < 0.001	590693411.3	p < 0.001
ed_h3	0.363907	p < 0.001	582541120.1	p < 0.001
ed_h4	0.351612	p < 0.001	533812181	p < 0.001
ed_h6	0.421403	p < 0.001	262336353.4	p < 0.001
eldrop16c	0.39438	p < 0.001	633665374.5	p < 0.001
flw_acum	0.364815	p < 0.001	236181467.4	p < 0.001
niccdcd	0.3572	p < 0.001	529018684.5	p < 0.001
rel_10c	0.483533	p < 0.001	640958328.5	p < 0.001
tpi_10c	0.482828	p < 0.001	652391381.1	p < 0.001
tpi_cls10c	0.417984	p < 0.001	642319759.3	p < 0.001
tpi_sd10c	0.483029	p < 0.001	652326275.4	p < 0.001
twi32c	0.412928	p < 0.001	184966690.9	p < 0.001
random	0.013023	p=0.042	415502733.8	p=0.628

Region 9/10 - Riverine Section 14				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_trail_dist	0.247957	p < 0.001	1025875931	p < 0.001
e_hyd_min	0.434936	p < 0.001	2036217109	p < 0.001
ed_h2	0.39778	p < 0.001	1952236060	p < 0.001
ed_h6	0.232057	p < 0.001	1551283192	p < 0.001
eldrop32c	0.217848	p < 0.001	1700107592	p < 0.001
elev_2_drainh	0.241557	p < 0.001	1732496605	p < 0.001
rel_10c	0.354752	p < 0.001	1921358584	p < 0.001
rng_16c	0.245468	p < 0.001	945069694.4	p < 0.001
slpvr_10c	0.344313	p < 0.001	770440604	p < 0.001
std_16c	0.237488	p < 0.001	980177257.2	p < 0.001
tpi_10c	0.368722	p < 0.001	1986923816	p < 0.001
tpi_cls10c	0.351432	p < 0.001	1909848195	p < 0.001
tpi_sd10c	0.369216	p < 0.001	1987270085	p < 0.001
tri_10c	0.337459	p < 0.001	787224083.9	p < 0.001
random	0.005389	p=0.480	1316889877	p=0.608

Region 9/10 - Riverine Section 14				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_trail_dist	0.239786	p < 0.001	298136756.4	p < 0.001
cd_h7	0.401928	p < 0.001	189941678.2	p < 0.001
drcdry	0.289729	p < 0.001	333992997.3	p < 0.001
e_hyd_min	0.349254	p < 0.001	364572995.9	p < 0.001
ed_h2	0.341209	p < 0.001	355271698.9	p < 0.001
ed_h4	0.394564	p < 0.001	151006327.6	p < 0.001
ed_h5	0.335438	p < 0.001	196273518.6	p < 0.001
eldrop32c	0.23005	p < 0.001	320324823.7	p < 0.001
elev_2_strm	0.309953	p < 0.001	212123994.4	p < 0.001
niccdcd	0.306752	p < 0.001	177266027.5	p < 0.001
rel_8c	0.337521	p < 0.001	362232821.6	p < 0.001
rng_32c	0.225198	p < 0.001	203826485.9	p < 0.001
slpvr_32c	0.327935	p < 0.001	169387939.9	p < 0.001
tpi_10c	0.391451	p < 0.001	388732015	p < 0.001
tpi_cls10c	0.363484	p < 0.001	369713338.2	p < 0.001
tpi_sd10c	0.391762	p < 0.001	388806211.6	p < 0.001
tri_32c	0.328725	p < 0.001	169305299.1	p < 0.001
random	0.014023	p=0.097	255124550.1	p=0.140

Region 9/10 - Upland Section 1				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.341062	p < 0.001	552649506.7	p < 0.001
c_hyd_min_wt	0.255279	p < 0.001	986014881.8	p < 0.001
c_trail_dist	0.318869	p < 0.001	813076443.2	p < 0.001
cd_drnh	0.342269	p < 0.001	1059218421	p < 0.001
cd_h5	0.37525	p < 0.001	1100702523	p < 0.001
ed_h1	0.405852	p < 0.001	375667663.7	p < 0.001
ed_h6	0.346696	p < 0.001	648198090.5	p < 0.001
eldrop10c	0.221665	p < 0.001	962803988.9	p < 0.001
elev_2_drainh	0.352557	p < 0.001	424904787.5	p < 0.001
niccdcd	0.494861	p < 0.001	1108502111	p < 0.001
rng_32c	0.443957	p < 0.001	1159523203	p < 0.001
slope_pct	0.306588	p < 0.001	1040437319	p < 0.001
slpvr_16c	0.400009	p < 0.001	1083638585	p < 0.001
std_32c	0.399037	p < 0.001	1142432958	p < 0.001
tri_16c	0.404158	p < 0.001	1087982428	p < 0.001
vrf_32c	0.273098	p < 0.001	981723240	p < 0.001
random	0.00766	p = 0.291	755127112.2	p = 0.447

Region 9/10 - Upland Section 2				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
cd_conf	0.401275	p < 0.001	514712188.4	p < 0.001
cd_h4	0.250035	p < 0.001	691012878.7	p < 0.001
cd_h5	0.240656	p < 0.001	712431419.9	p < 0.001
cd_h6	0.726828	p < 0.001	192852091.2	p < 0.001
e_hyd_min	0.268975	p < 0.001	684940887.1	p < 0.001
e_trail_dist	0.325039	p < 0.001	1309021211	p < 0.001
ed_drnh	0.462499	p < 0.001	1520526812	p < 0.001
ed_h2	0.253367	p < 0.001	700476048	p < 0.001
elev_2_conf	0.347419	p < 0.001	595966301.5	p < 0.001
elev_2_drainh	0.359438	p < 0.001	554901083.5	p < 0.001
elev_2_strm	0.601701	p < 0.001	269339071.3	p < 0.001
slpvr_32c	0.282393	p < 0.001	1310035464	p < 0.001
tpi_250c	0.518766	p < 0.001	421062662.8	p < 0.001
tpi_cls250c	0.300457	p < 0.001	581984643.6	p < 0.001
tpi_sd250c	0.51942	p < 0.001	419980128.4	p < 0.001
tri_32c	0.28323	p < 0.001	1304026500	p < 0.001
random	0.009212	p = 0.100	980897705.2	p = 0.395

Region 9/10 - Upland Section 3				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.446652	p < 0.001	545543797.9	p < 0.001
c_trail_dist	0.583184	p < 0.001	196083340.1	p < 0.001
cd_conf	0.585149	p < 0.001	130800280.2	p < 0.001
cd_h2	0.431816	p < 0.001	188371984.8	p < 0.001
cd_h4	0.566211	p < 0.001	121242189.5	p < 0.001
cd_h5	0.443401	p < 0.001	183229915.2	p < 0.001
cd_h7	0.592025	p < 0.001	104558565.2	p < 0.001
e_hyd_min_wt	0.516608	p < 0.001	136606369.6	p < 0.001
ed_drnh	0.458698	p < 0.001	595655754.7	p < 0.001
elev_2_conf	0.484519	p < 0.001	169594558.3	p < 0.001
elev_2_drainh	0.551136	p < 0.001	137602138.2	p < 0.001
elev_2_strm	0.619531	p < 0.001	107644626.5	p < 0.001
rel_32c	0.495005	p < 0.001	190964287.3	p < 0.001
tpi_100c	0.597687	p < 0.001	110293434.1	p < 0.001
tpi_cls100c	0.592527	p < 0.001	149851642.1	p < 0.001
tpi_sd100c	0.597496	p < 0.001	110421736.3	p < 0.001
random	0.013303	p = 0.044	387982172	p = 0.155

Region 9/10 - Upland Section 4				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.290188	p < 0.001	310092477.3	p < 0.001
c_trail_dist	0.346525	p < 0.001	161870876	p < 0.001
cd_conf	0.373442	p < 0.001	128820575.8	p < 0.001
cd_drnh	0.366507	p < 0.001	143796094.1	p < 0.001
cd_h2	0.455731	p < 0.001	107128234.8	p < 0.001
cd_h4	0.337547	p < 0.001	134457059.9	p < 0.001
cd_h5	0.416801	p < 0.001	120497322.7	p < 0.001
cd_h7	0.313968	p < 0.001	158896894.8	p < 0.001
e_hyd_min	0.47242	p < 0.001	107706715.7	p < 0.001
eldrop32c	0.333291	p < 0.001	139131248.2	p < 0.001
elev_2_conf	0.415427	p < 0.001	124657444.9	p < 0.001
elev_2_drainh	0.357122	p < 0.001	142368303.1	p < 0.001
elev_2_strm	0.346043	p < 0.001	144967295	p < 0.001
niccdcd	0.348253	p < 0.001	171015828.8	p < 0.001
rng_32c	0.443941	p < 0.001	113795574.9	p < 0.001
slope_deg	0.256674	p < 0.001	173970438.1	p < 0.001
slpvr_32c	0.280726	p < 0.001	179693070.4	p < 0.001
std_32c	0.443911	p < 0.001	123776600.8	p < 0.001
tpi_250c	0.410934	p < 0.001	141557624.7	p < 0.001
tpi_cls250c	0.346693	p < 0.001	152257249.1	p < 0.001
tpi_sd250c	0.410624	p < 0.001	141645933.7	p < 0.001
tri_32c	0.290286	p < 0.001	176617019.2	p < 0.001
random	0.010931	p = 0.325	245590577.4	p = 0.212

Region 9/10 - Upland Section 5				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.339517	p < 0.001	2095331309	p < 0.001
c_hyd_min_wt	0.480761	p < 0.001	734094393.4	p < 0.001
c_trail_dist	0.521479	p < 0.001	735213353.4	p < 0.001
cd_conf	0.63352	p < 0.001	428654734	p < 0.001
cd_drnh	0.472979	p < 0.001	724067069.5	p < 0.001
cd_h2	0.4552	p < 0.001	812237748.2	p < 0.001
cd_h4	0.534539	p < 0.001	537510623	p < 0.001
cd_h5	0.537955	p < 0.001	622624536	p < 0.001
ed_h6	0.714194	p < 0.001	499565432.3	p < 0.001
eldrop32c	0.391645	p < 0.001	955825591.8	p < 0.001
elev_2_conf	0.551277	p < 0.001	653397917.7	p < 0.001
elev_2_strm	0.64078	p < 0.001	368413002.4	p < 0.001
rng_32c	0.431485	p < 0.001	797259805.5	p < 0.001
slope_deg	0.392221	p < 0.001	908055410.9	p < 0.001
std_32c	0.415329	p < 0.001	862361296.3	p < 0.001
tpi_250c	0.599692	p < 0.001	545290647	p < 0.001
tpi_cls100c	0.520785	p < 0.001	740784817.1	p < 0.001
tpi_sd250c	0.599288	p < 0.001	545721138.2	p < 0.001
random	0.00634	p = 0.286	1822082846	p = 0.403

Region 9/10 - Upland Section 6				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_trail_dist	0.216372	p < 0.001	2080212575	p < 0.001
cd_conf	0.414782	p < 0.001	792571040.8	p < 0.001
cd_h4	0.230532	p < 0.001	1250081137	p < 0.001
cd_h5	0.204336	p < 0.001	1336258909	p < 0.001
cd_h7	0.195428	p < 0.001	1505265720	p < 0.001
e_hyd_min	0.375506	p < 0.001	915298935.4	p < 0.001
ed_drnh	0.363884	p < 0.001	2424734786	p < 0.001
ed_h2	0.403784	p < 0.001	867823560.3	p < 0.001
elev_2_conf	0.396523	p < 0.001	866364017	p < 0.001
elev_2_drainh	0.296913	p < 0.001	1060733966	p < 0.001
elev_2_strm	0.383532	p < 0.001	904198509.3	p < 0.001
rel_32c	0.329278	p < 0.001	1003258979	p < 0.001
slpvr_16c	0.272157	p < 0.001	2276546857	p < 0.001
tpi_250c	0.506166	p < 0.001	676482195	p < 0.001
tpi_cls250c	0.503076	p < 0.001	813757742.1	p < 0.001
tpi_sd250c	0.506129	p < 0.001	676685339.7	p < 0.001
tri_16c	0.266274	p < 0.001	2263202133	p < 0.001
random	0.009347	p = 0.059	1700178427	p = 0.274

Region 9/10 - Upland Section 7				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.258952	p < 0.001	672351569.5	p < 0.001
c_trail_dist	0.301279	p < 0.001	423899749.6	p < 0.001
drcdry	0.269316	p < 0.001	718115536.3	p < 0.001
e_hyd_min	0.264264	p < 0.001	364362339.7	p < 0.001
ed_conf	0.297825	p < 0.001	329375922.9	p < 0.001
ed_h2	0.257249	p < 0.001	364754512.1	p < 0.001
ed_h4	0.267292	p < 0.001	347294270.5	p < 0.001
ed_h5	0.302649	p < 0.001	330194267.4	p < 0.001
ed_h7	0.617858	p < 0.001	180966546.1	p < 0.001
elev_2_drainh	0.34851	p < 0.001	334419152.1	p < 0.001
elev_2_strm	0.536734	p < 0.001	198809097.4	p < 0.001
niccdcd	0.270046	p < 0.001	375912619.7	p < 0.001
rng_32c	0.244106	p < 0.001	683600421.3	p < 0.001
slpvr_32c	0.469401	p < 0.001	871980980.4	p < 0.001
std_32c	0.260781	p < 0.001	701183501.2	p < 0.001
tpi_250c	0.564194	p < 0.001	179719381.9	p < 0.001
tpi_cls250c	0.564426	p < 0.001	231475913.6	p < 0.001
tpi_sd250c	0.564114	p < 0.001	179709995	p < 0.001
tri_32c	0.46966	p < 0.001	872035050.9	p < 0.001
random	0.008288	p = 0.303	548076004.2	p = 0.263

Region 9/10 - Upland Section 8				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_trail_dist	0.320805	p < 0.001	226976614.8	p < 0.001
cd_conf	0.448164	p < 0.001	172611643.9	p < 0.001
cd_h4	0.339756	p < 0.001	264197121.3	p < 0.001
cd_h5	0.278809	p < 0.001	266121261.2	p < 0.001
cd_h7	0.474245	p < 0.001	182369401.7	p < 0.001
e_hyd_min_wt	0.407229	p < 0.001	186946811.5	p < 0.001
ed_drnh	0.304044	p < 0.001	511860461.3	p < 0.001
ed_h2	0.441408	p < 0.001	171312250.8	p < 0.001
eldrop32c	0.273722	p < 0.001	259525942.6	p < 0.001
elev_2_conf	0.406893	p < 0.001	184163527	p < 0.001
elev_2_drainh	0.344261	p < 0.001	212082744.5	p < 0.001
elev_2_strm	0.463039	p < 0.001	177901435.1	p < 0.001
rel_32c	0.29418	p < 0.001	237937189	p < 0.001
rng_16c	0.24622	p < 0.001	318937062.7	p < 0.001
std_32c	0.251609	p < 0.001	317673466.4	p < 0.001
tpi_250c	0.459008	p < 0.001	191543595.7	p < 0.001
tpi_cls250c	0.297221	p < 0.001	234028977.1	p < 0.001
tpi_sd250c	0.458816	p < 0.001	191628074.9	p < 0.001
random	0.009149	p = 0.314	382667672.7	p = 0.720

Region 9/10 - Upland Section 9				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.73025	p < 0.001	20961292.78	p < 0.001
c_trail_dist	0.718951	p < 0.001	20501779.83	p < 0.001
cd_h6	0.687312	p < 0.001	19982315.2	p < 0.001
drcdry	0.557991	p < 0.001	17649777.56	p < 0.001
e_hyd_min	0.716082	p < 0.001	3134040.85	p < 0.001
ed_conf	0.733643	p < 0.001	2328500.415	p < 0.001
ed_drnh	0.385611	p < 0.001	11295397.34	p < 0.010
ed_h2	0.766802	p < 0.001	2649988.025	p < 0.001
ed_h4	0.4548	p < 0.001	10552051.88	p < 0.001
ed_h5	0.693129	p < 0.001	18649808.39	p < 0.001
elev_2_conf	0.515084	p < 0.001	6857702.49	p < 0.001
elev_2_drainh	0.721458	p < 0.001	2872926.165	p < 0.001
elev_2_strm	0.493201	p < 0.001	16233393.58	p < 0.001
niccdcd	0.726273	p < 0.001	3465243.93	p < 0.001
rng_16c	0.390821	p < 0.001	15673118.36	p < 0.001
slpvr_16c	0.513052	p < 0.001	15728811.8	p < 0.001
std_8c	0.411541	p < 0.001	16288772.21	p < 0.001
tpi_50c	0.555029	p < 0.001	4946562.9	p < 0.001
tpi_sd50c	0.555362	p < 0.001	4941133.135	p < 0.001
tri_16c	0.51116	p < 0.001	15684980.88	p < 0.001
vrf_16c	0.425706	p < 0.001	18348341.51	p < 0.001
random	0.045009	p = 0.287	12572625.41	p = 0.217

Region 9/10 - Upland Section 10				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.38779	p < 0.001	662459562.4	p < 0.001
cd_conf	0.492483	p < 0.001	201382512	p < 0.001
cd_h4	0.44662	p < 0.001	204455373.8	p < 0.001
cd_h5	0.279538	p < 0.001	326555011.9	p < 0.001
drcdry	0.298844	p < 0.001	548424193.8	p < 0.001
e_hyd_min_wt	0.45701	p < 0.001	216117695.5	p < 0.001
ed_h2	0.491906	p < 0.001	190465008.5	p < 0.001
ed_h6	0.468741	p < 0.001	679153752.2	p < 0.001
elev_2_conf	0.396967	p < 0.001	250511471.9	p < 0.001
elev_2_drainh	0.263885	p < 0.001	348303757.8	p < 0.001
elev_2_strm	0.466822	p < 0.001	255088022.2	p < 0.001
niccdcd	0.313643	p < 0.001	338478544.3	p < 0.001
rng_32c	0.297665	p < 0.001	317292167.2	p < 0.001
std_32c	0.293795	p < 0.001	316587527.3	p < 0.001
tpi_100c	0.293349	p < 0.001	359325070.1	p < 0.001
tpi_sd100c	0.292874	p < 0.001	359633885	p < 0.001
random	0.010575	p=0.138	468340244.4	p=0.266

Region 9/10 - Upland Section 11				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.151386	p < 0.001	2345465693	p < 0.001
e_hyd_min	0.372779	p < 0.001	1189279660	p < 0.001
e_trail_dist	0.251253	p < 0.001	1827674373	p < 0.001
ed_conf	0.192211	p < 0.001	1760576988	p < 0.001
ed_drnh	0.156542	p < 0.001	2433082473	p < 0.001
ed_h2	0.363354	p < 0.001	1198301950	p < 0.001
ed_h5	0.268744	p < 0.001	1375227582	p < 0.001
ed_h6	0.329381	p < 0.001	1546627188	p < 0.001
elev_2_conf	0.21238	p < 0.001	1622568932	p < 0.001
elev_2_drainh	0.275331	p < 0.001	1376517153	p < 0.001
rel_32c	0.1706	p < 0.001	1727624045	p < 0.001
rng_32c	0.194363	p < 0.001	2713214773	p < 0.001
slpvr_32c	0.266147	p < 0.001	2843881813	p < 0.001
std_32c	0.207893	p < 0.001	2741689935	p < 0.001
tpi_250c	0.341852	p < 0.001	1246219845	p < 0.001
tpi_cls250c	0.334743	p < 0.001	1371653029	p < 0.001
tpi_sd250c	0.341385	p < 0.001	1247262773	p < 0.001
tri_32c	0.266315	p < 0.001	2843841119	p < 0.001
random	0.005041	p=0.443	2169474534	p=0.529

Region 9/10 - Upland Section 12				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.475446	p < 0.001	1321105725	p < 0.001
c_hyd_min_wt	0.519555	p < 0.001	327531379.9	p < 0.001
c_trail_dist	0.61936	p < 0.001	268560509	p < 0.001
cd_conf	0.662892	p < 0.001	190753328.9	p < 0.001
cd_drnh	0.565361	p < 0.001	299831149.7	p < 0.001
cd_h2	0.573907	p < 0.001	272010160.7	p < 0.001
cd_h4	0.689238	p < 0.001	215927965.3	p < 0.001
cd_h5	0.58703	p < 0.001	293679248.5	p < 0.001
drcdry	0.380895	p < 0.001	1137210916	p < 0.001
ed_h7	0.472395	p < 0.001	1272527032	p < 0.001
eldrop32c	0.333357	p < 0.001	490956299.9	p < 0.001
elev_2_conf	0.485906	p < 0.001	397961657.6	p < 0.001
elev_2_drainh	0.555186	p < 0.001	309942426.9	p < 0.001
elev_2_strm	0.395369	p < 0.001	618692643.4	p < 0.001
niccdcd	0.377951	p < 0.001	636266606.9	p < 0.001
rel_32c	0.341646	p < 0.001	556123824.8	p < 0.001
rng_32c	0.505322	p < 0.001	407163342.7	p < 0.001
slpvr_32c	0.302301	p < 0.001	754373636.3	p < 0.001
std_32c	0.450399	p < 0.001	465281682.2	p < 0.001
tpi_250c	0.609401	p < 0.001	292357790.4	p < 0.001
tpi_cls100c	0.399144	p < 0.001	571291799	p < 0.001
tpi_sd250c	0.60886	p < 0.001	292907482.3	p < 0.001
tri_32c	0.303653	p < 0.001	752461139	p < 0.001
random	0.01383	p=0.001	921990350	p=0.440

Region 9/10 - Upland Section 13				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.248285	p < 0.001	604501403.3	p < 0.001
c_trail_dist	0.536756	p < 0.001	201581541.2	p < 0.001
e_hyd_min	0.263588	p < 0.001	353440754.8	p < 0.001
ed_h1	0.438901	p < 0.001	297506950	p < 0.001
ed_h4	0.406207	p < 0.001	674243907.3	p < 0.001
ed_h5	0.246781	p < 0.001	377512491.5	p < 0.001
ed_h6	0.552392	p < 0.001	203512441.6	p < 0.001
eldrop32c	0.200604	p < 0.001	546145832.9	p < 0.001
elev_2_strm	0.287629	p < 0.001	329461611.2	p < 0.001
slpvr_16c	0.295876	p < 0.001	674531318.4	p < 0.001
std_32c	0.196602	p < 0.001	581492007.3	p < 0.001
tpi_10c	0.199182	p < 0.001	585169662.4	p < 0.001
tpi_cls10c	0.199245	p < 0.001	582985738.7	p < 0.001
tpi_sd10c	0.199148	p < 0.001	585092180.8	p < 0.001
tri_32c	0.29427	p < 0.001	674456697	p < 0.001
random	0.011853	p=0.066	497341899.7	p=0.397

Region 9/10 - Upland Section 14				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
c_hyd_min	0.384342	p < 0.001	1297338462	p < 0.001
c_trail_dist	0.259697	p < 0.001	1778370598	p < 0.001
cd_conf	0.386696	p < 0.001	1190783600	p < 0.001
cd_h2	0.3989	p < 0.001	1245926752	p < 0.001
cd_h4	0.331382	p < 0.001	1412614618	p < 0.001
cd_h5	0.32578	p < 0.001	1459397371	p < 0.001
ed_h6	0.258484	p < 0.001	2809205853	p < 0.001
eldrop32c	0.258802	p < 0.001	1617125962	p < 0.001
elev_2_conf	0.397654	p < 0.001	1196920090	p < 0.001
elev_2_drainh	0.246225	p < 0.001	1638884483	p < 0.001
rel_32c	0.280749	p < 0.001	1503022839	p < 0.001
rng_32c	0.204378	p < 0.001	1789538259	p < 0.001
std_32c	0.177287	p < 0.001	1875284074	p < 0.001
tpi_100c	0.345298	p < 0.001	1318004030	p < 0.001
tpi_cls100c	0.291916	p < 0.001	1495916316	p < 0.001
tpi_sd100c	0.345509	p < 0.001	1316737095	p < 0.001
random	0.004851	p=0.478	2445251547	p=0.603

Region 9/10 - Upland Section 15				
Predictor	**Mean D**	**Mean KS p**	**Mean U**	**Mean MW p**
aws050	0.218444	p < 0.001	454766403.8	p < 0.001
cd_conf	0.492339	p < 0.001	155452025.5	p < 0.001
cd_h4	0.56803	p < 0.001	119055118.3	p < 0.001
cd_h5	0.388979	p < 0.001	199849247.1	p < 0.001
cd_h7	0.441905	p < 0.001	230445404.8	p < 0.001
e_hyd_min_wt	0.402842	p < 0.001	224889474.4	p < 0.001
e_trail_dist	0.214765	p < 0.001	498974173.9	p < 0.001
ed_drnh	0.207394	p < 0.001	500815821.2	p < 0.001
ed_h2	0.388275	p < 0.001	234016800.6	p < 0.001
eldrop32c	0.252217	p < 0.001	271182386.3	p < 0.001
elev_2_conf	0.443574	p < 0.001	183479886.1	p < 0.001
elev_2_drainh	0.346771	p < 0.001	224879822.3	p < 0.001
elev_2_strm	0.433882	p < 0.001	218718755.5	p < 0.001
niccdcd	0.310223	p < 0.001	274678885.1	p < 0.001
rel_32c	0.392931	p < 0.001	201810859.3	p < 0.001
tpi_250c	0.567598	p < 0.001	150324875.6	p < 0.001
tpi_cls250c	0.538061	p < 0.001	172869614.8	p < 0.001
tpi_sd250c	0.567734	p < 0.001	150198312.5	p < 0.001
random	0.009346	p=0.272	413857610.7	p=0.302

APPENDIX E

VARIABLE IMPORTANCE

FOR SELECTED RF MODELS

WITHIN REGIONS 7, 8, AND 9/10

Chart 1. Region 7 All - Riverine Section 1

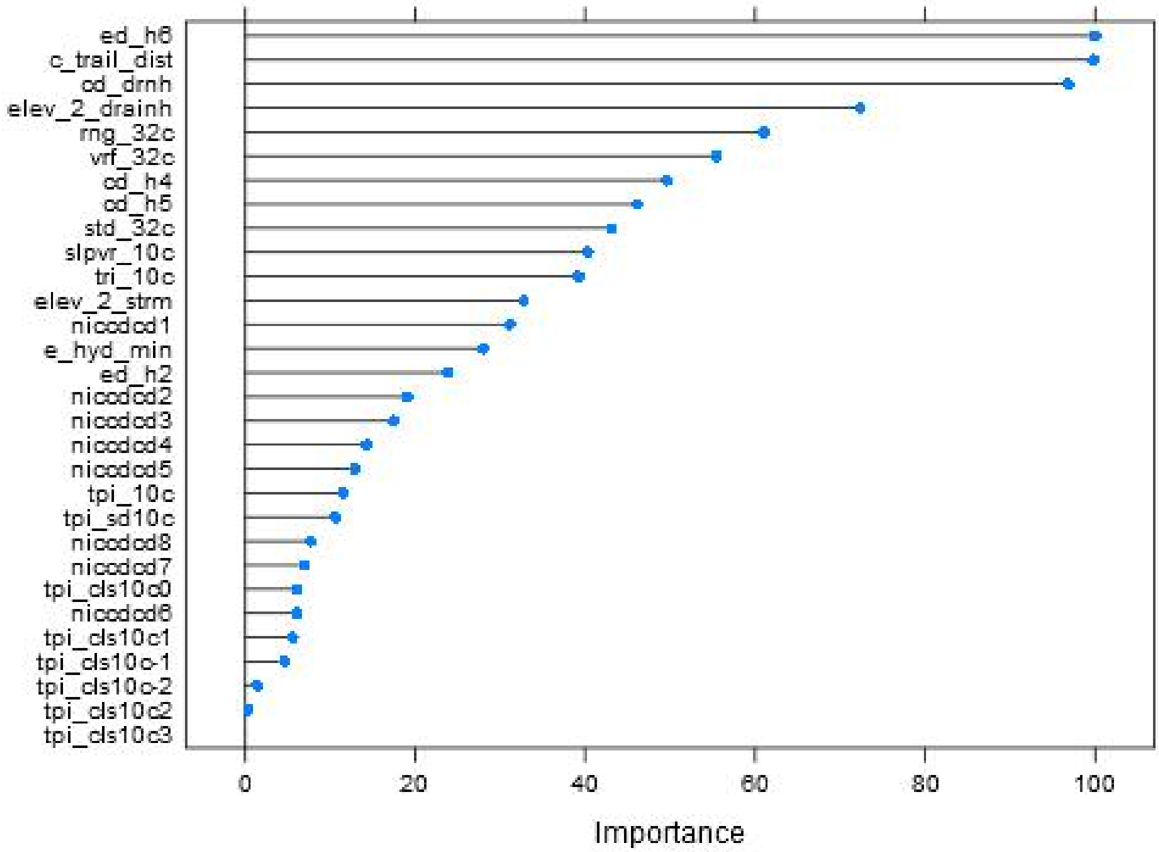


Chart 2. Region 7 All - Riverine Section 2

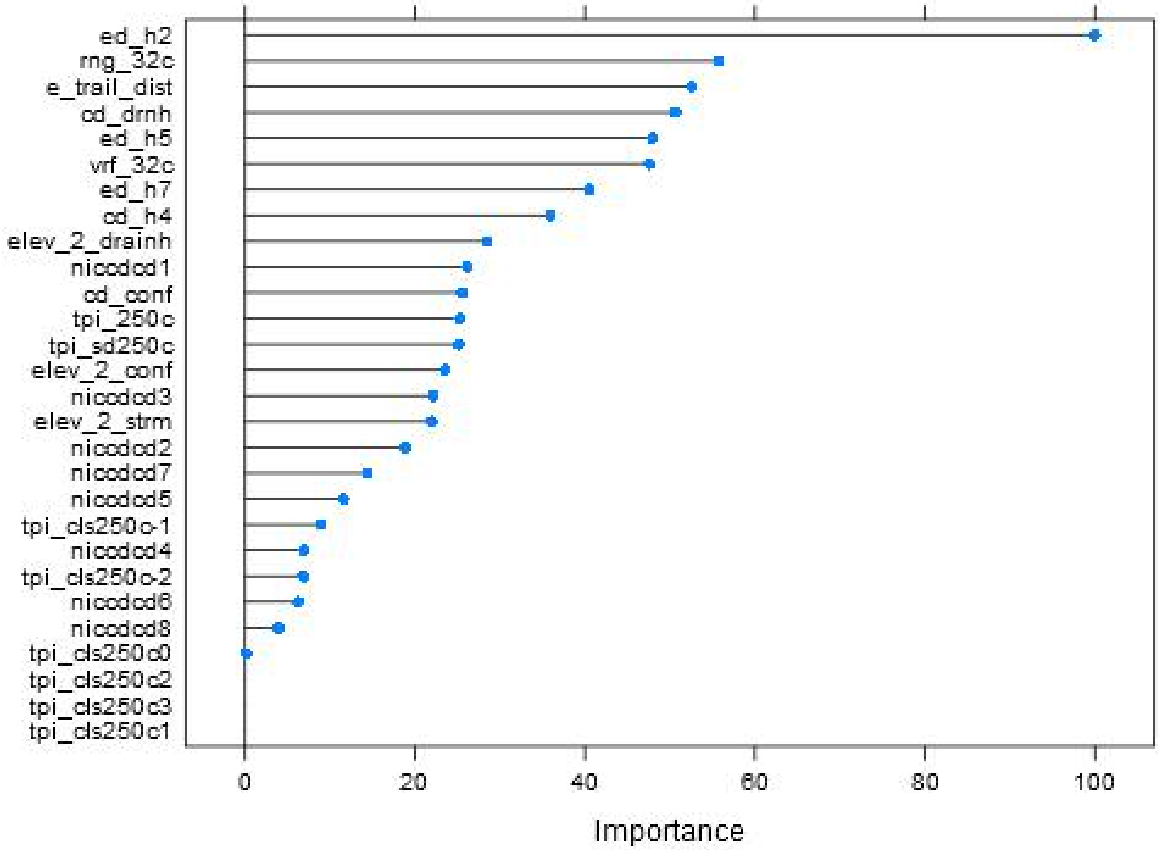


Chart 3. Region 7 All - Riverine Section

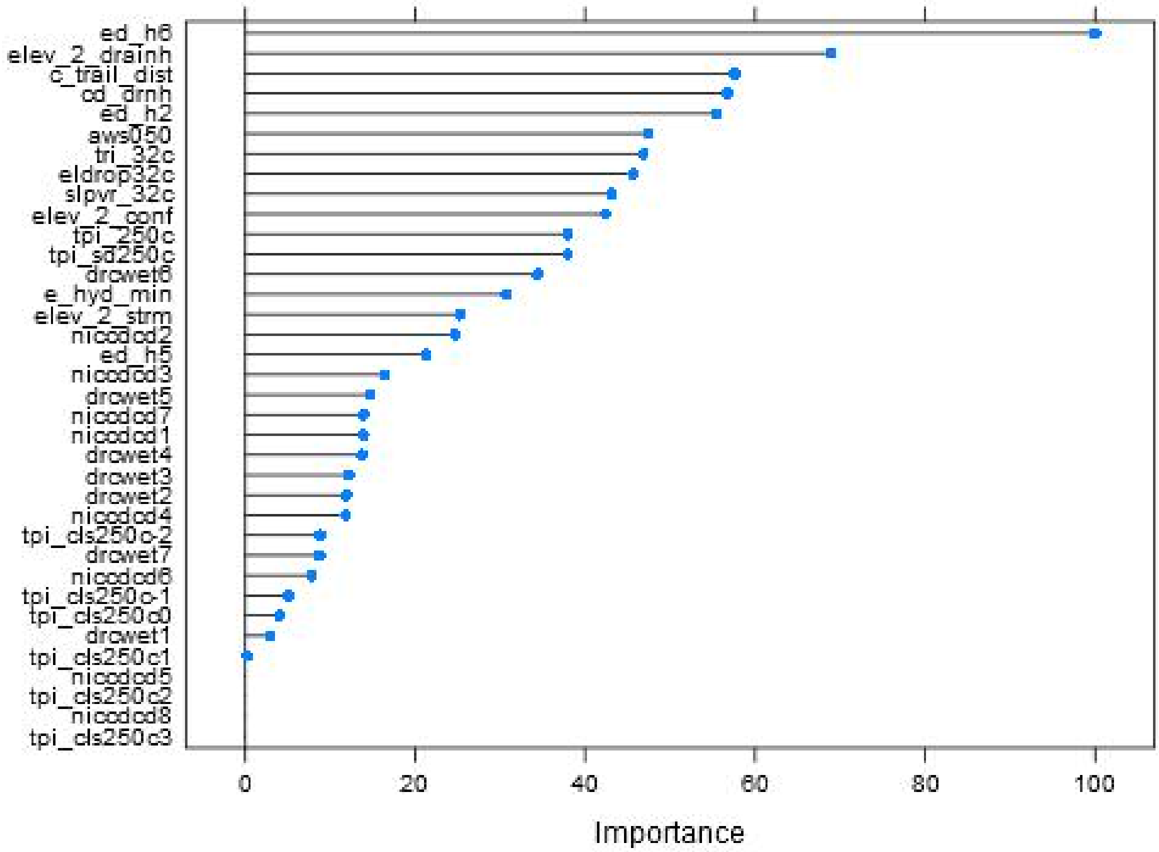


Chart 4. Region 7 All – Riverine Section 4

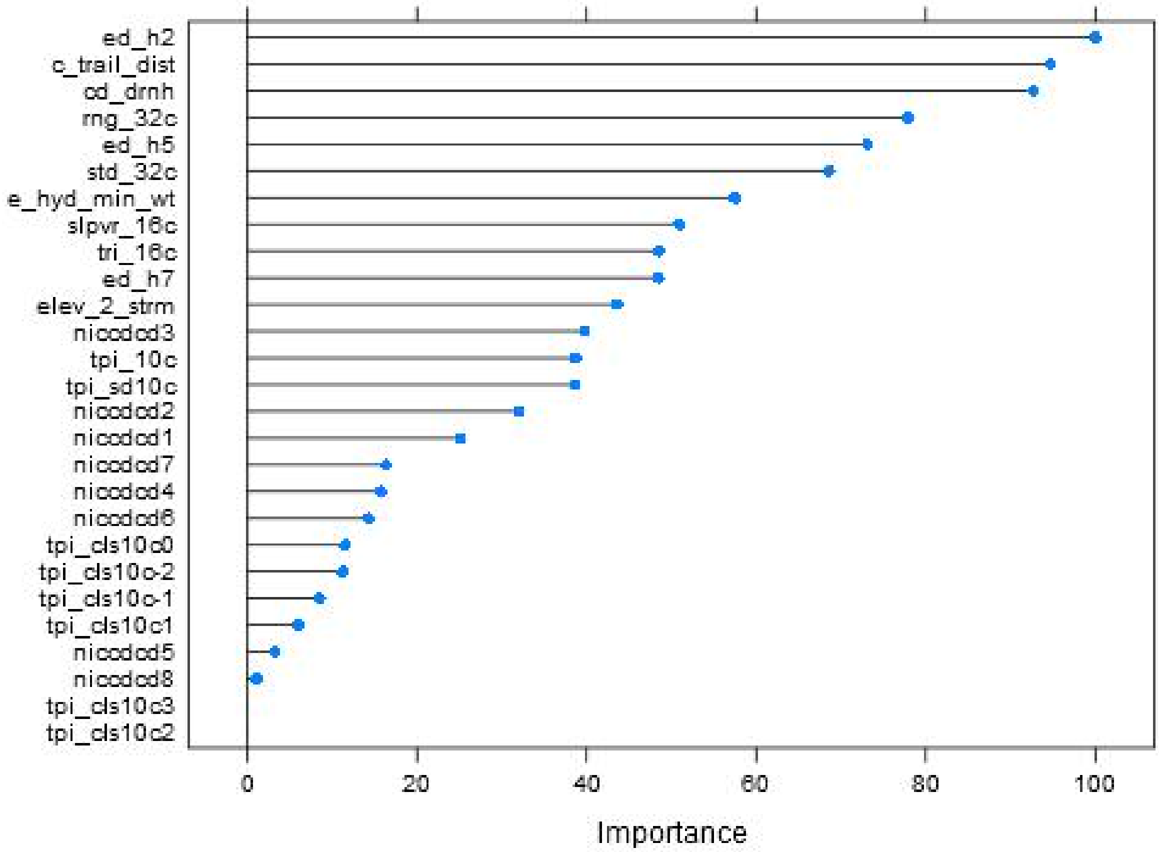


Chart 5. Region 7 All - Riverine Section 5

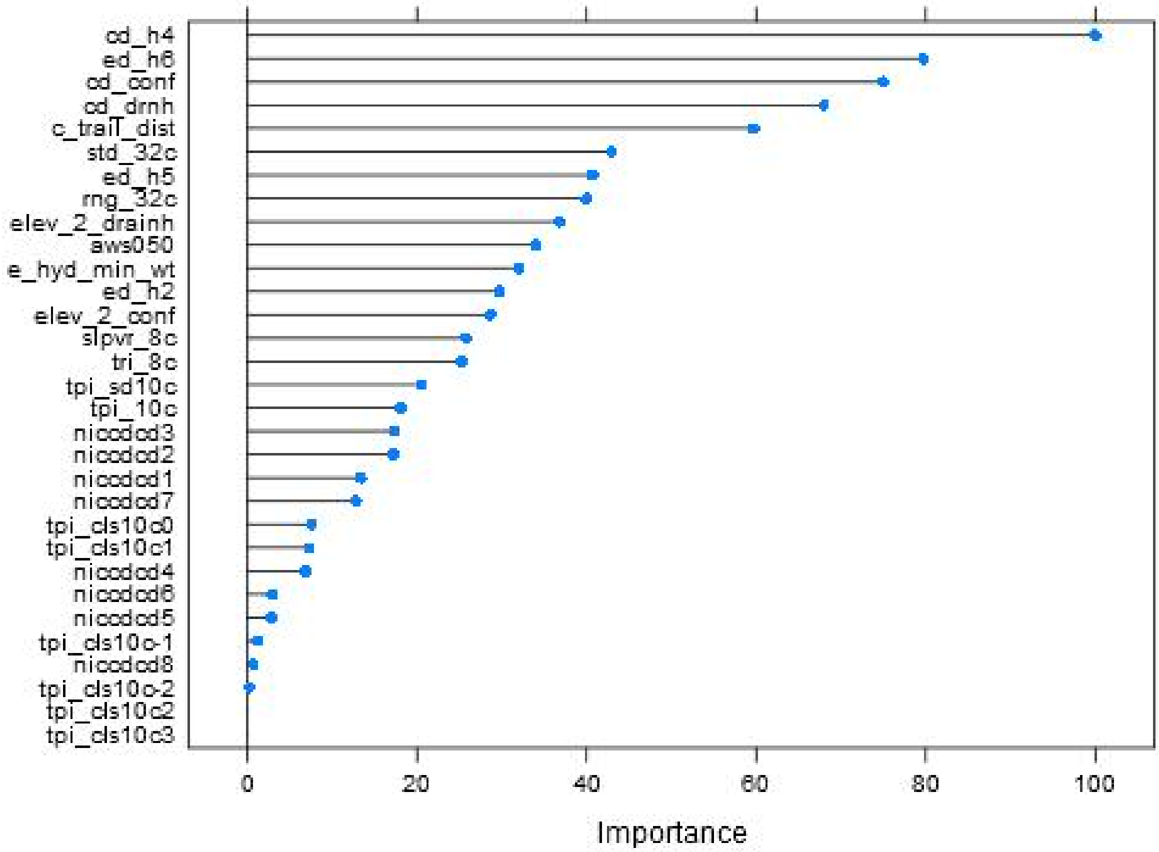


Chart 6. Region 7 All - Riverine Section 6

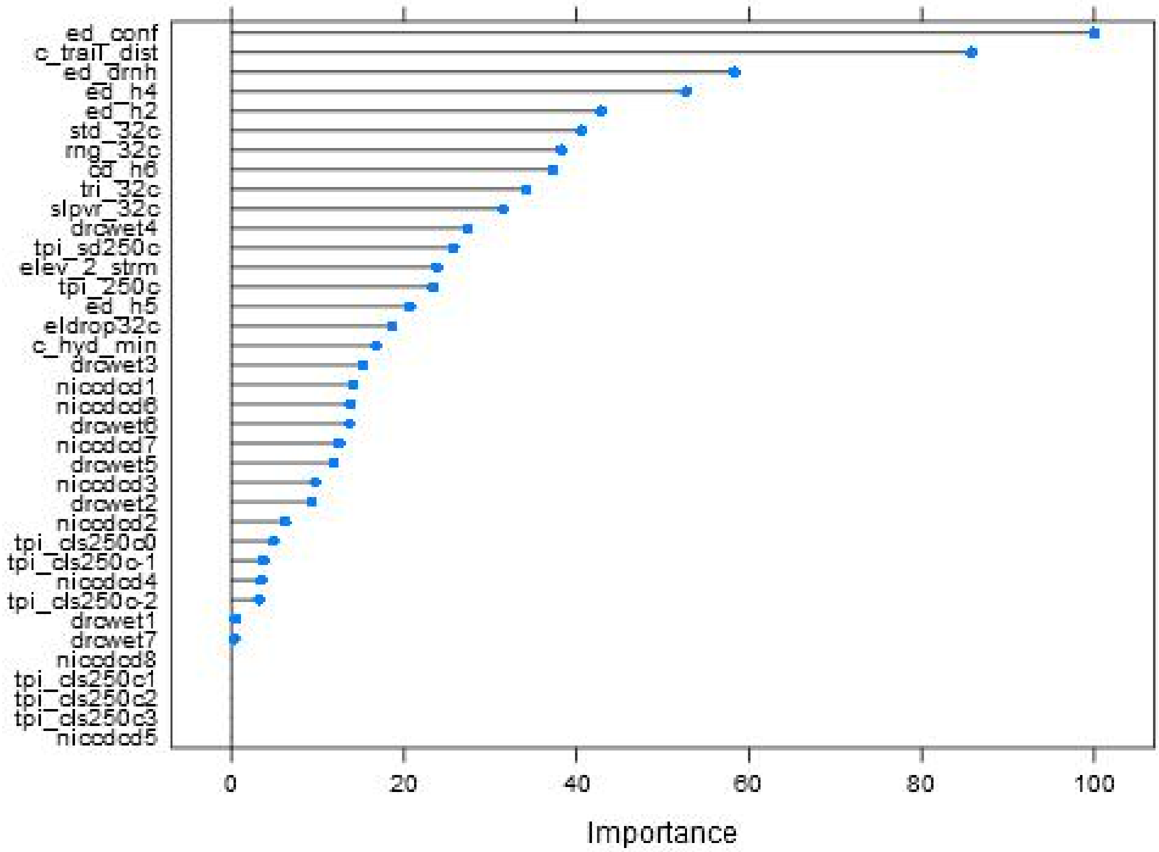


Chart 7. Region 7 All - Riverine Section 7

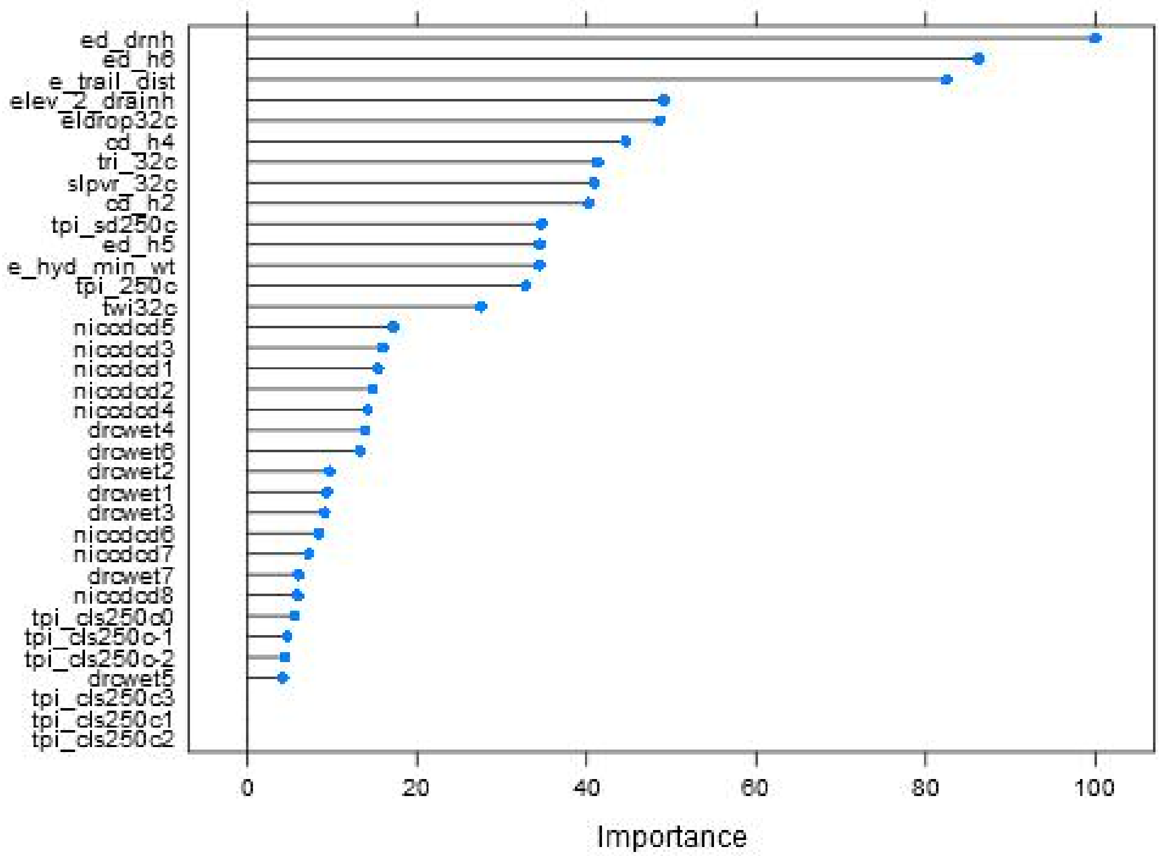


Chart 8. Region 7 All - Riverine Section 8

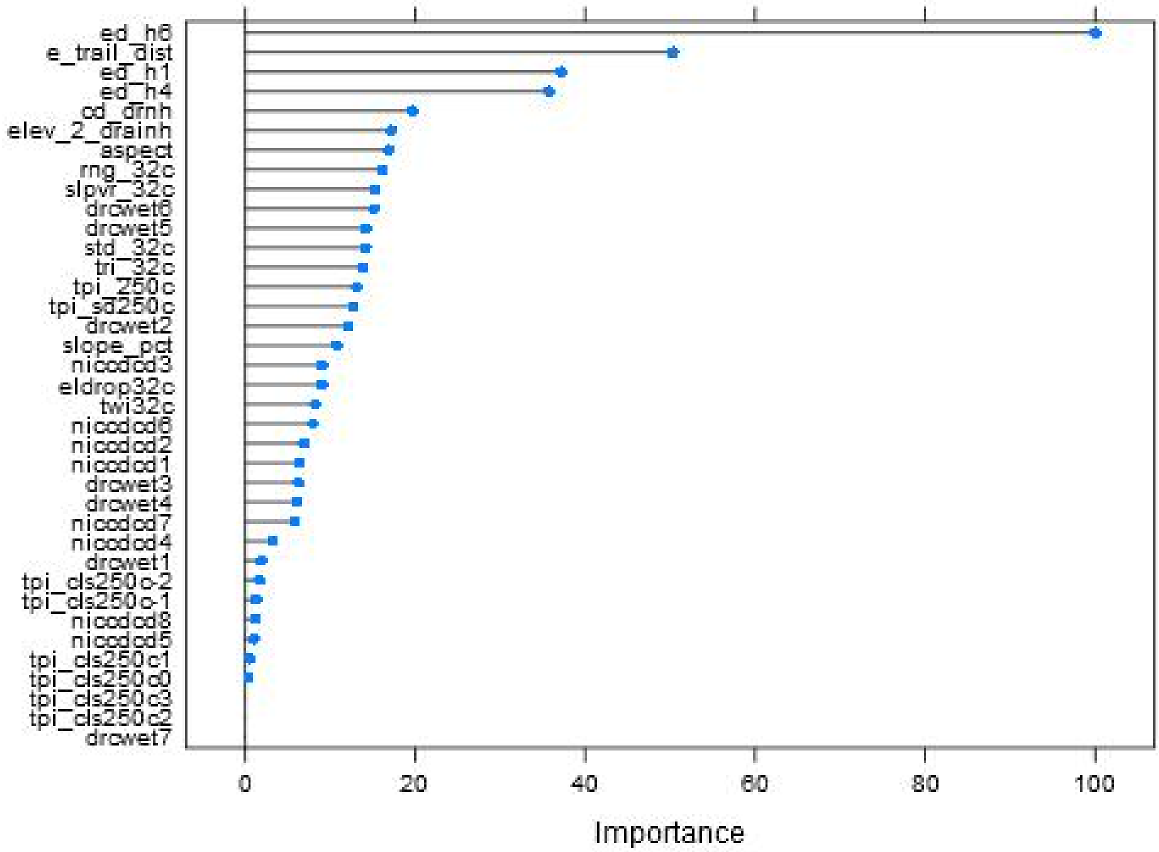


Chart 9. Region 7 All – Riverine Section 9

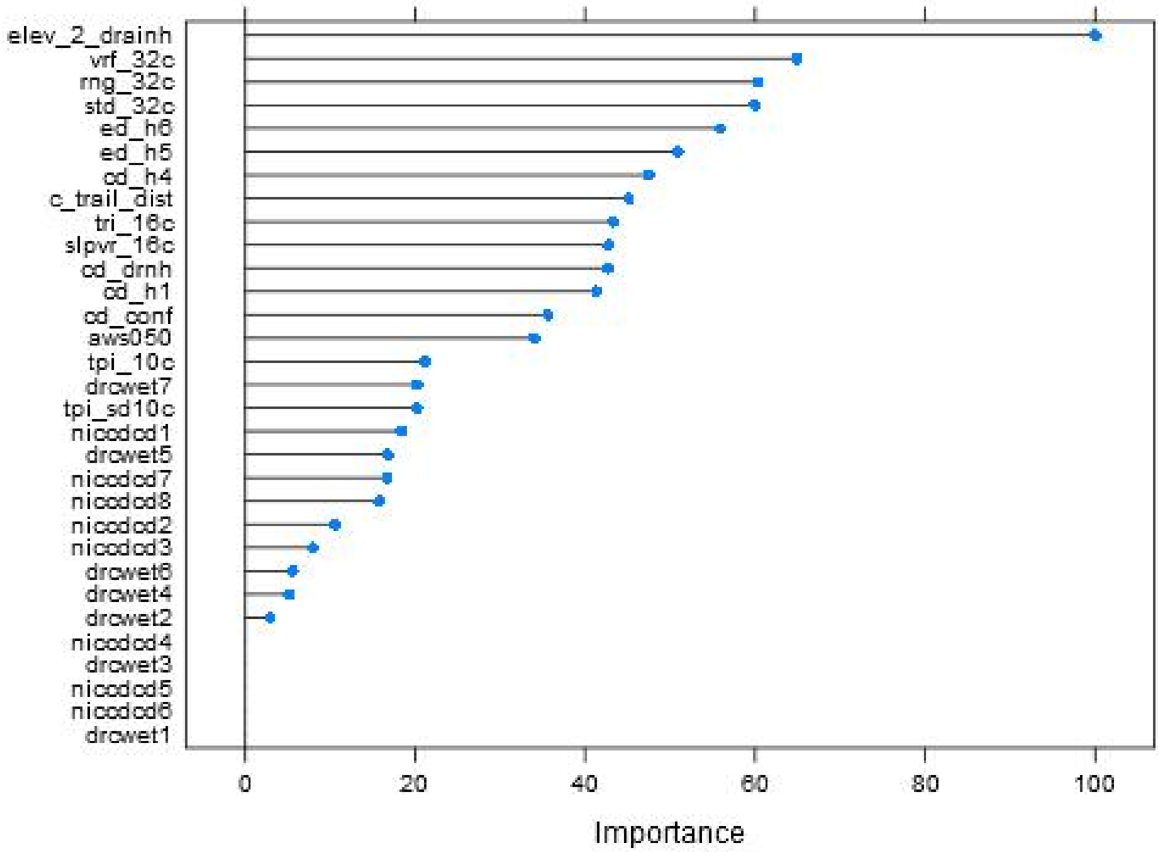


Chart 10. Region 7 All - Upland Section 1

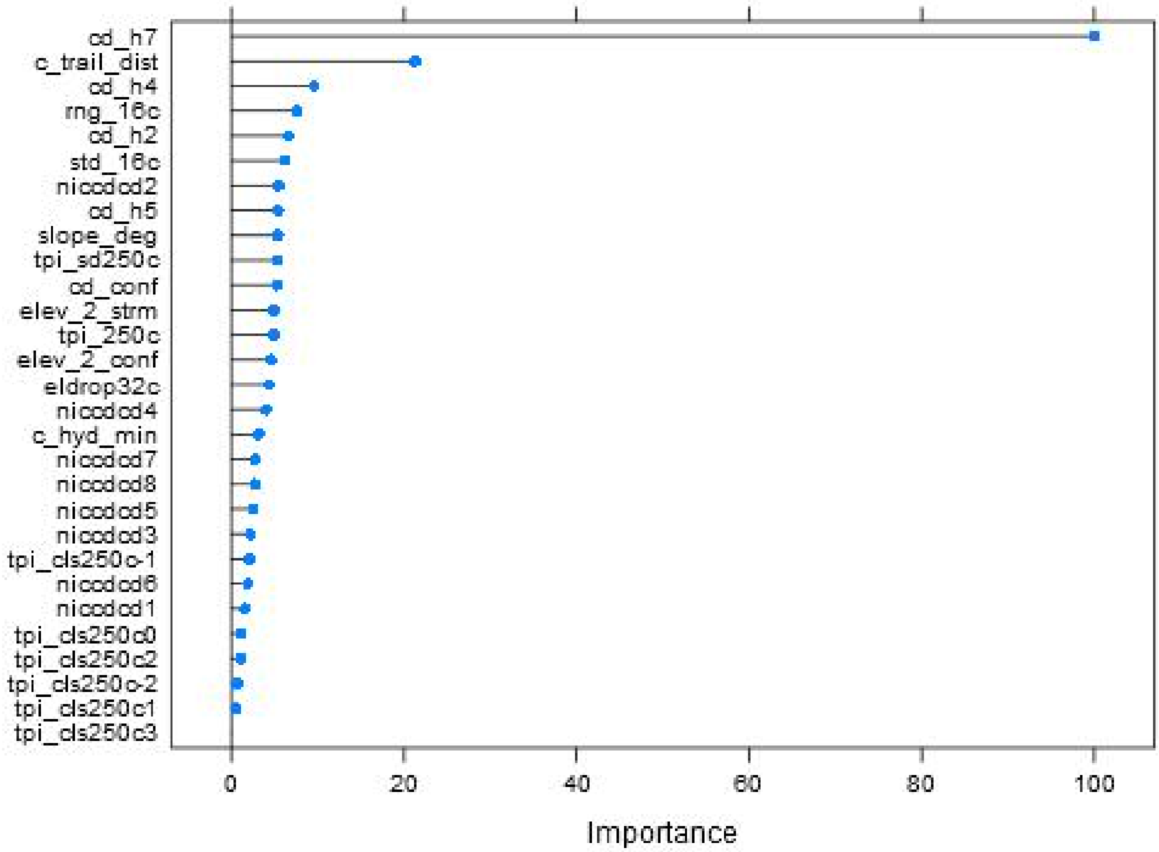


Chart 11. Region 7 All - Upland Section 2

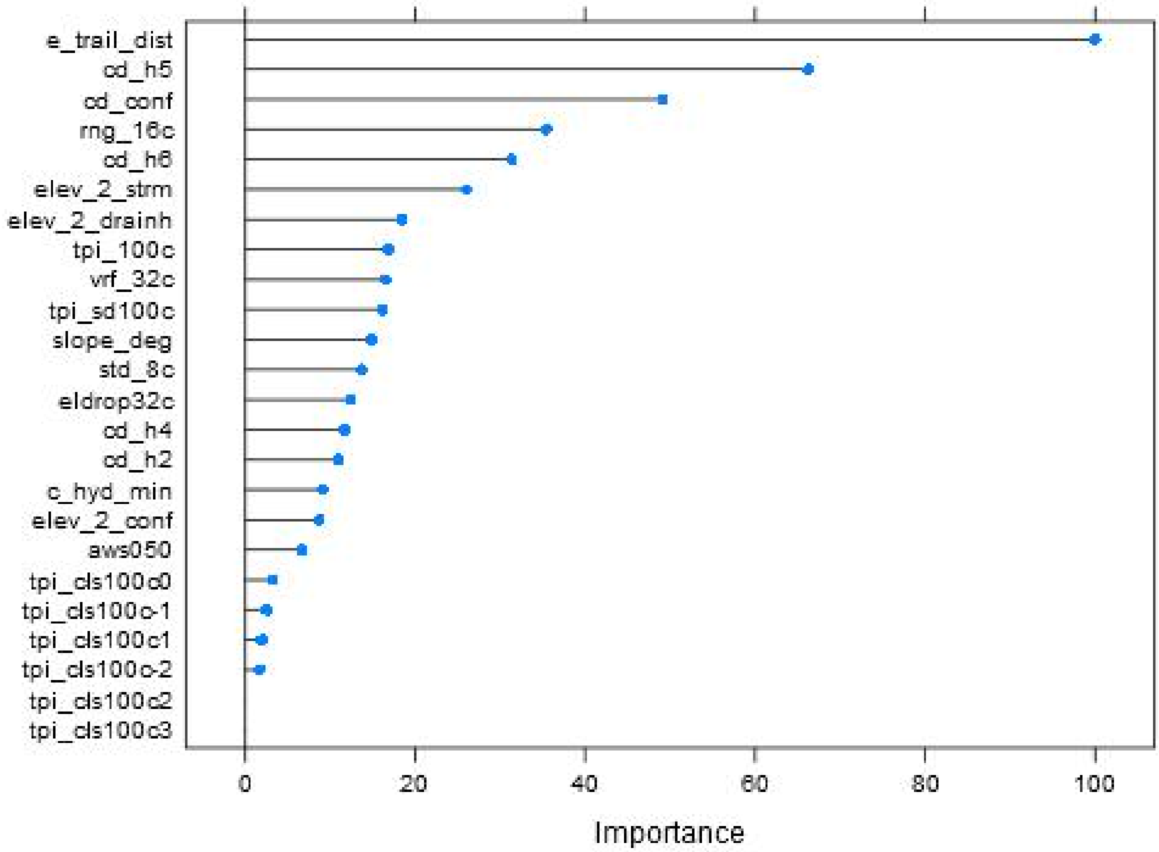


Chart 12. Region 7 All - Upland Section 3

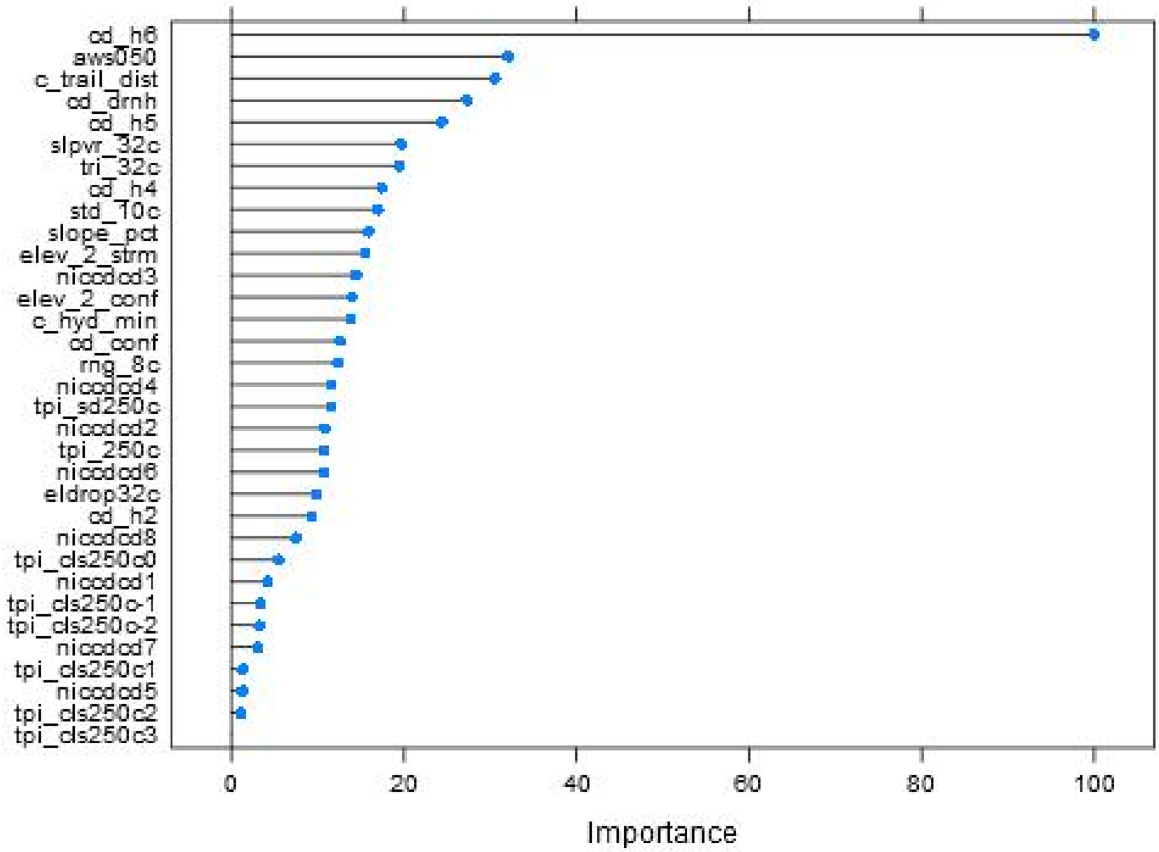


Chart 13. Region 7 All - Upland Section 4

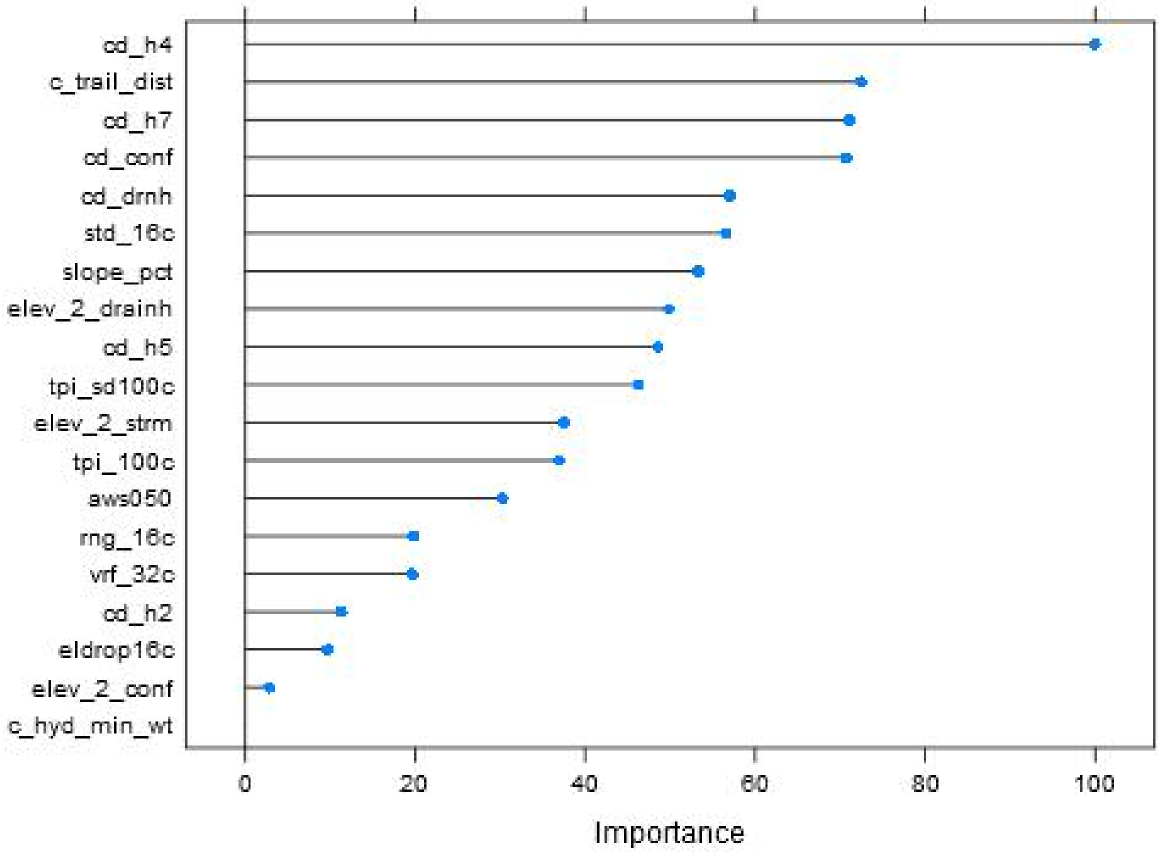


Chart 14. Region 7 All – Upland Section 5

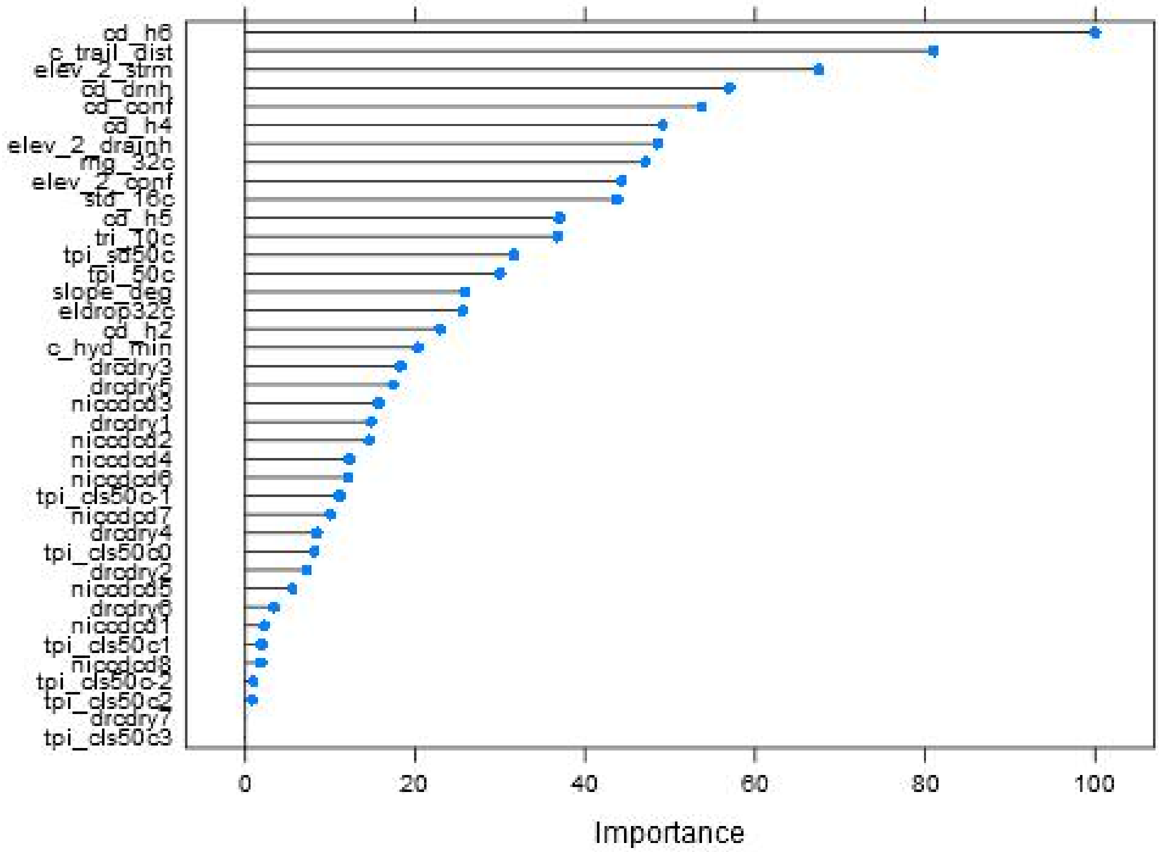


Chart 15. Region 7 All – Upland Section 6

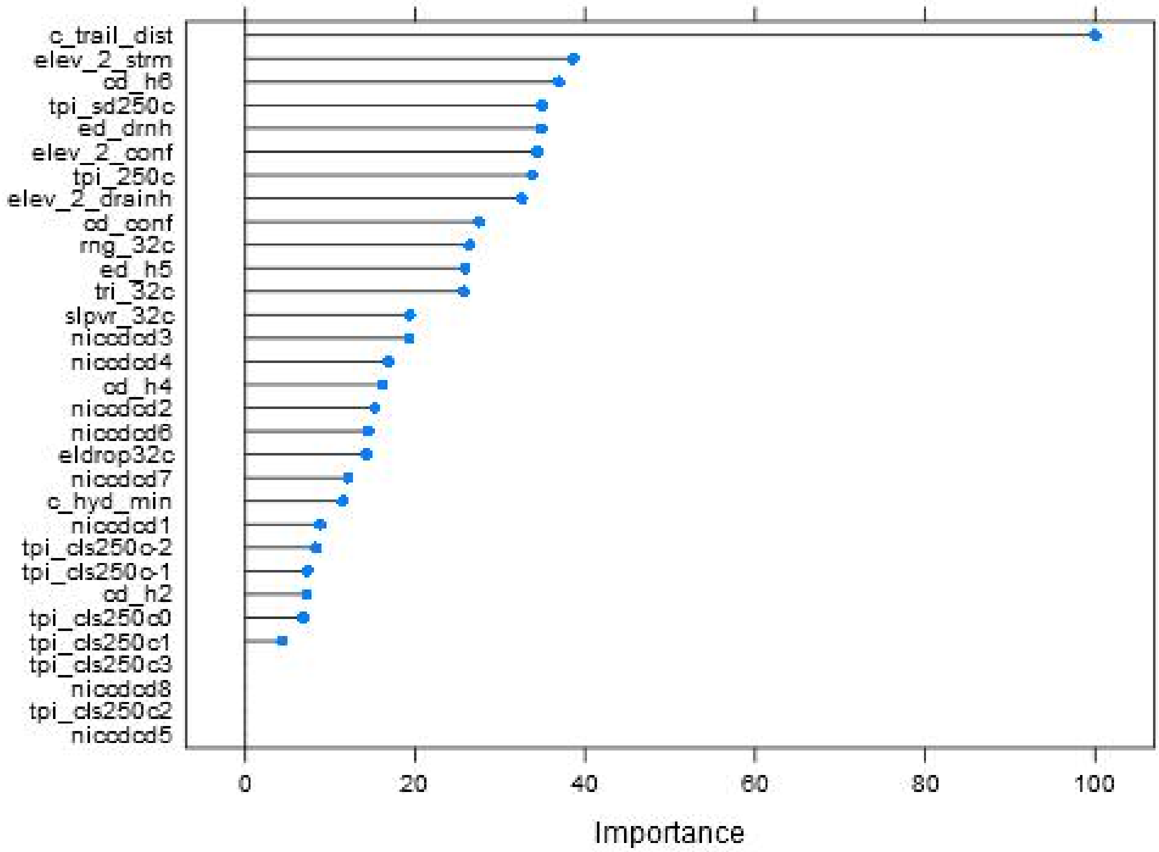


Chart 16. Region 7 All – Upland Section 6

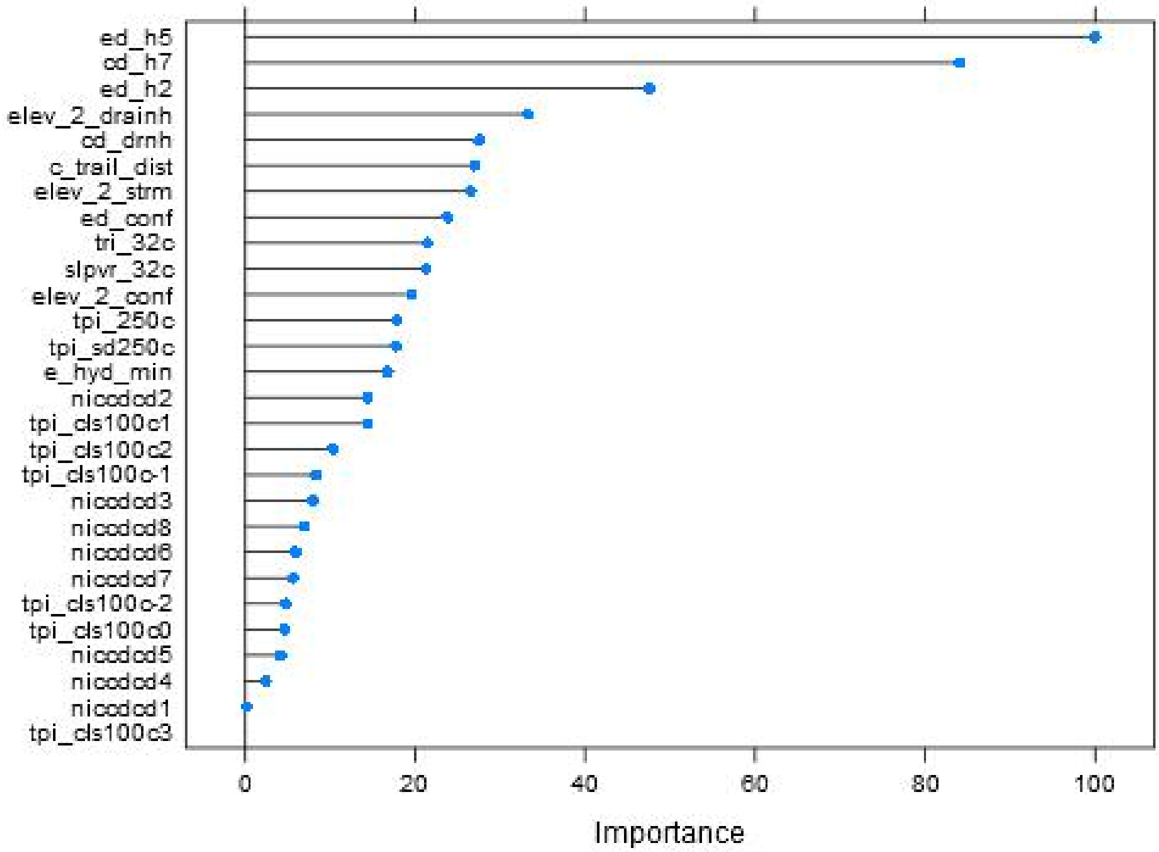


Chart 17. Region 7 All – Upland Section 8

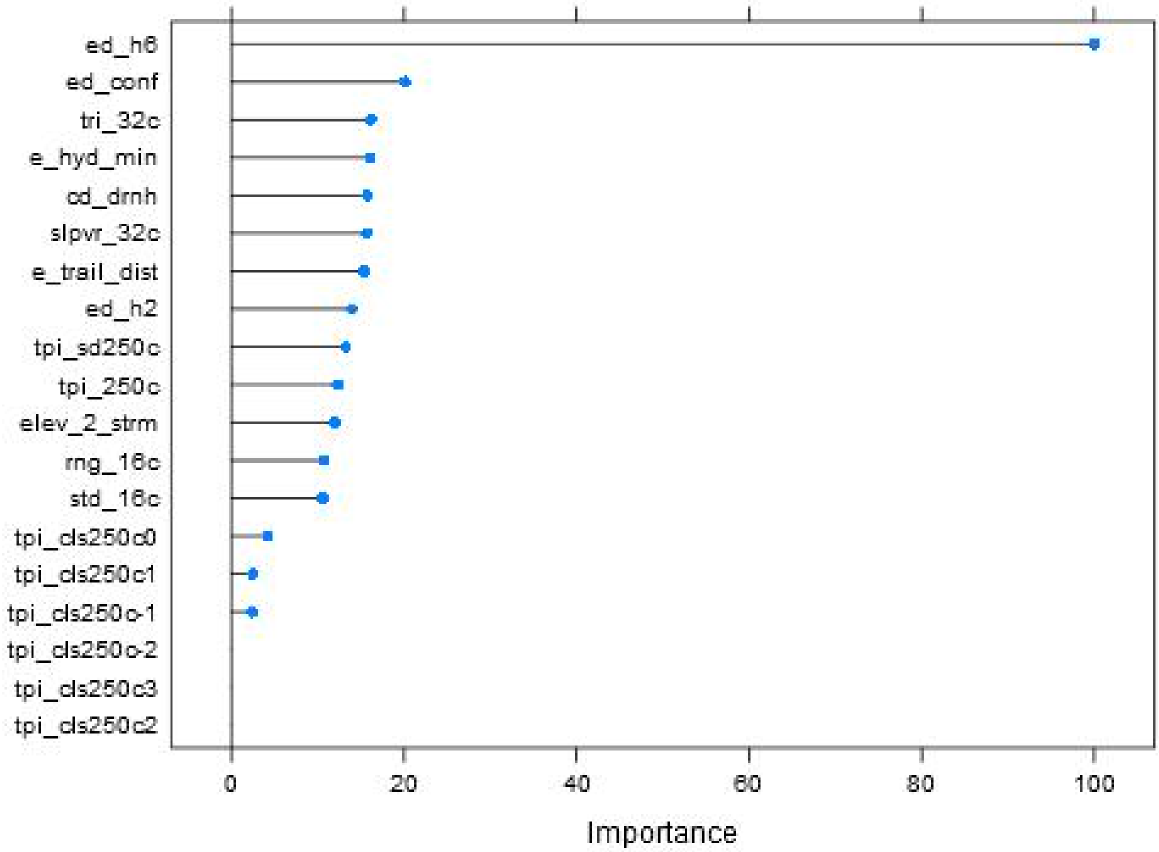


Chart 18. Region 7 All – Upland Section 9

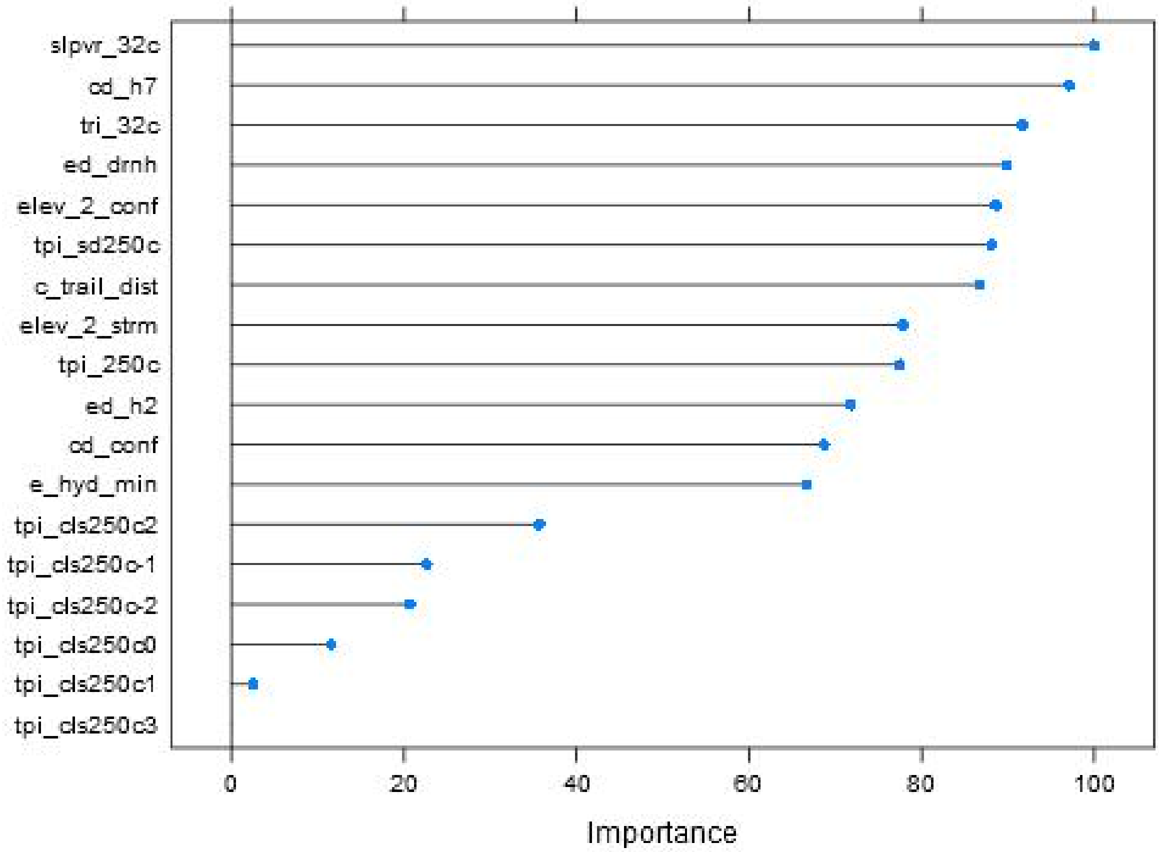


Chart 19. Region 8 All – Riverine Section 1

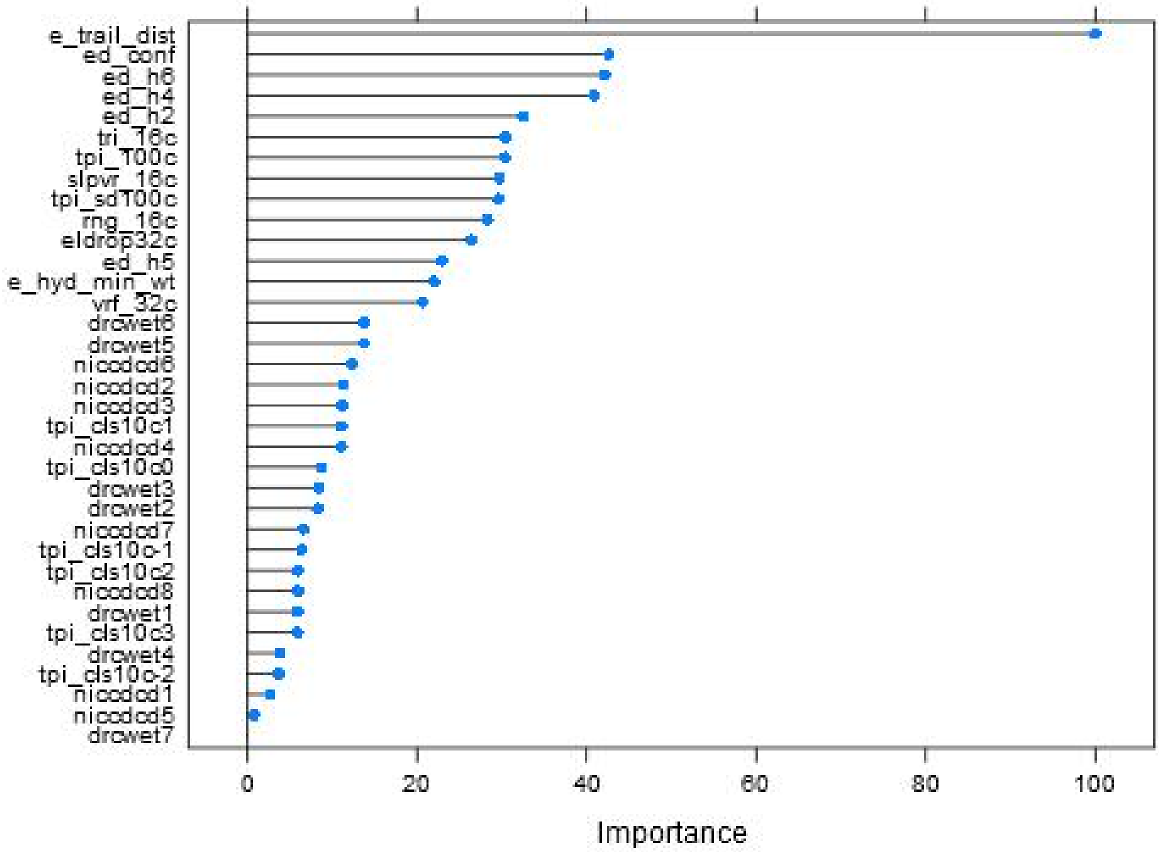


Chart 20. Region 8 All – Riverine Section 2

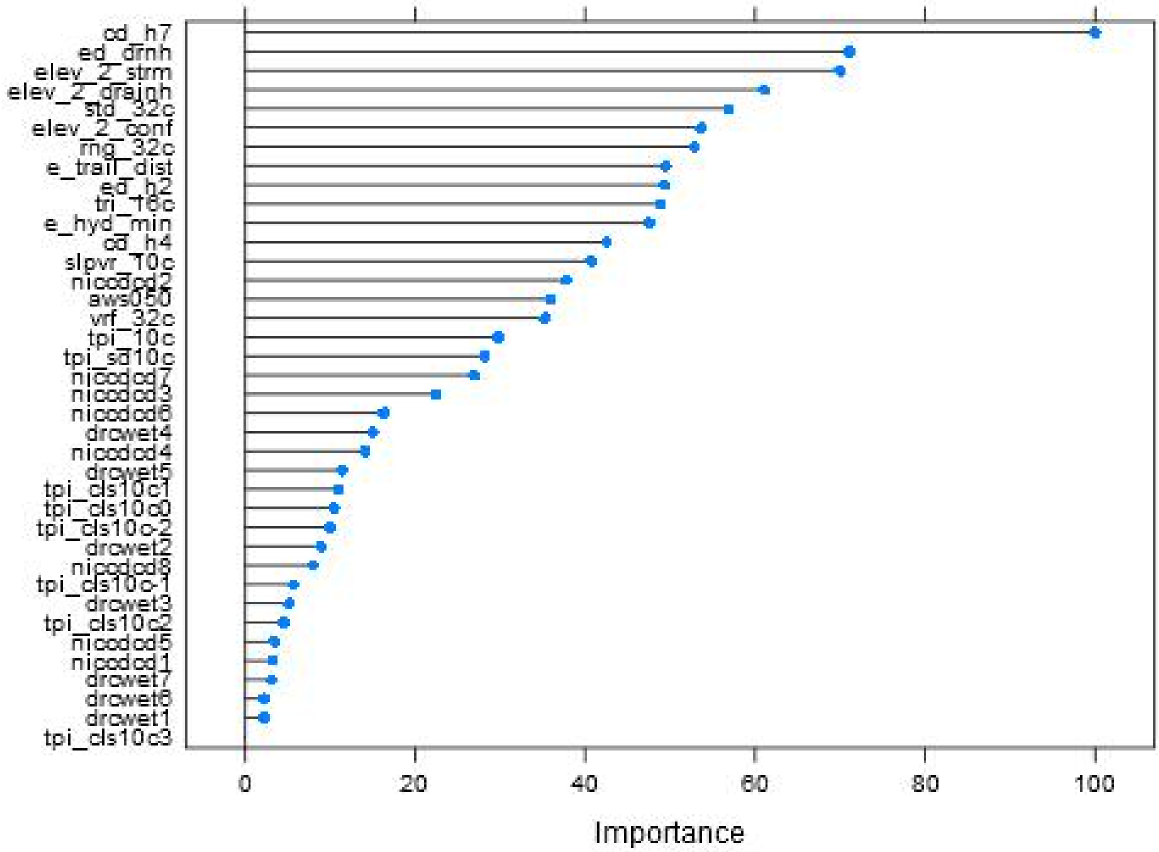


Chart 21. Region 8 All – Riverine Section 3

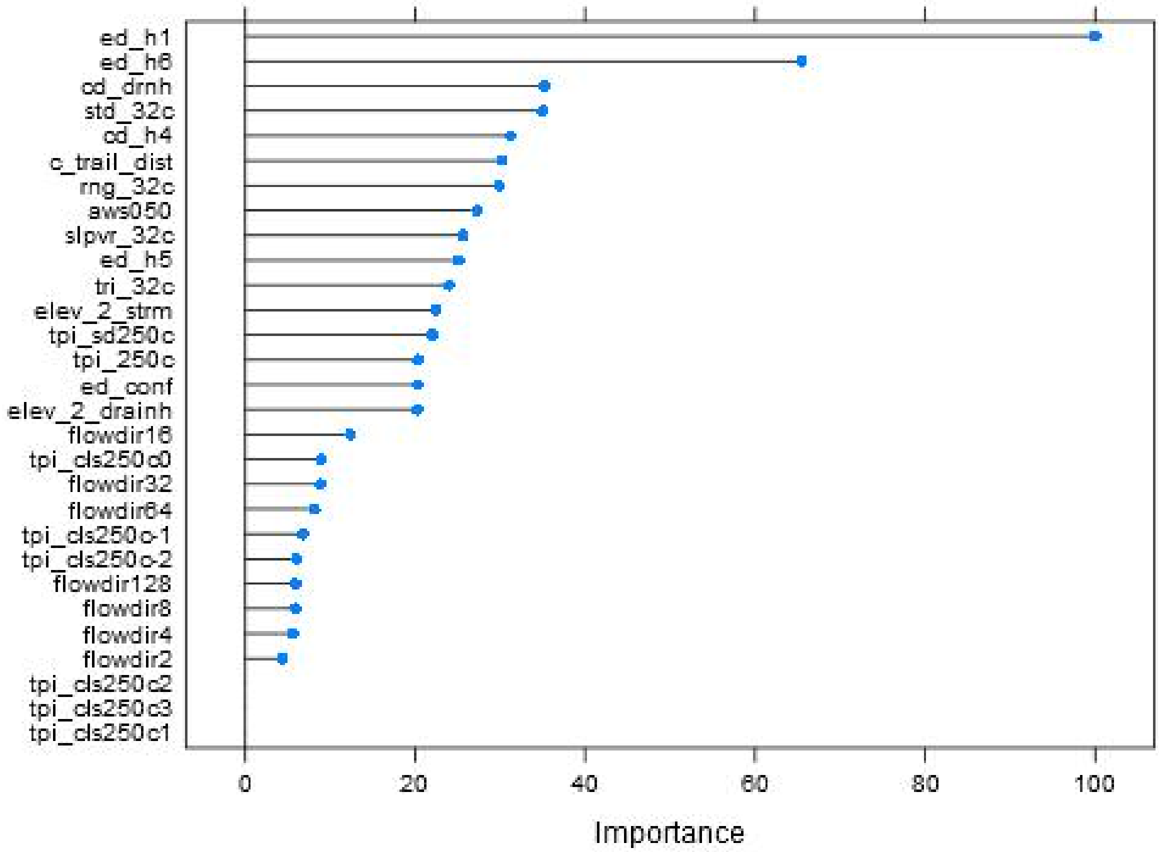


Chart 22. Region 8 All – Riverine Section 4

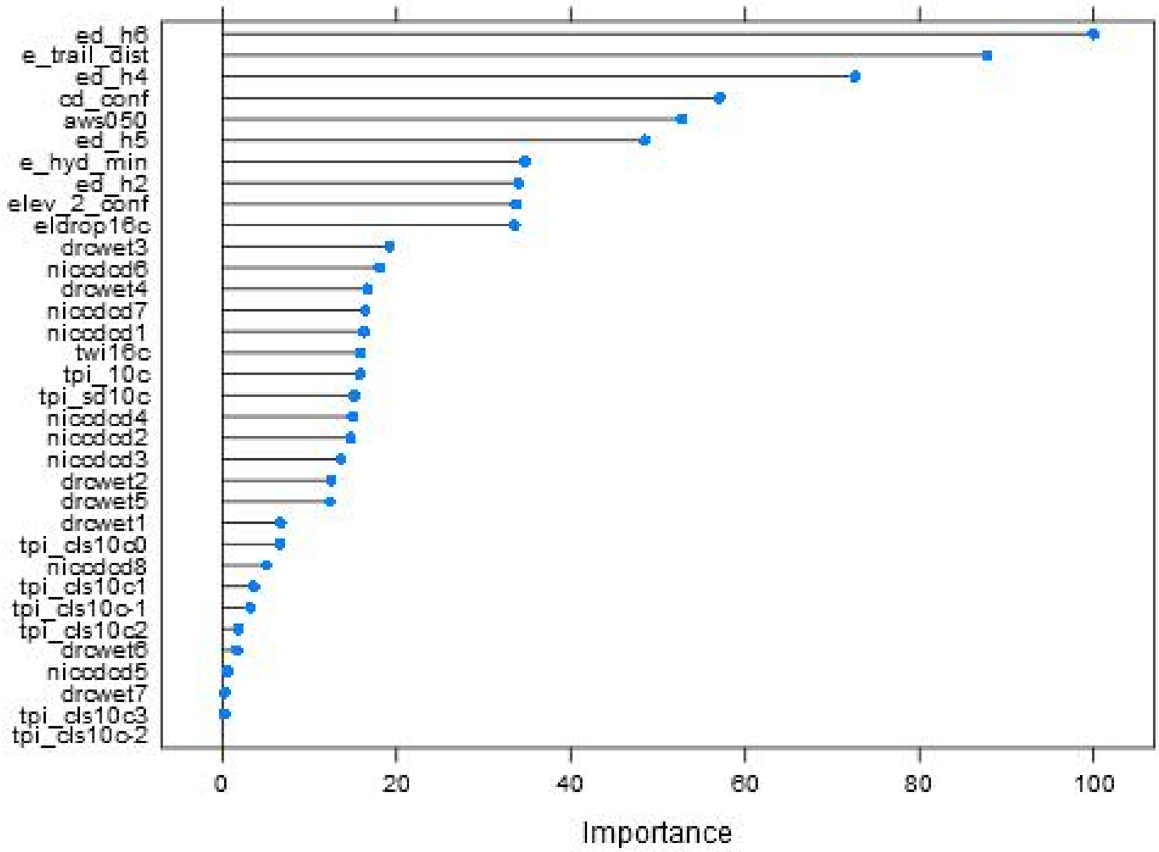


Chart 23. Region 8 All – Riverine Section 5

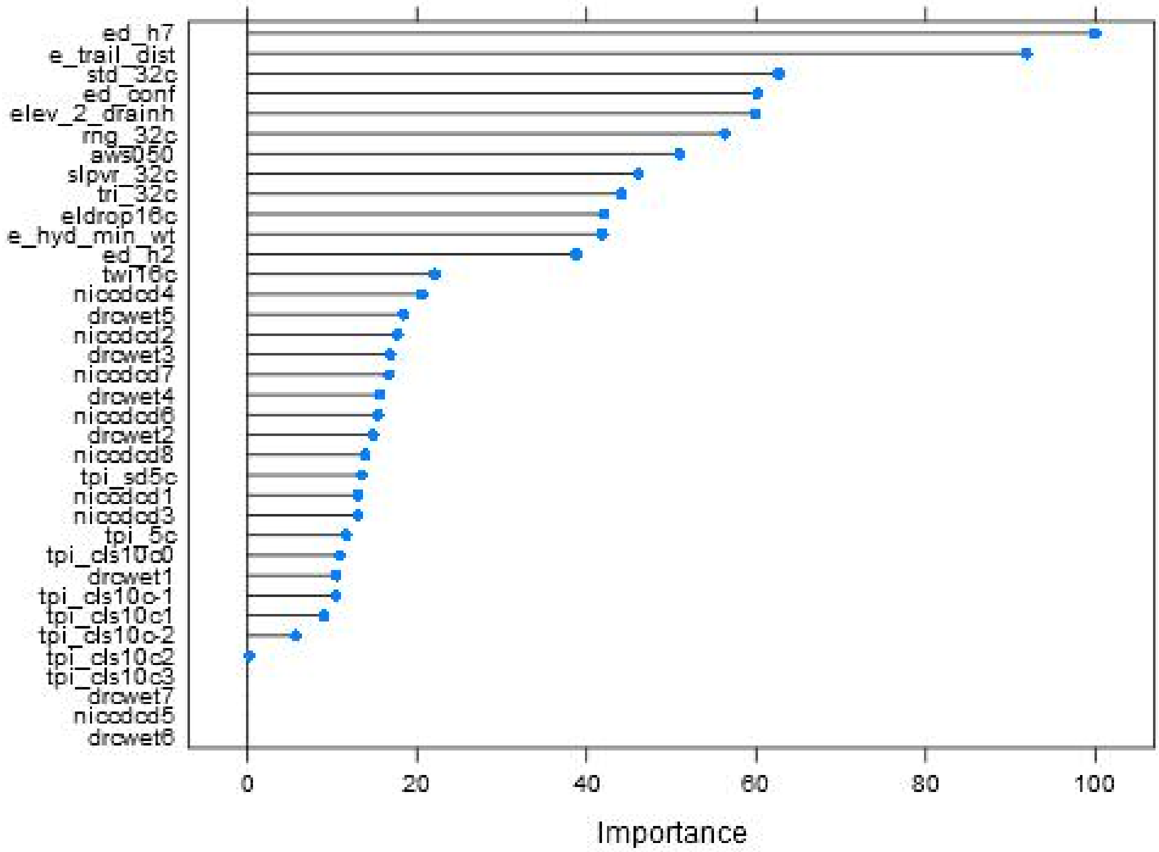


Chart 24. Region 8 All – Riverine Section 6

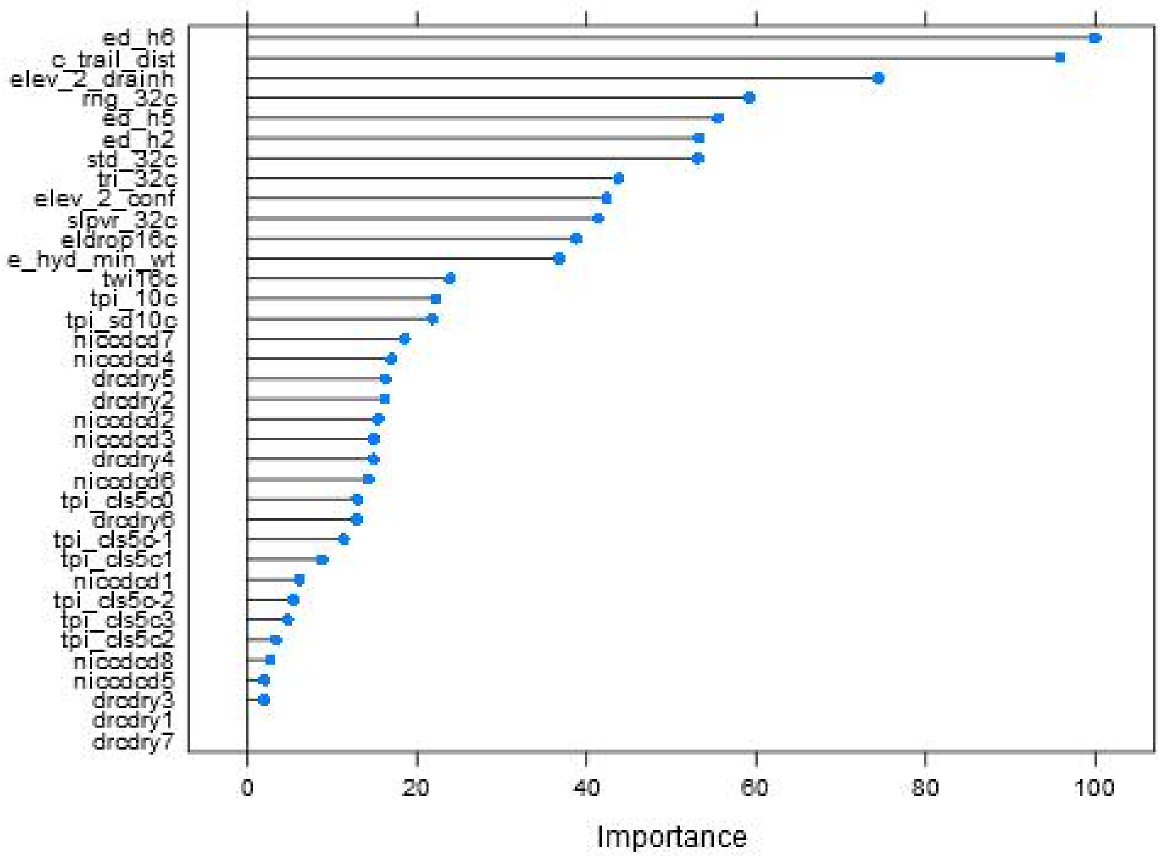


Chart 25. Region 8 All – Riverine Section 7

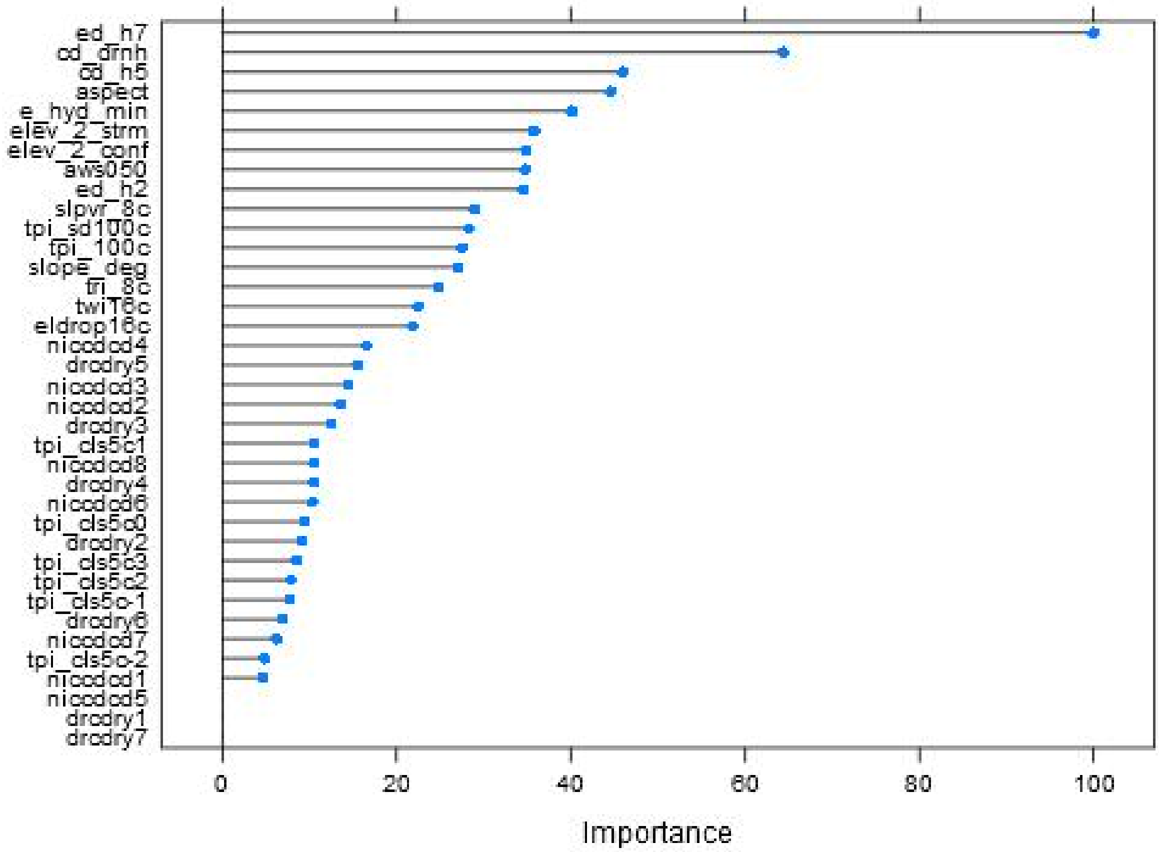


Chart 26. Region 8 All – Riverine Section 8

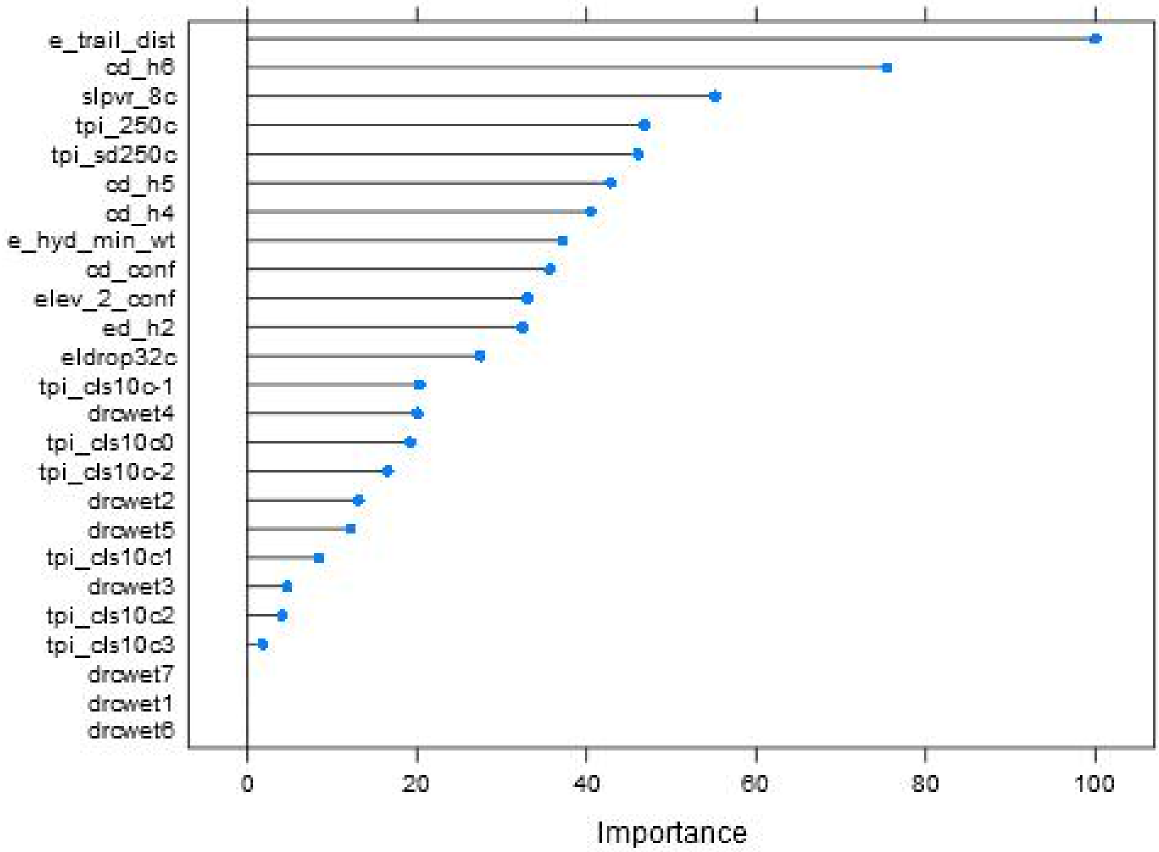


Chart 27. Region 8 All – Riverine Section 9

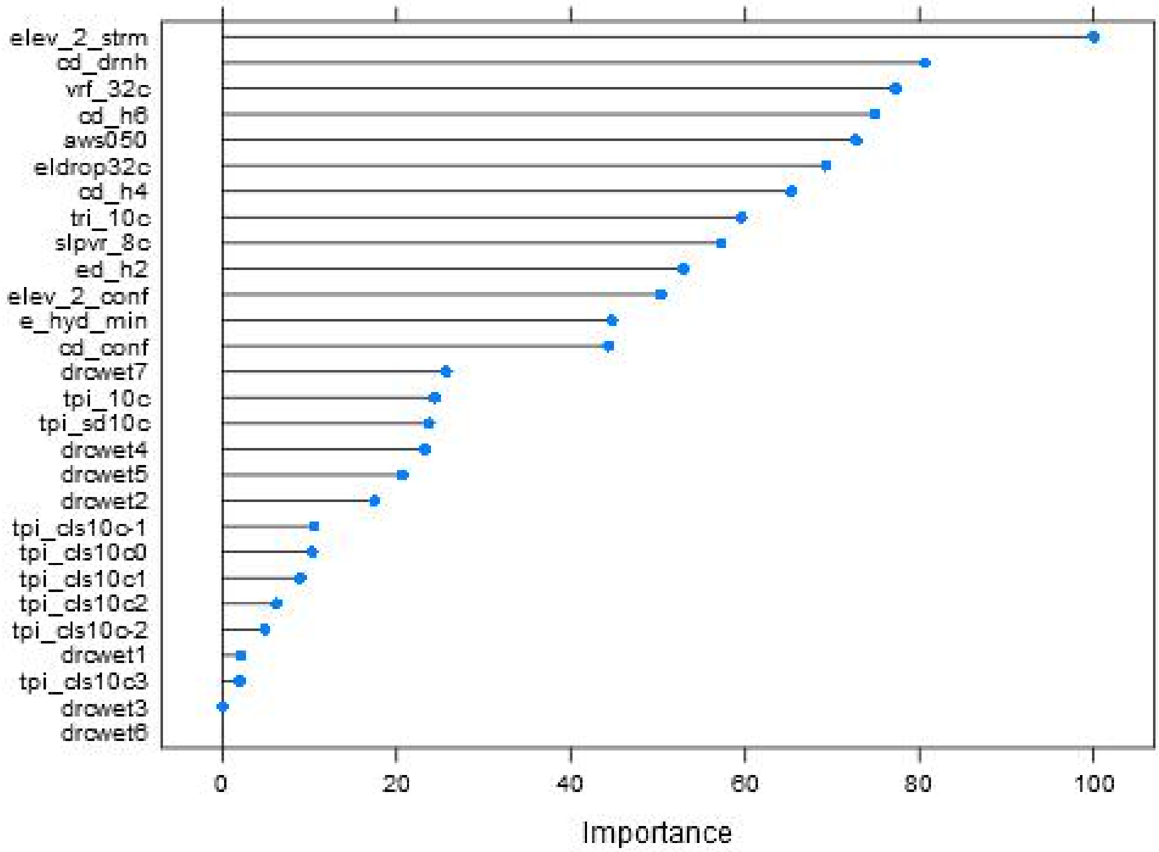


Chart 28. Region 8 All – Upland Section 1

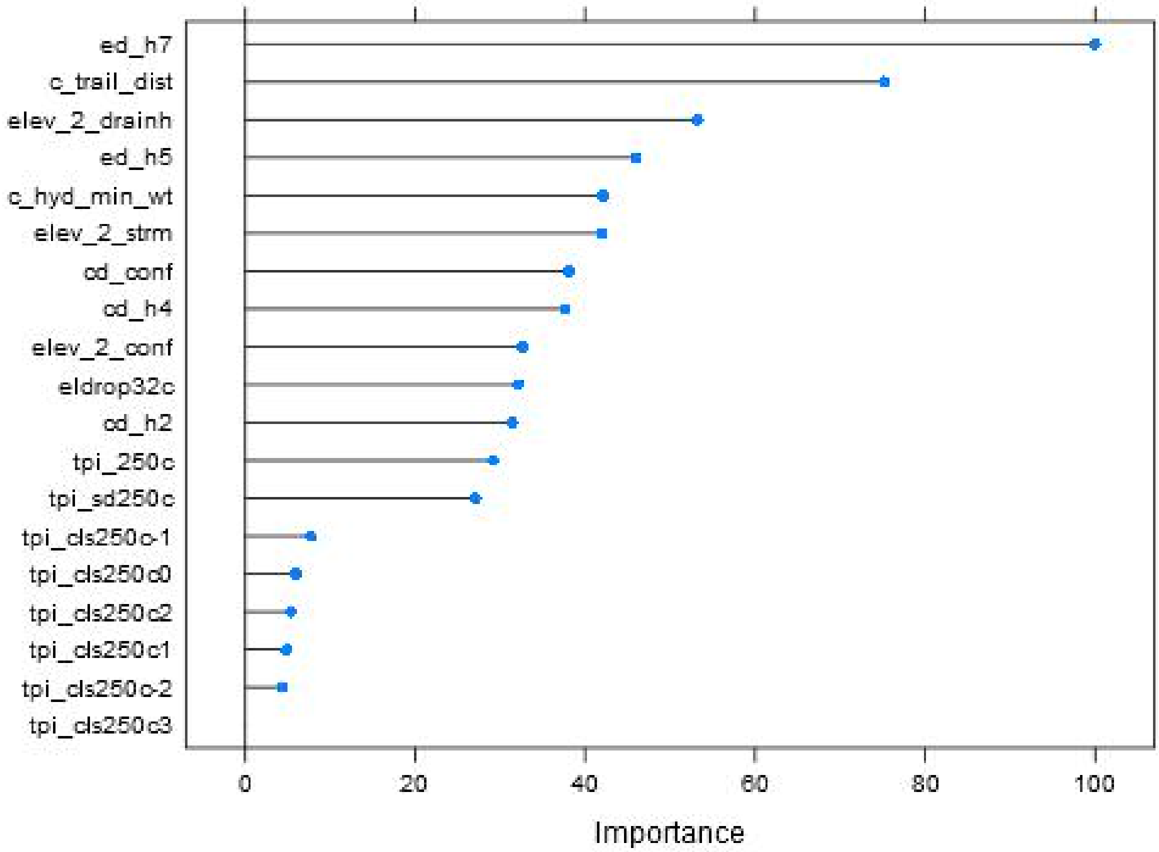


Chart 29. Region 8 All – Upland Section 2

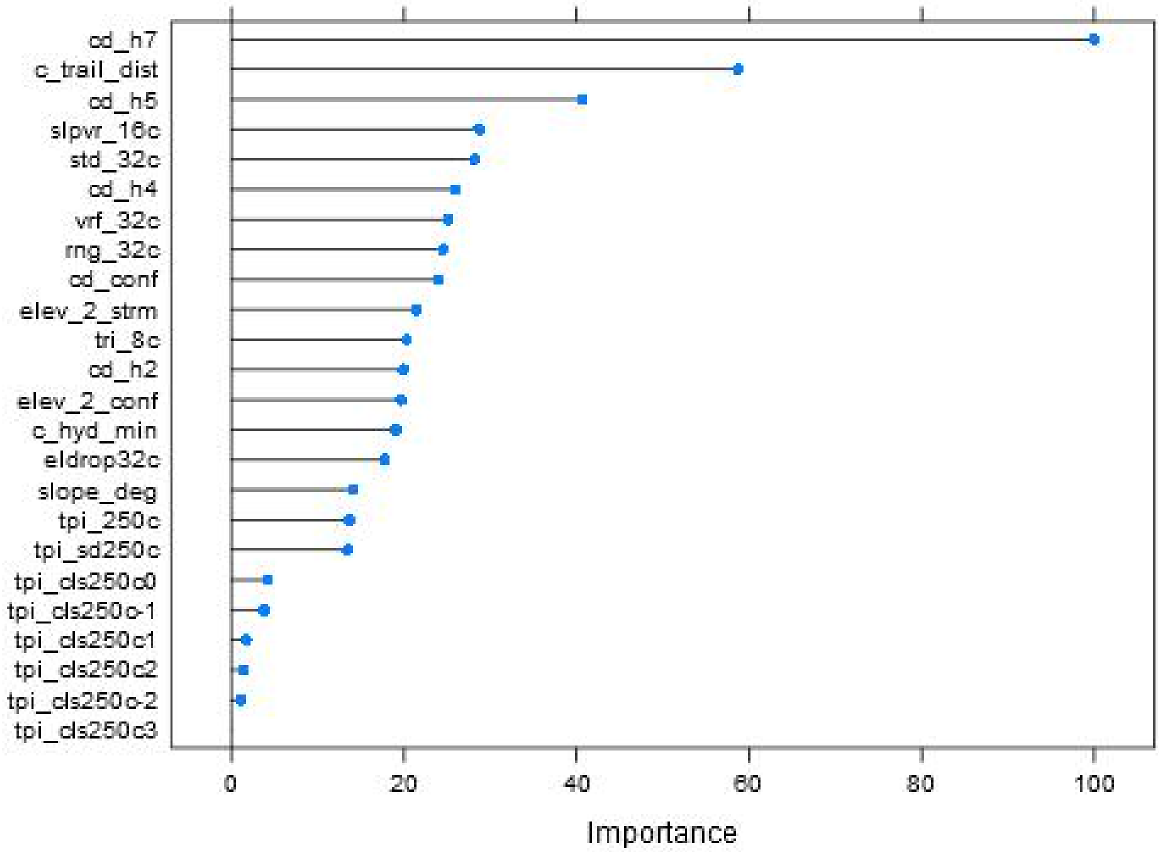


Chart 30. Region 8 All – Upland Section 3

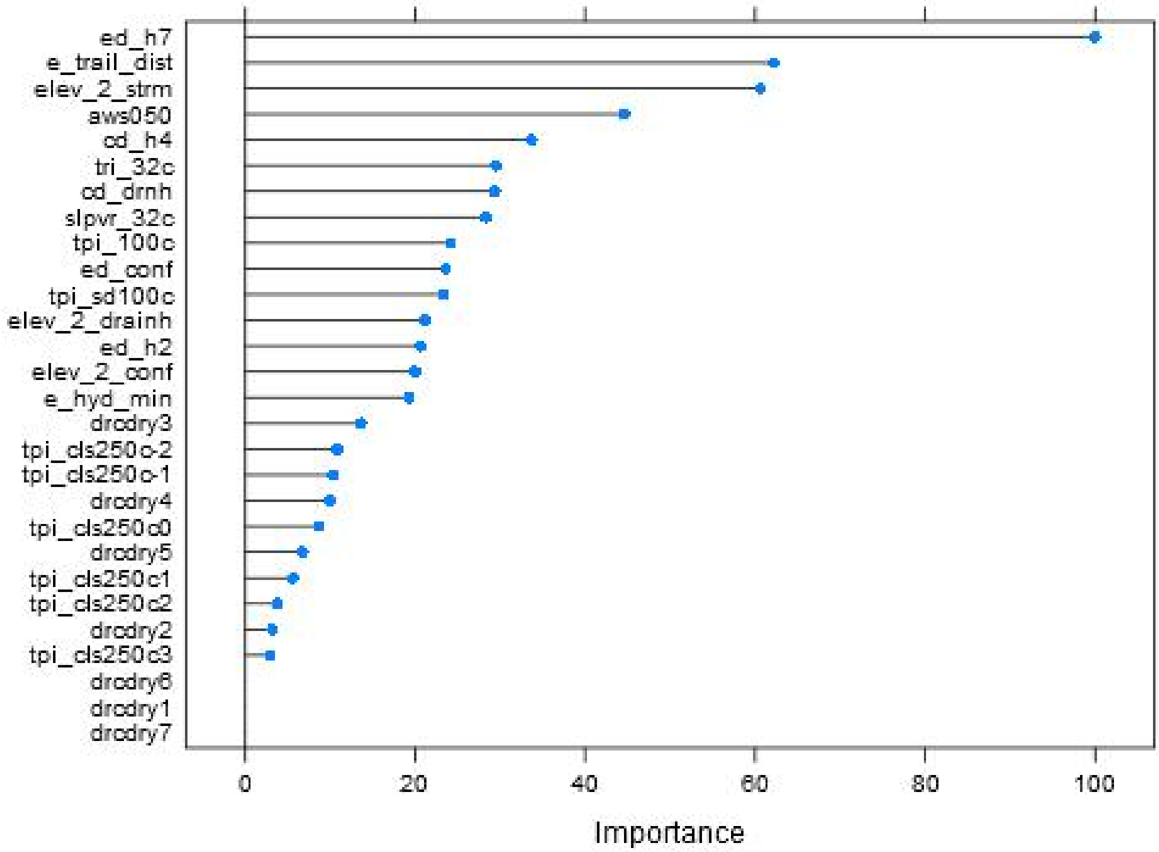


Chart 31. Region 8 All – Upland Section 4

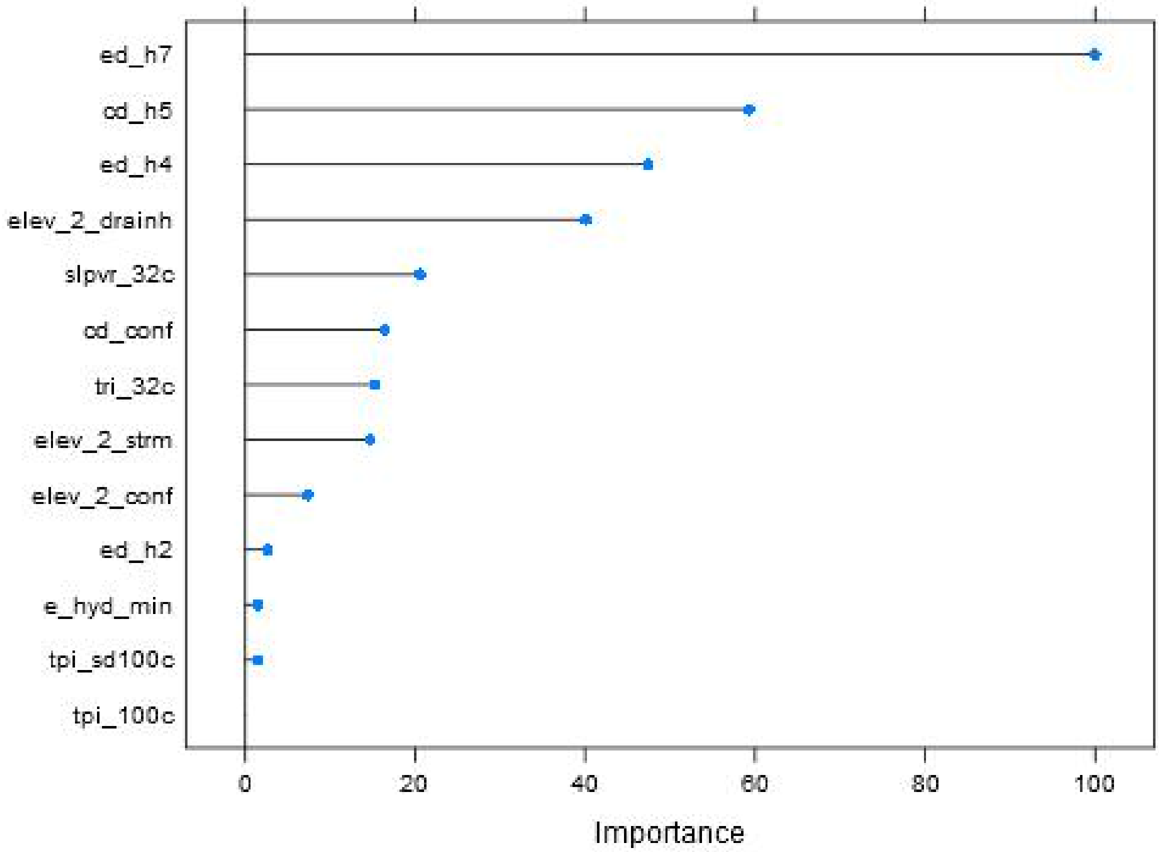


Chart 32. Region 8 All – Upland Section 5

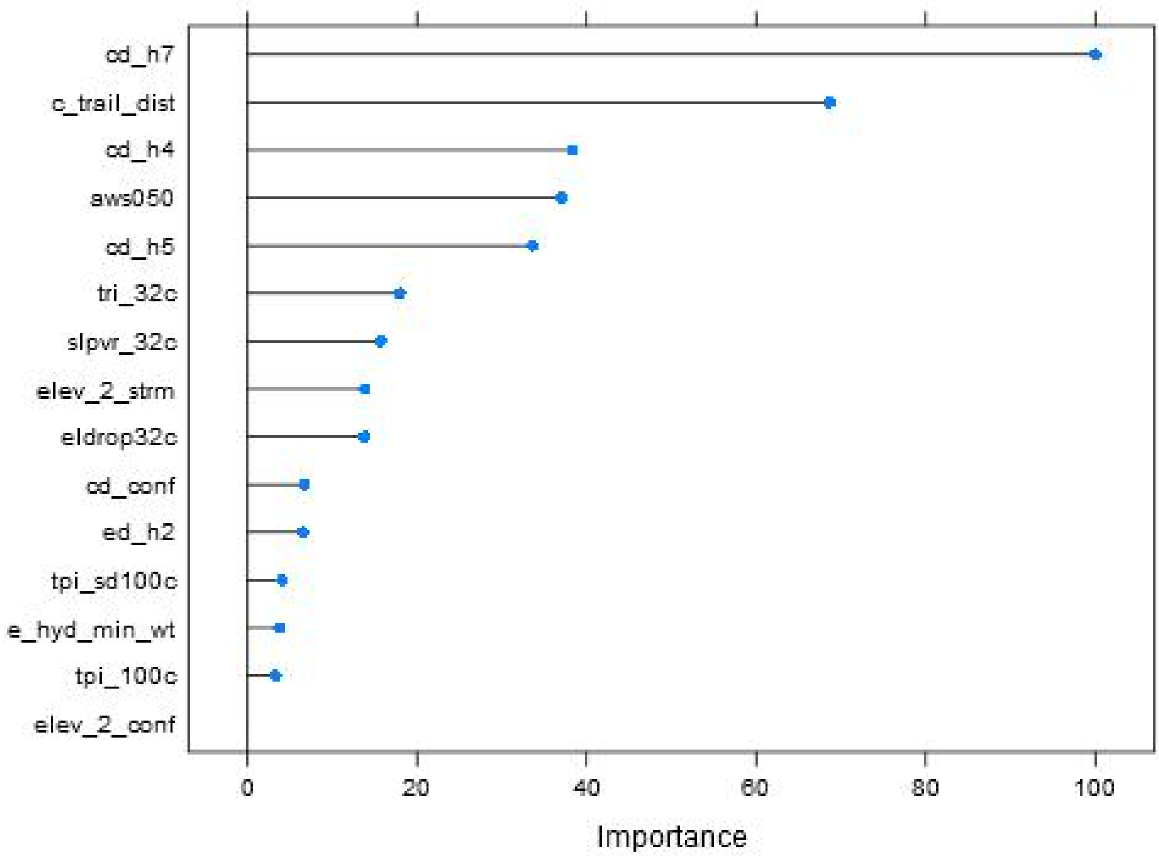


Chart 33. Region 8 All – Upland Section 6

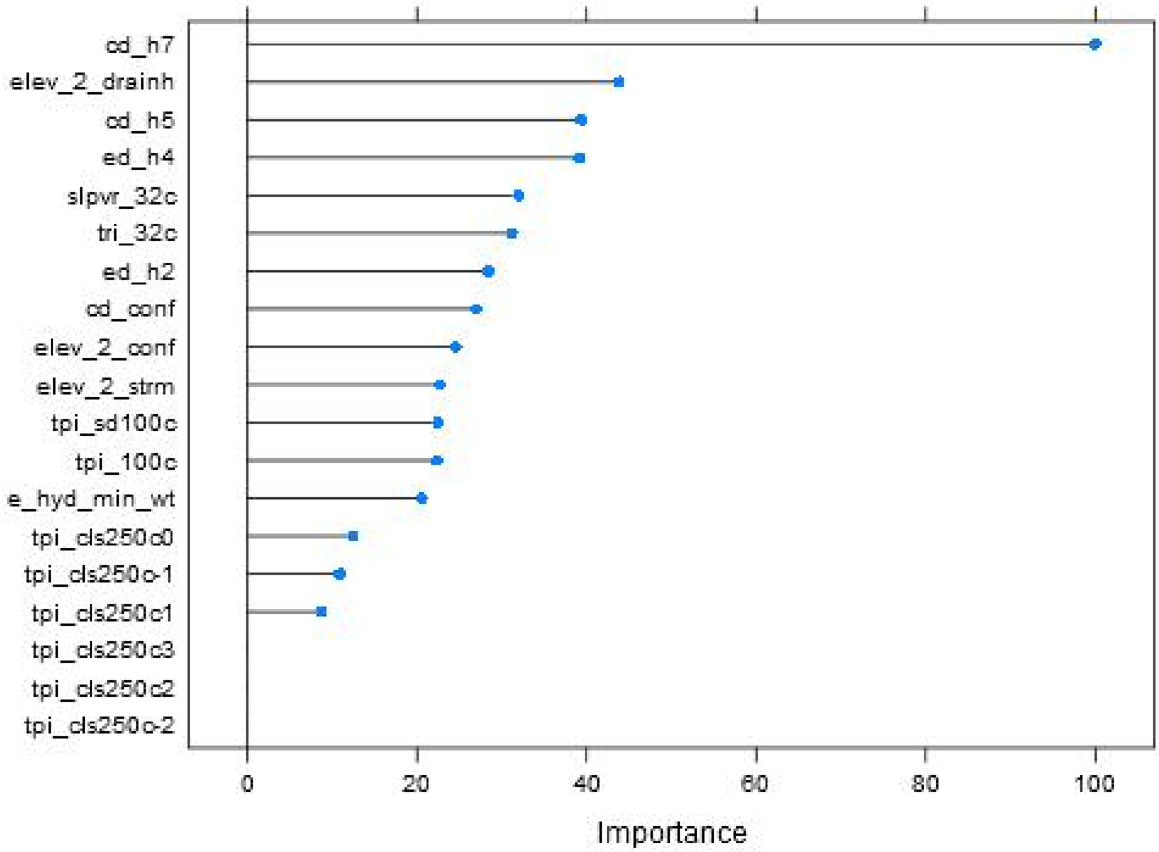


Chart 34. Region 8 All – Upland Section 7

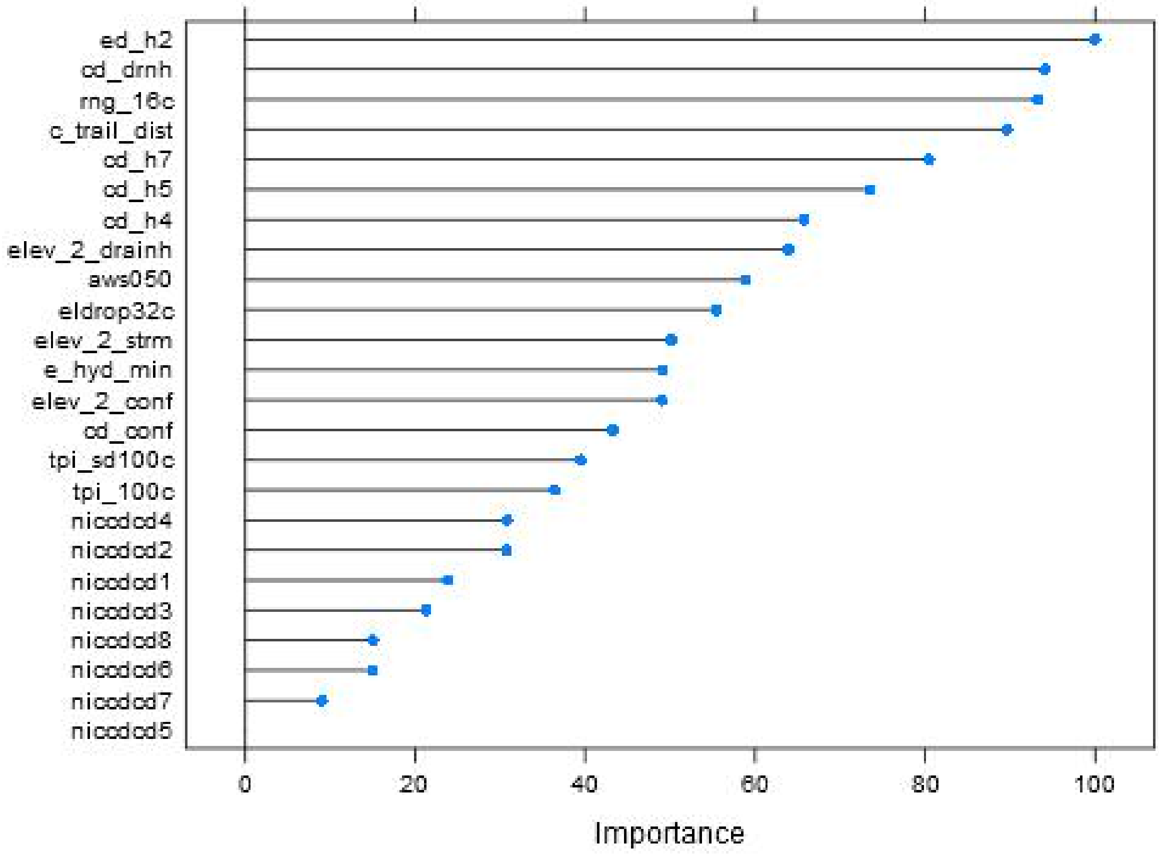


Chart 35. Region 8 All – Upland Section 8

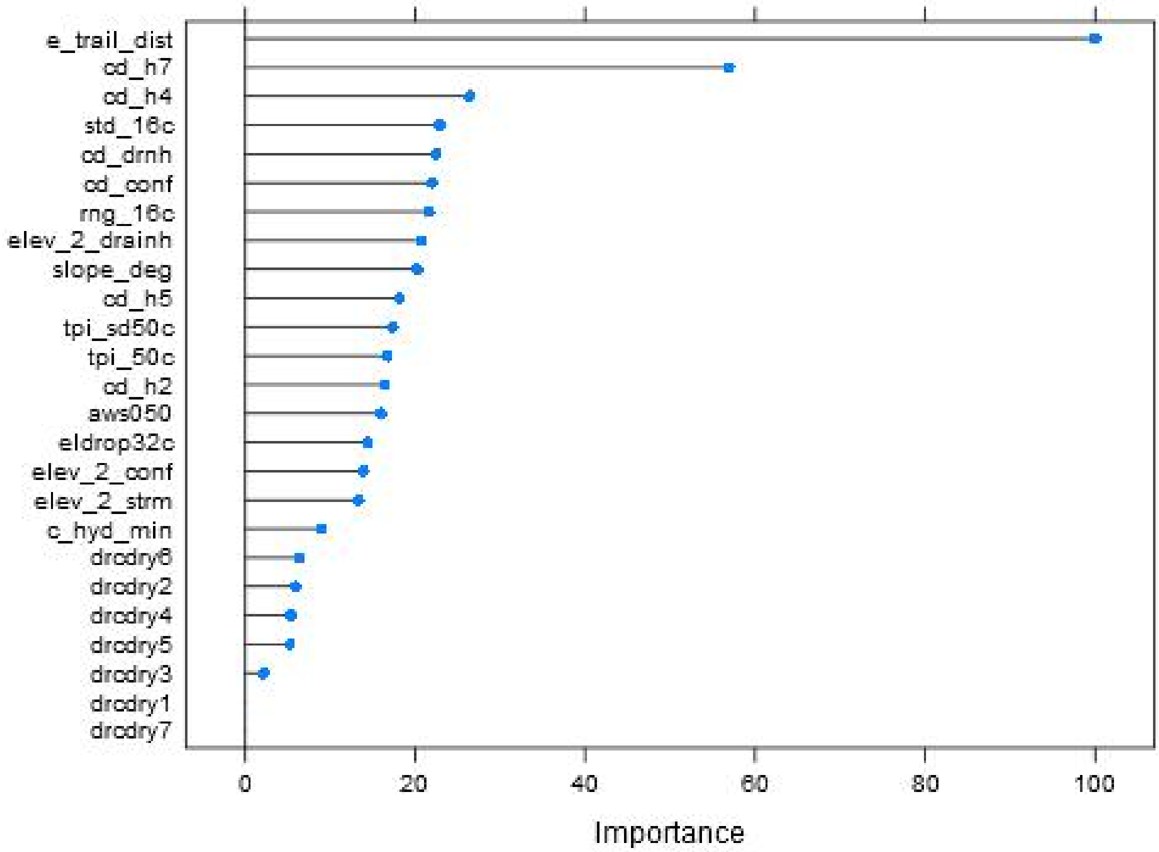


Chart 36. Region 8 All – Upland Section 9

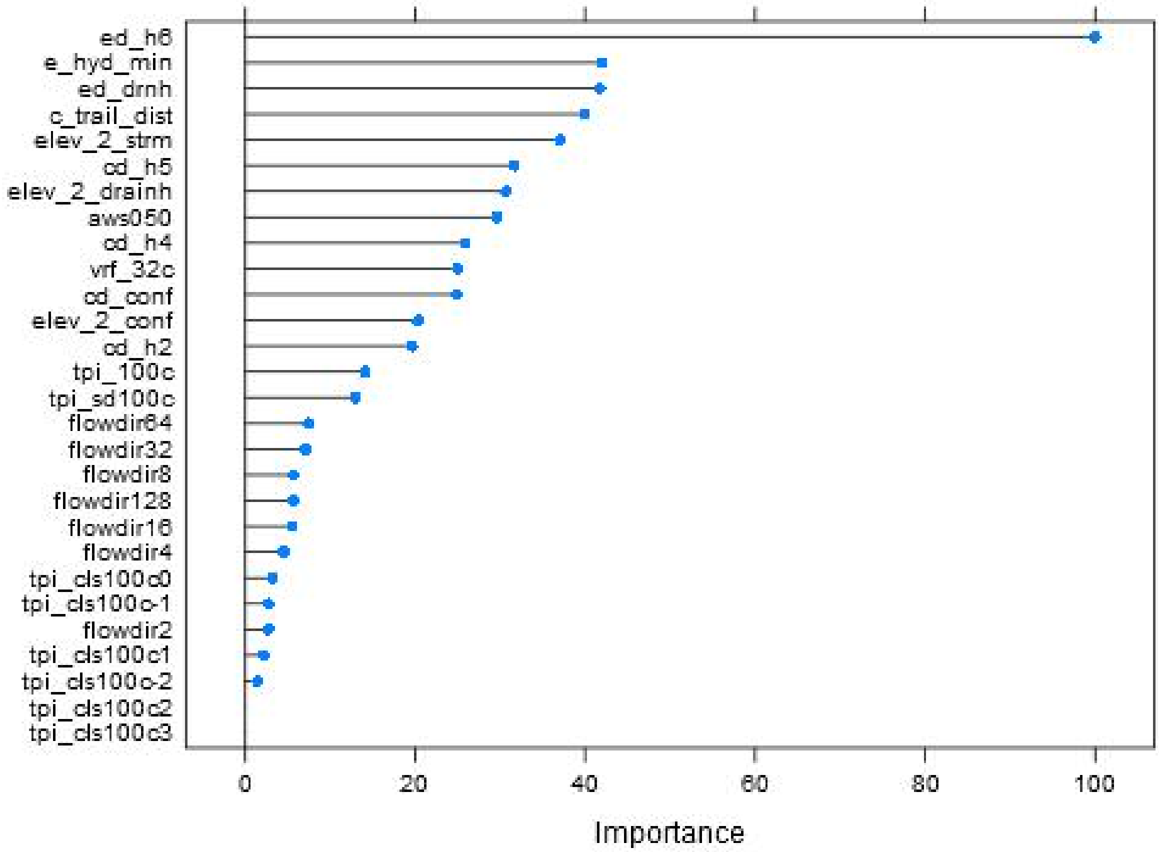


Chart 37. Region 9/10 All – Riverine Section 1

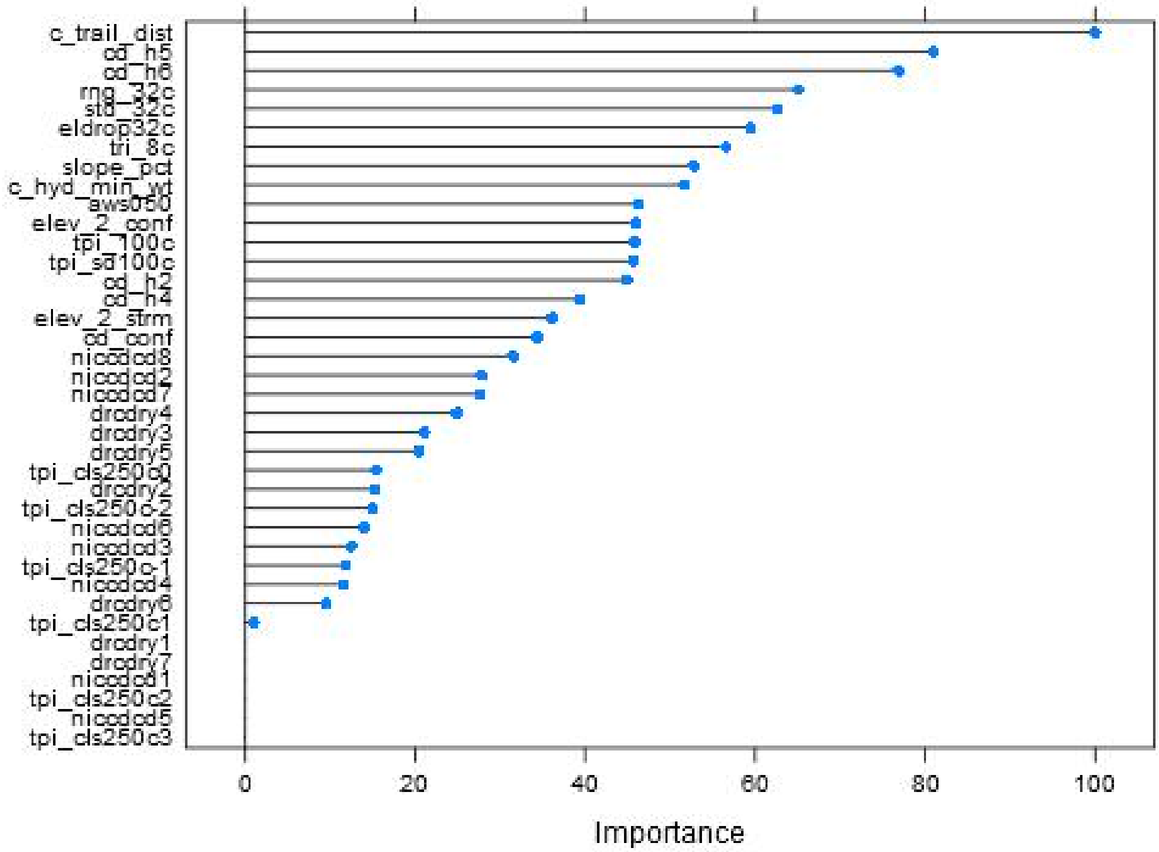


Chart 38. Region 9/10 All – Riverine Section 2

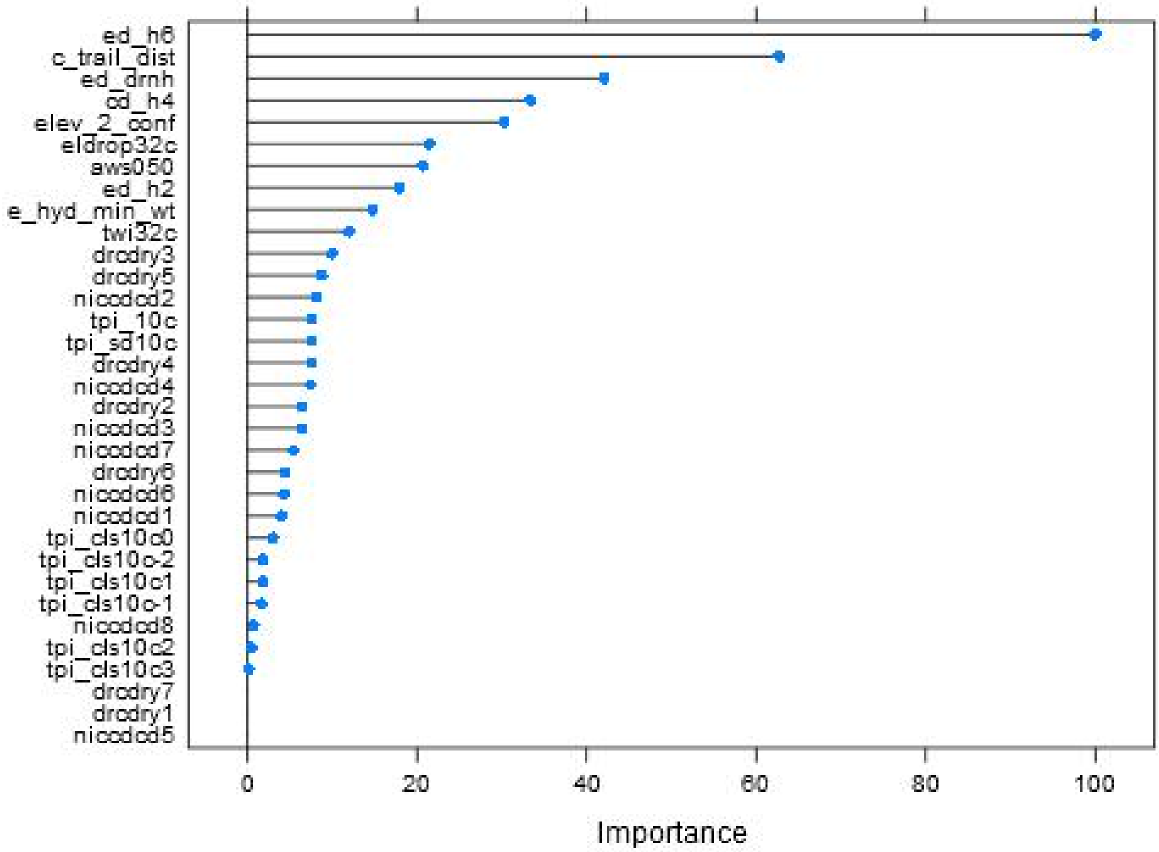


Chart 39. Region 9/10 All – Riverine Section 3

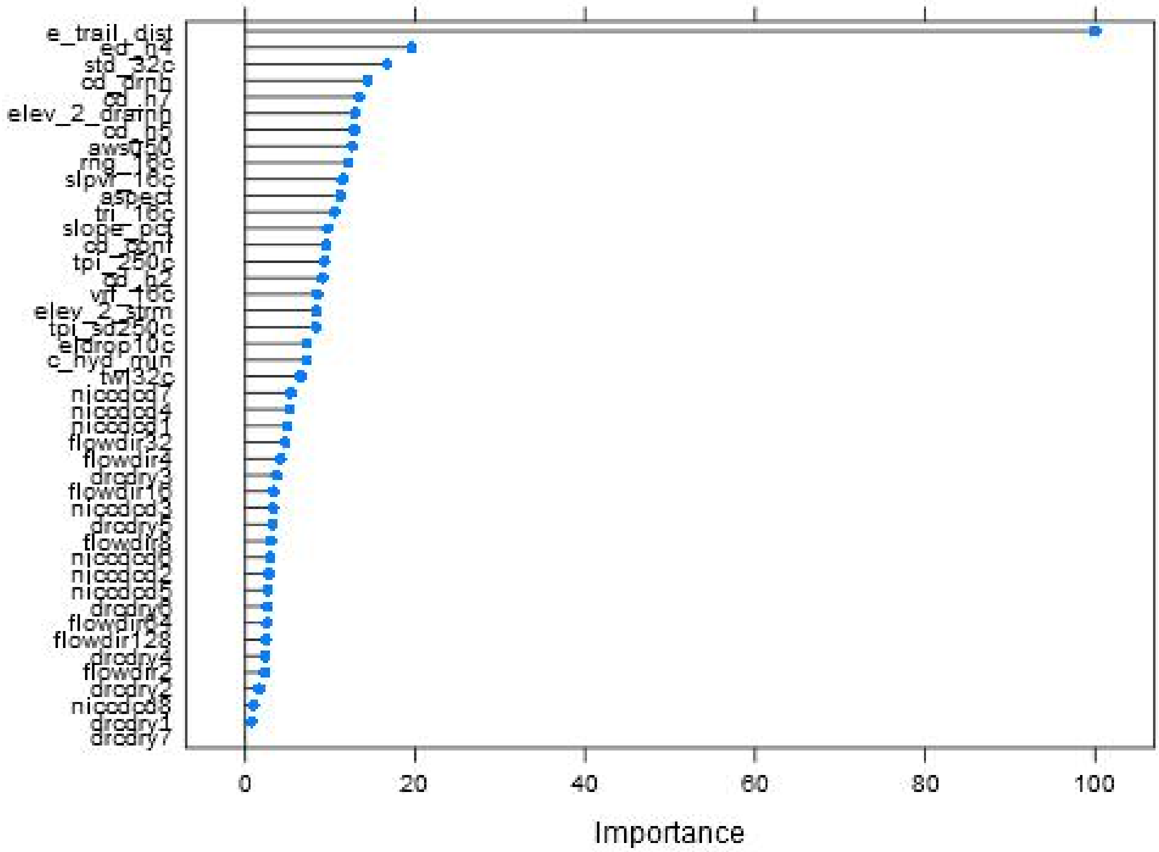


Chart 40. Region 9/10 All – Riverine Section 4

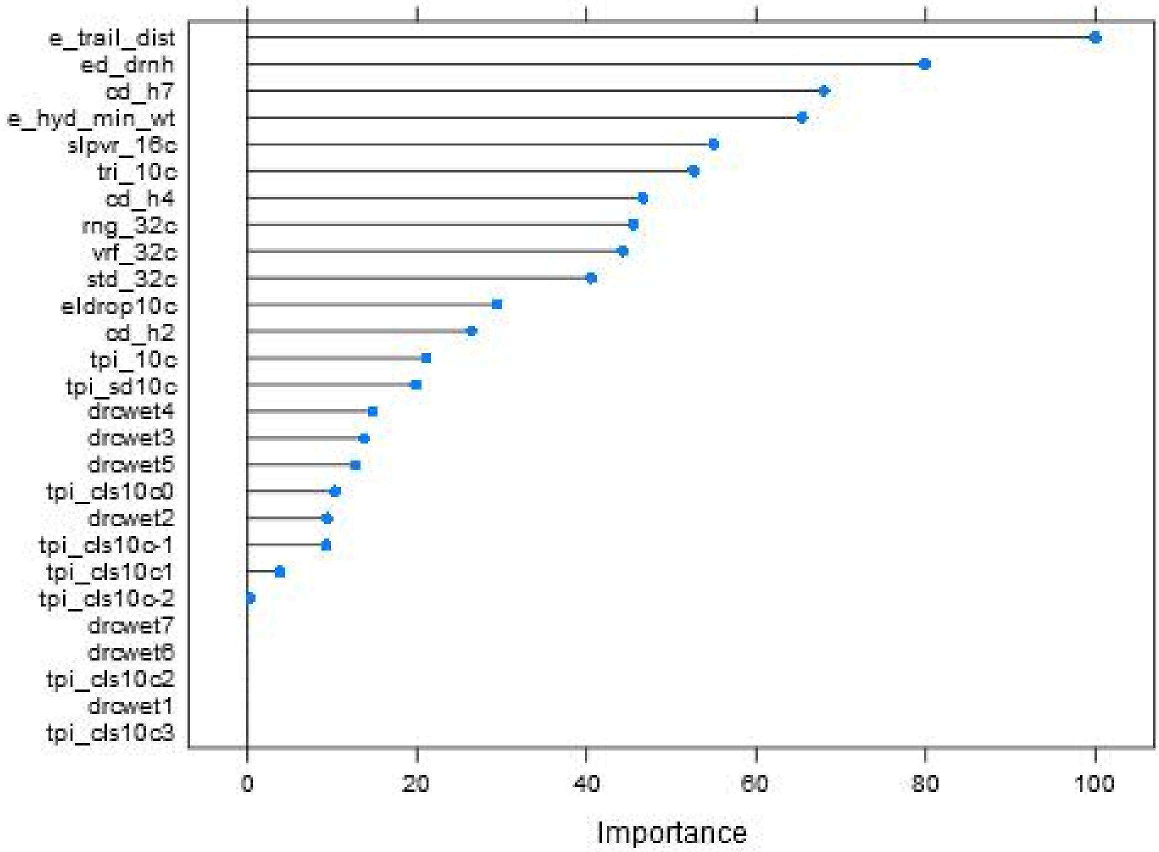


Chart 41. Region 9/10 All – Riverine Section 5

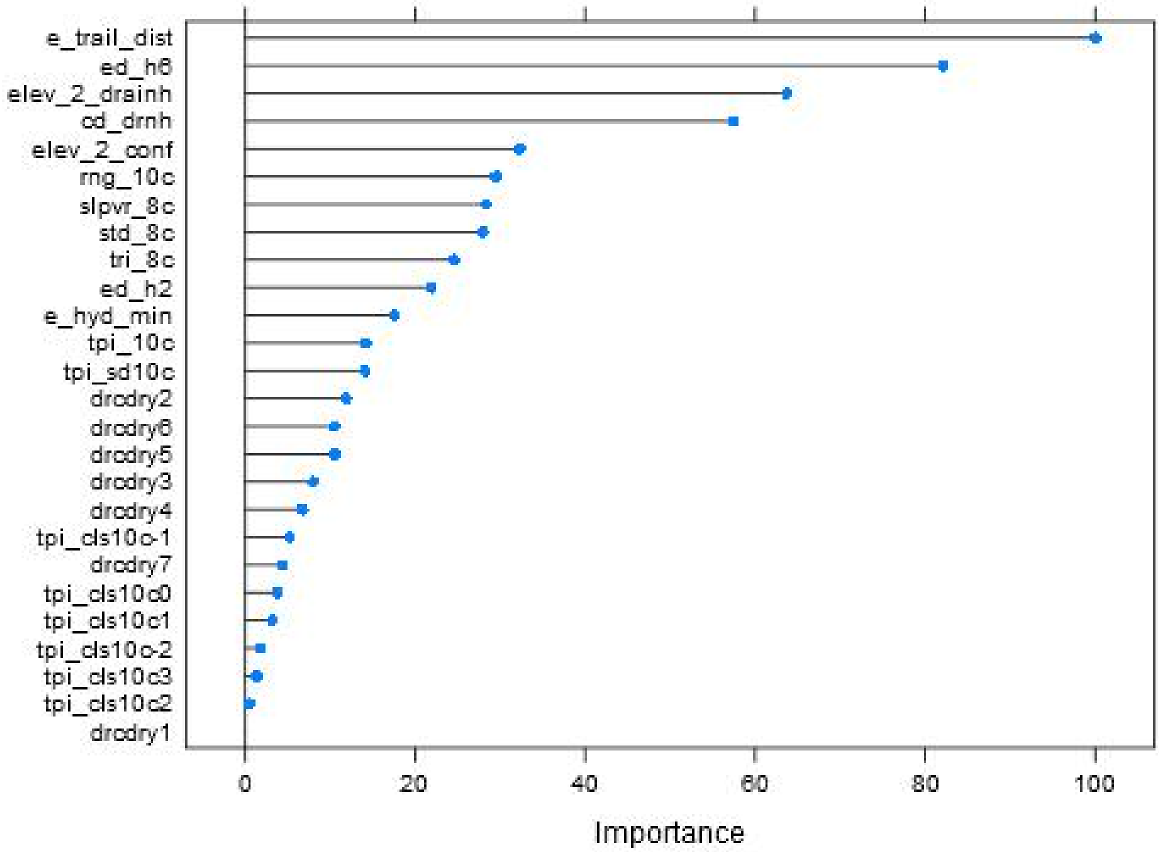


Chart 42. Region 9/10 All – Riverine Section 6

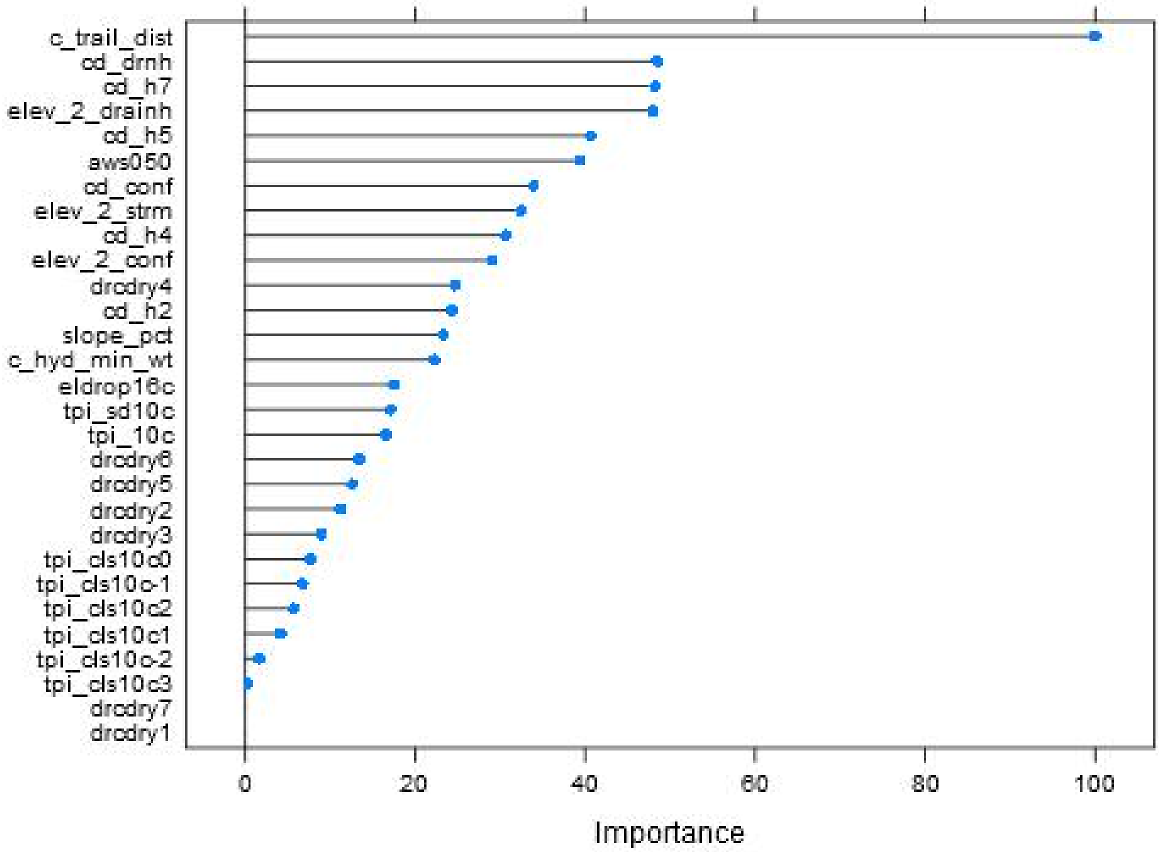


Chart 43. Region 9/10 All – Riverine Section 7

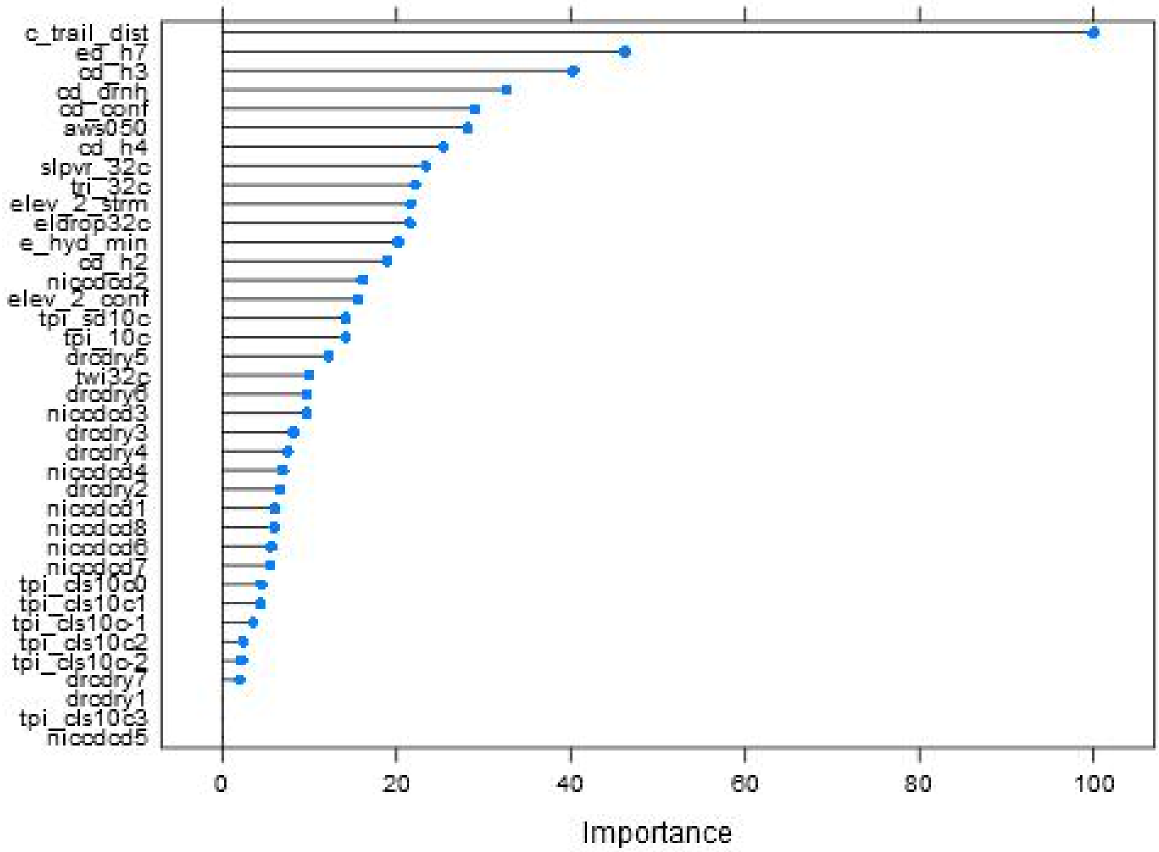


Chart 44. Region 9/10 All – Riverine Section 8

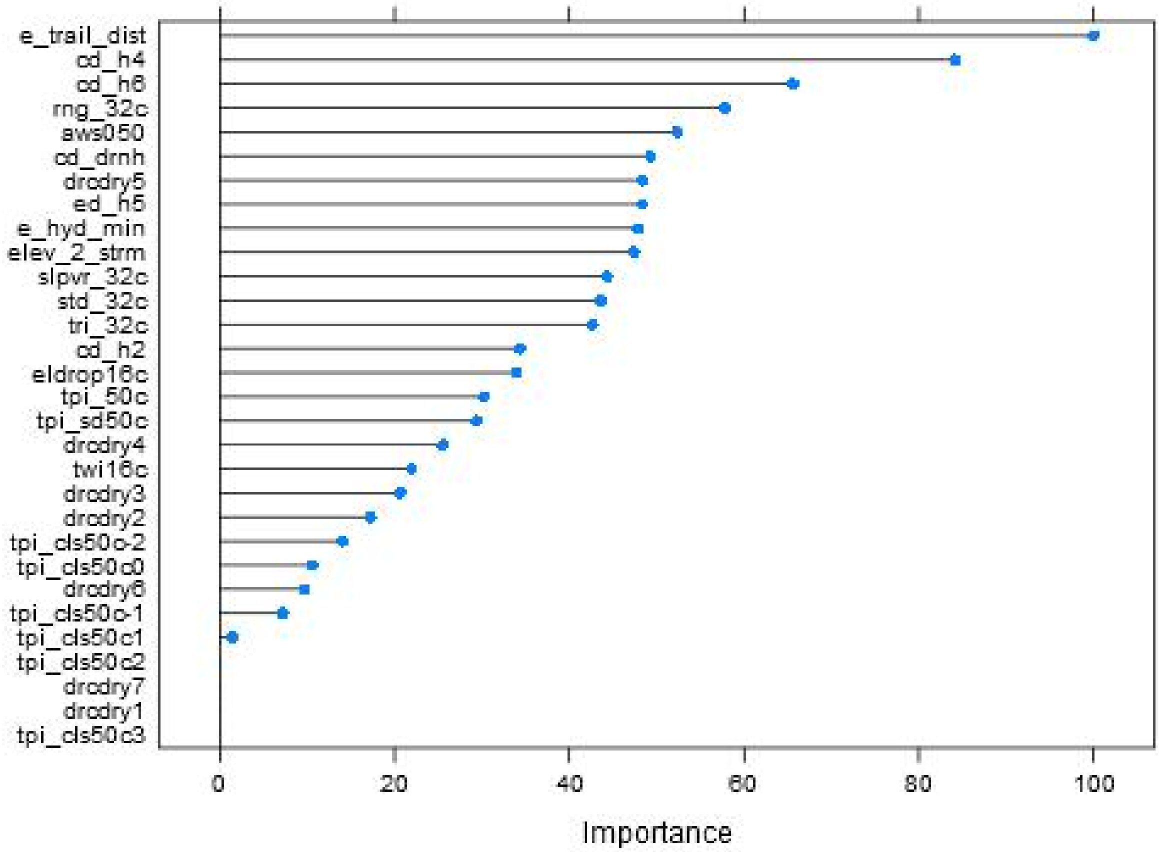


Chart 45. Region 9/10 All – Riverine Section 9

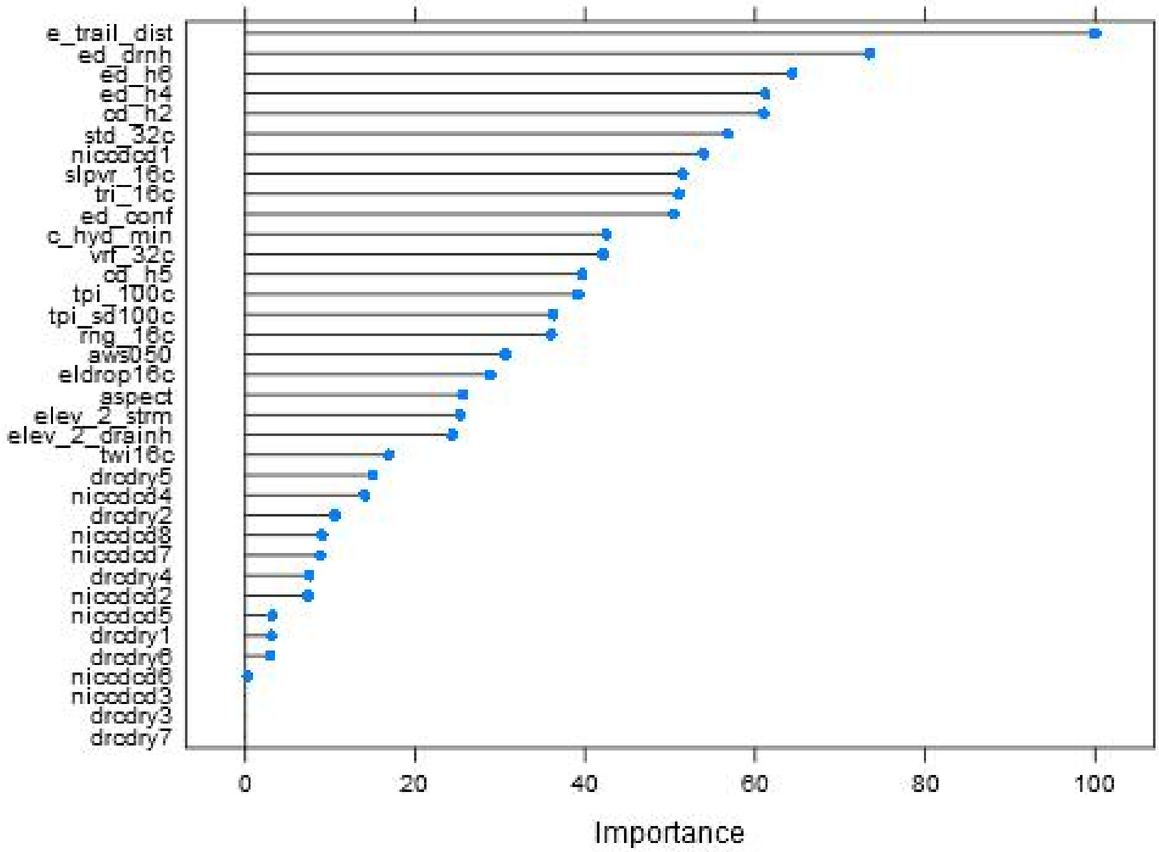


Chart 46. Region 9/10 All – Riverine Section 10

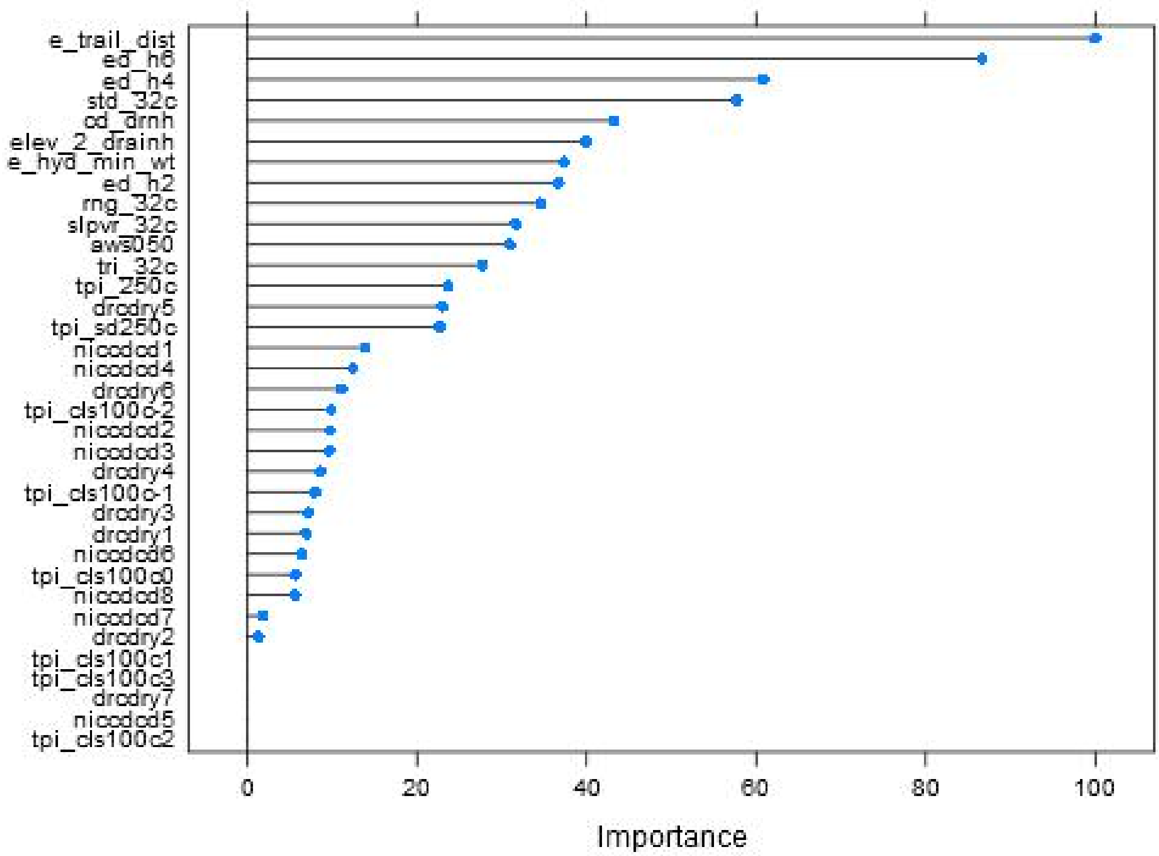


Chart 47. Region 9/10 All – Riverine Section 11

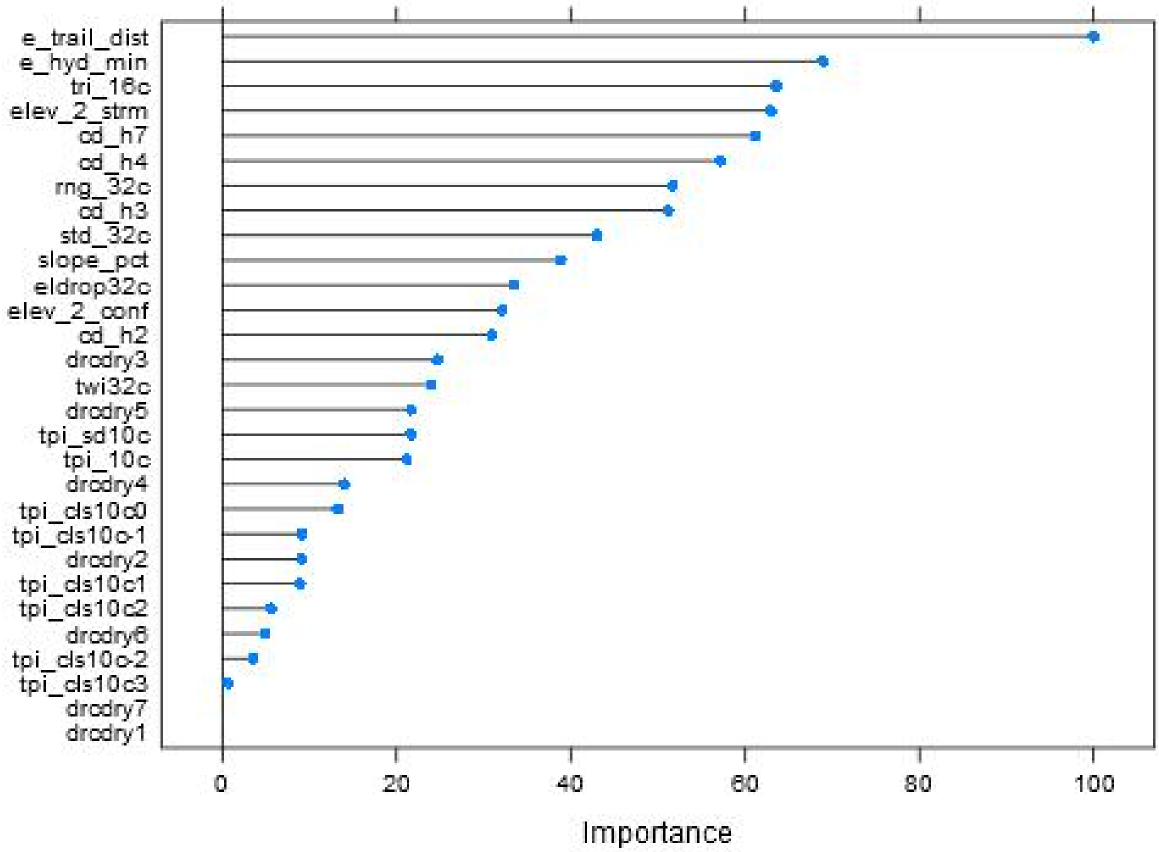


Chart 48. Region 9/10 All – Riverine Section 12

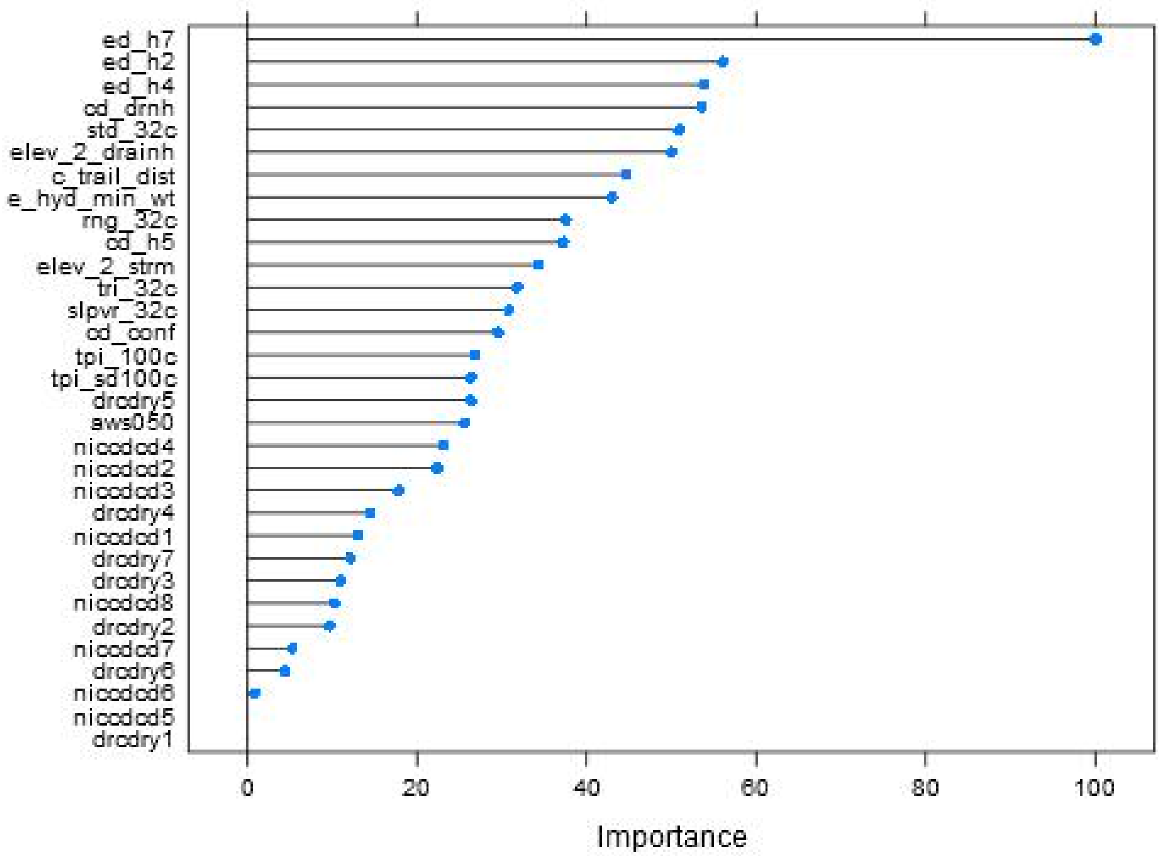


Chart 49. Region 9/10 All – Riverine Section 13

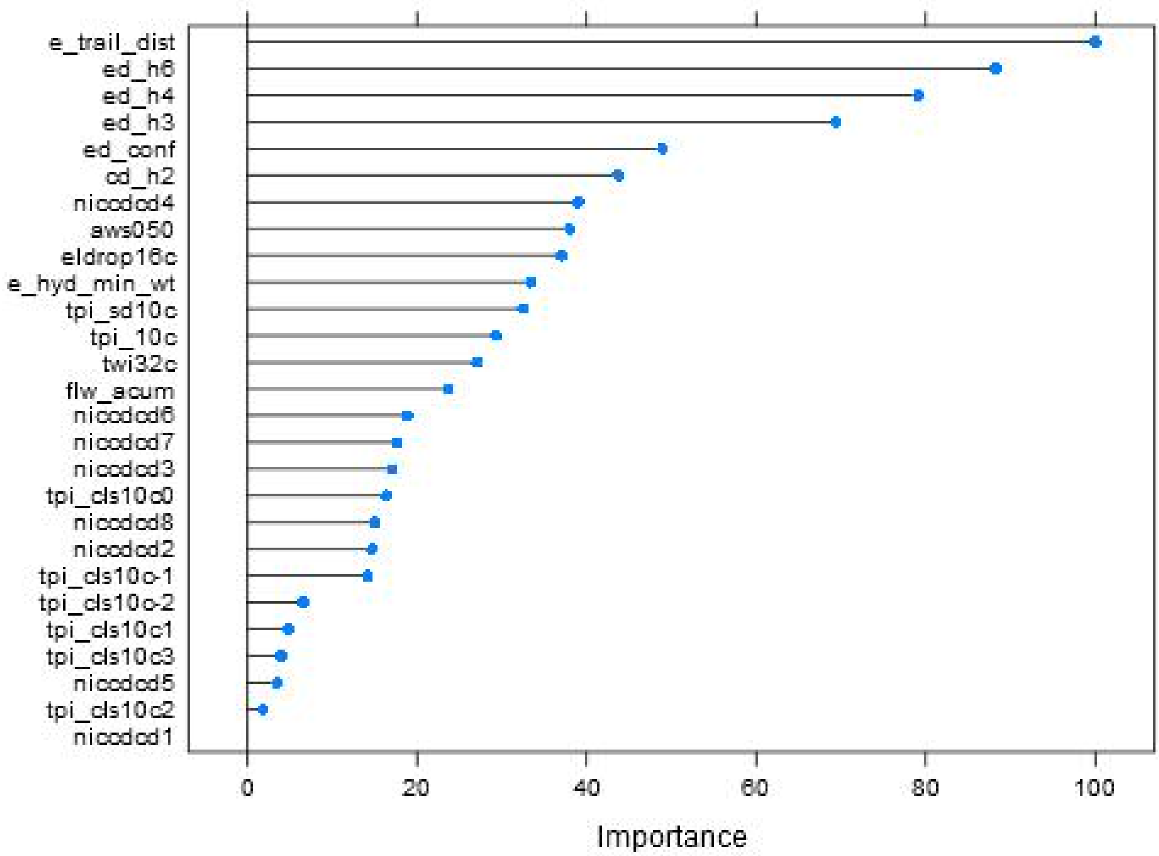


Chart 50. Region 9/10 All – Riverine Section 14

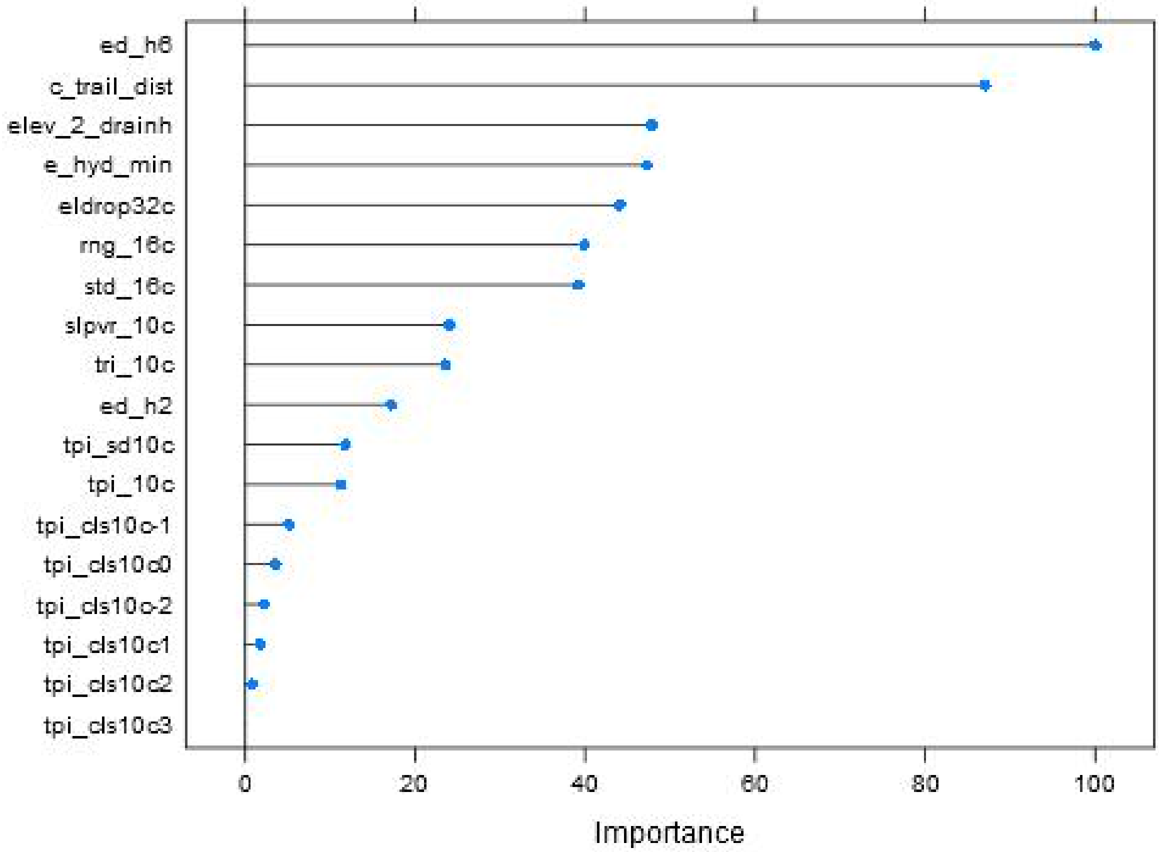


Chart 51. Region 9/10 All – Riverine Section 15

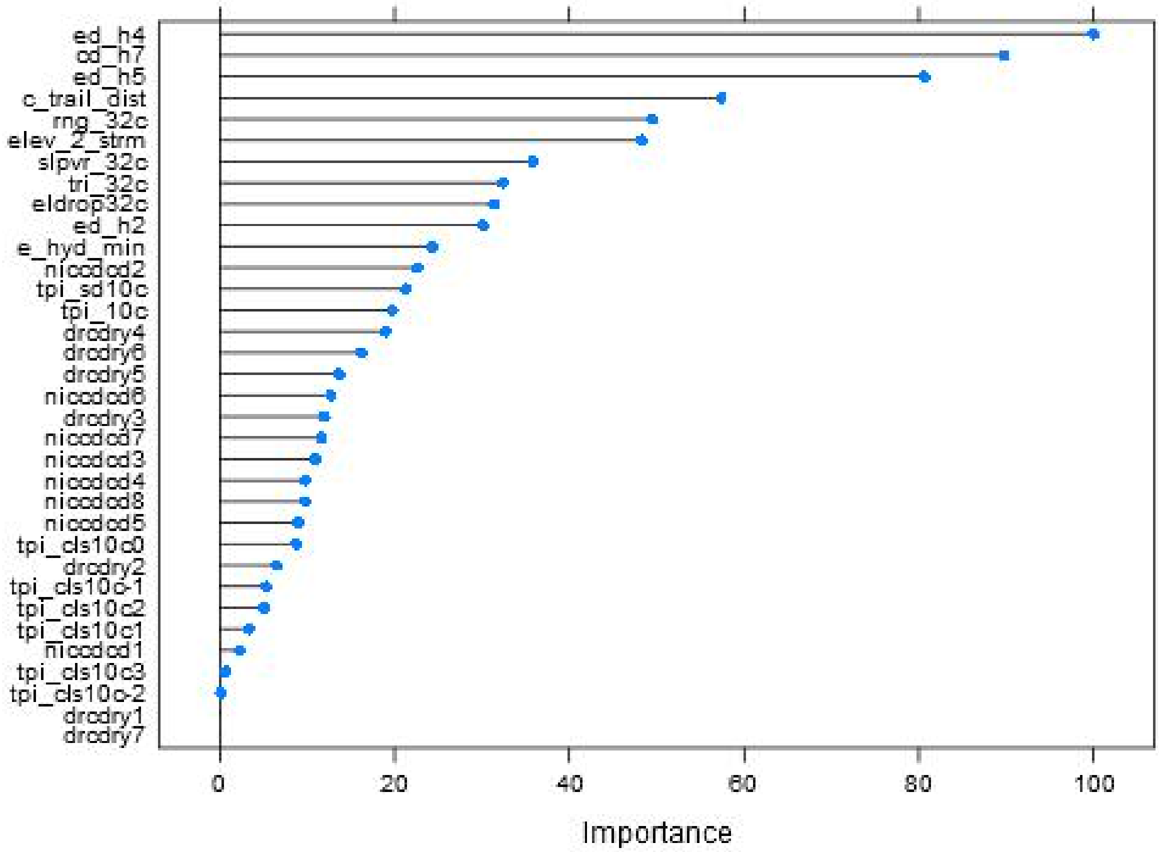


Chart 52. Region 9/10 All – Upland Section 1

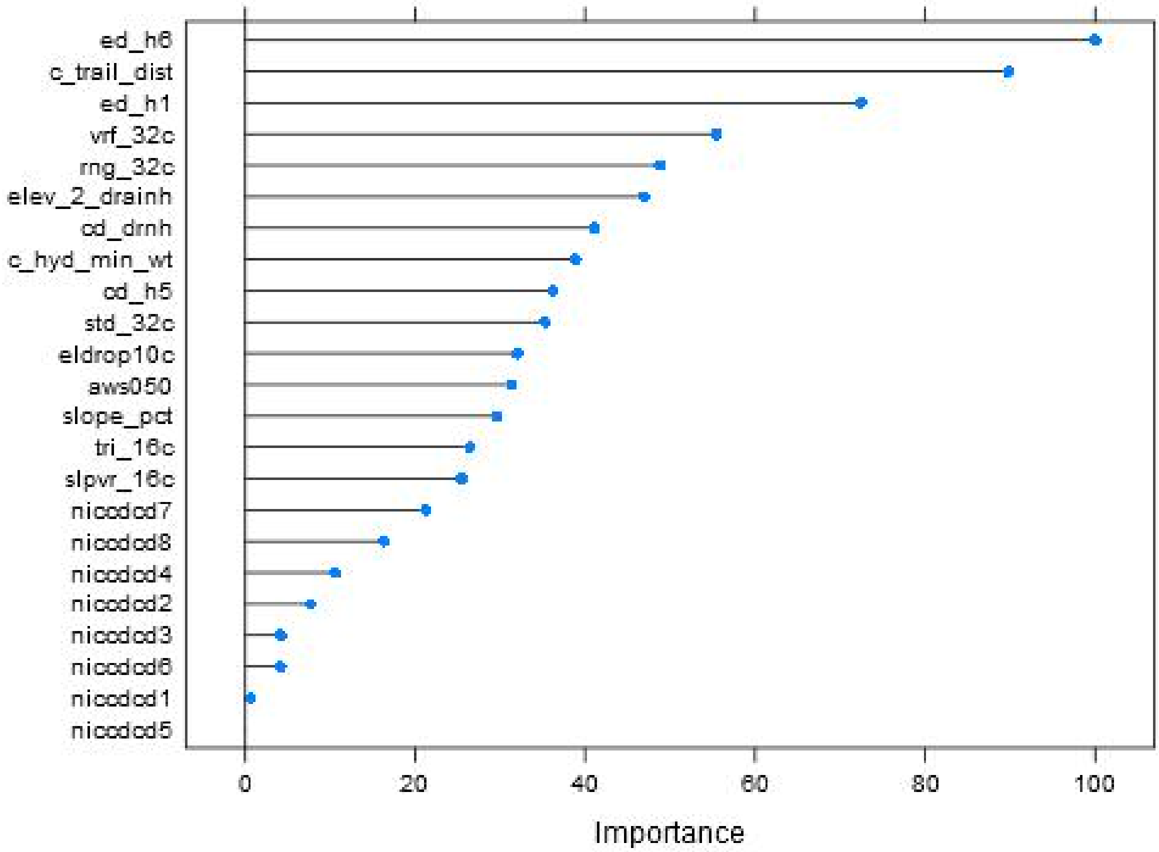


Chart 53. Region 9/10 All – Upland Section 2

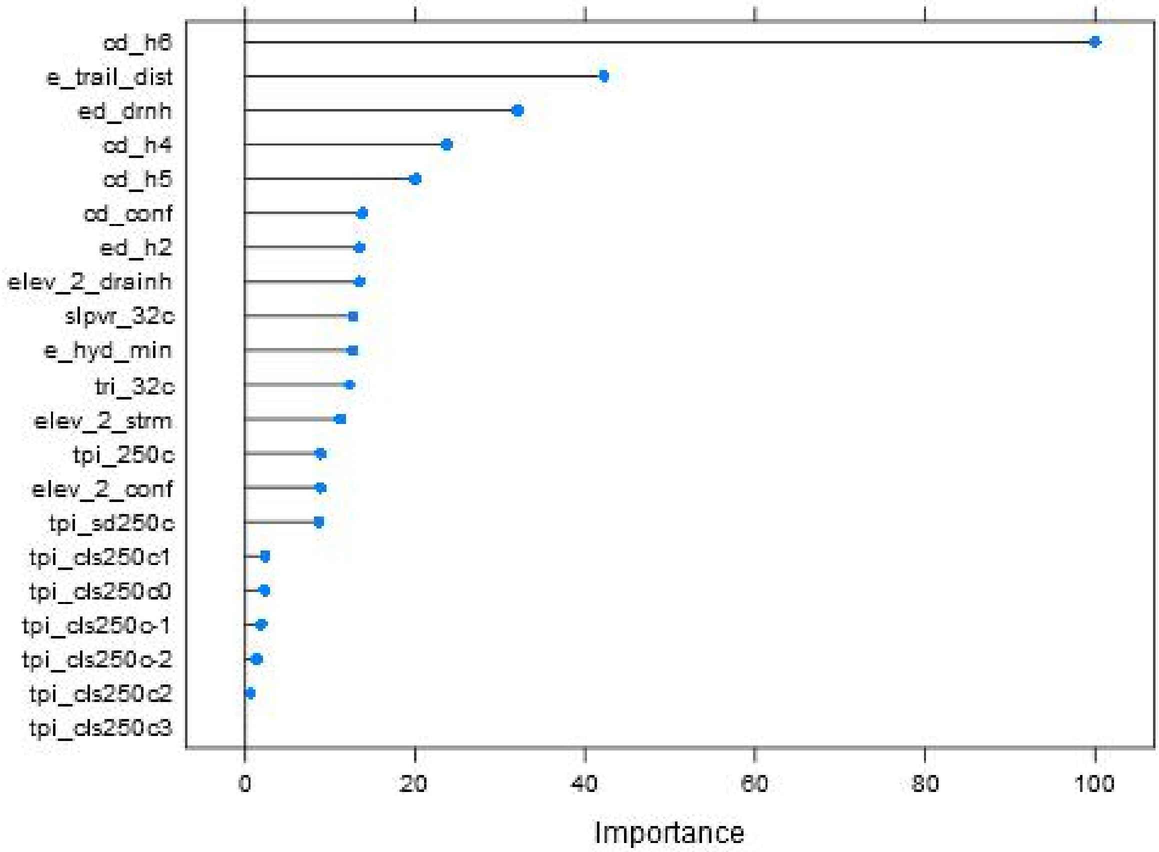


Chart 54. Region 9/10 All – Upland Section 3

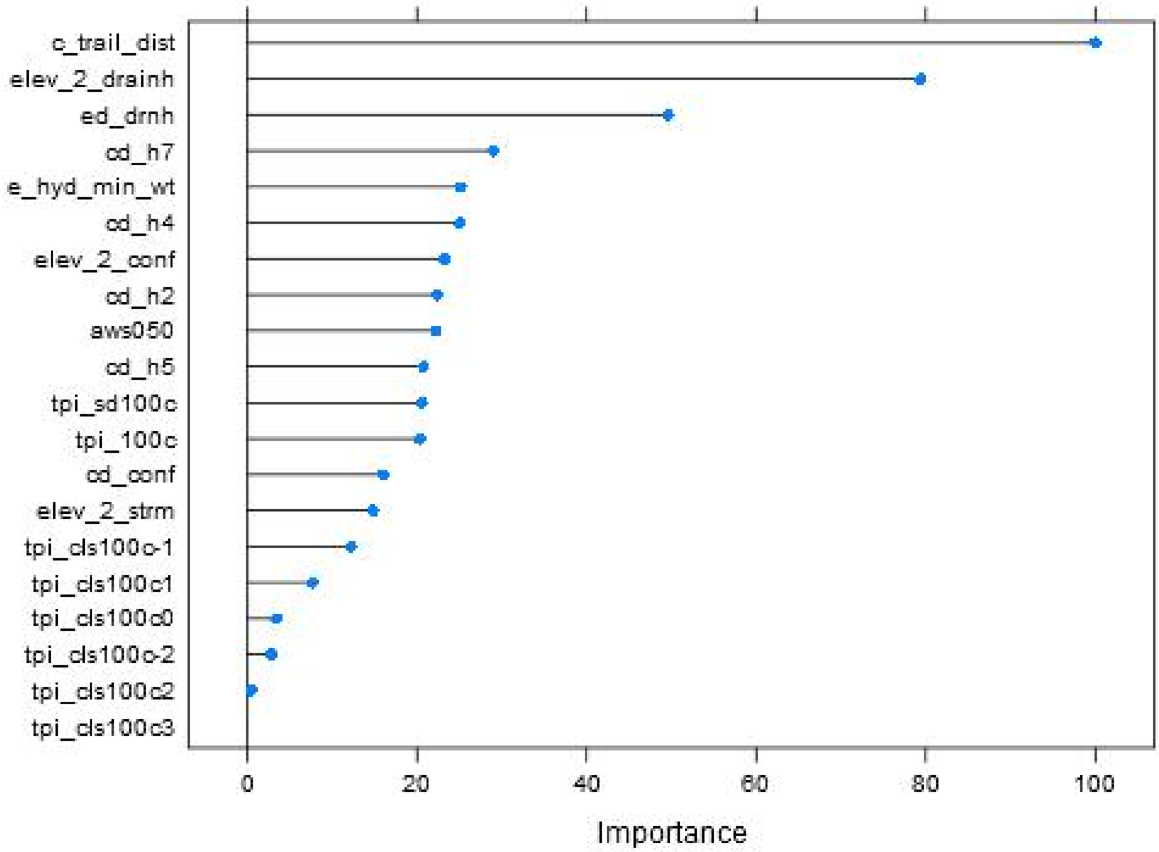


Chart 55. Region 9/10 All – Upland Section 4

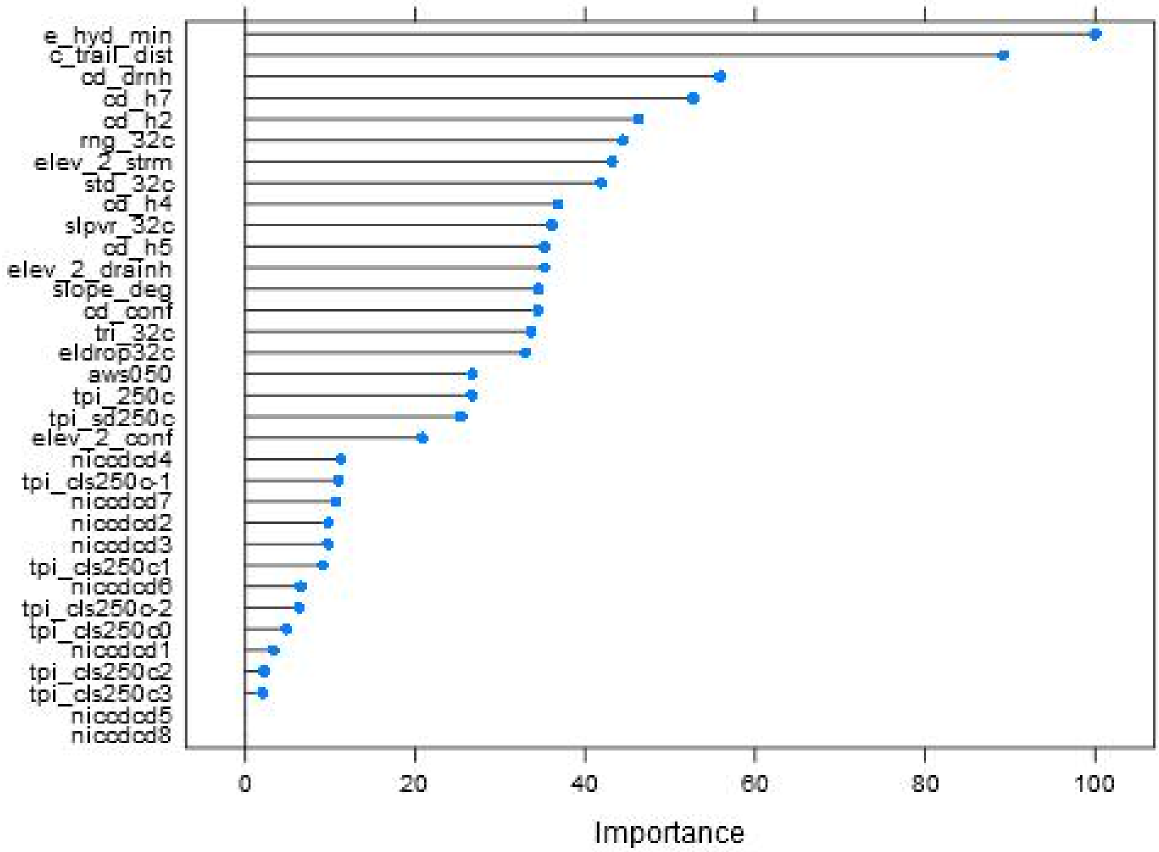


Chart 56. Region 9/10 All – Upland Section 5

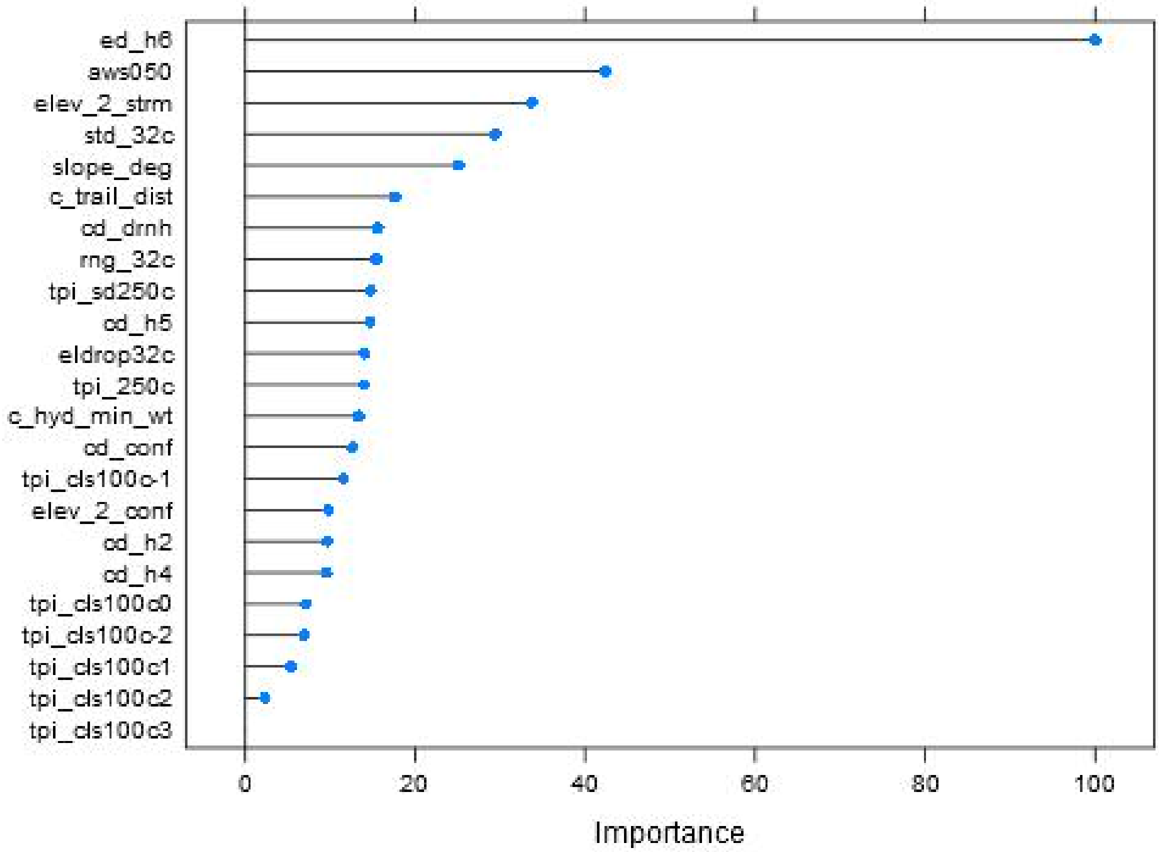


Chart 57. Region 9/10 All – Upland Section 6

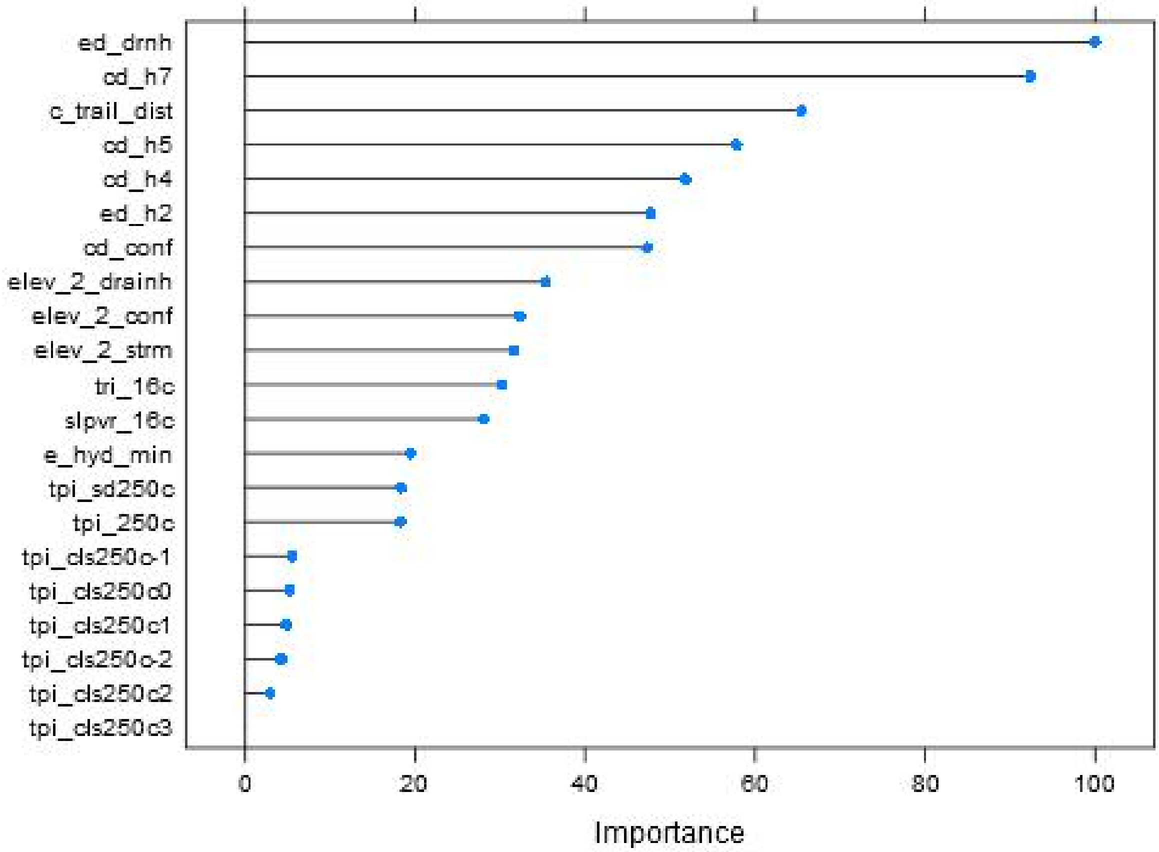


Chart 58. Region 9/10 All – Upland Section 7

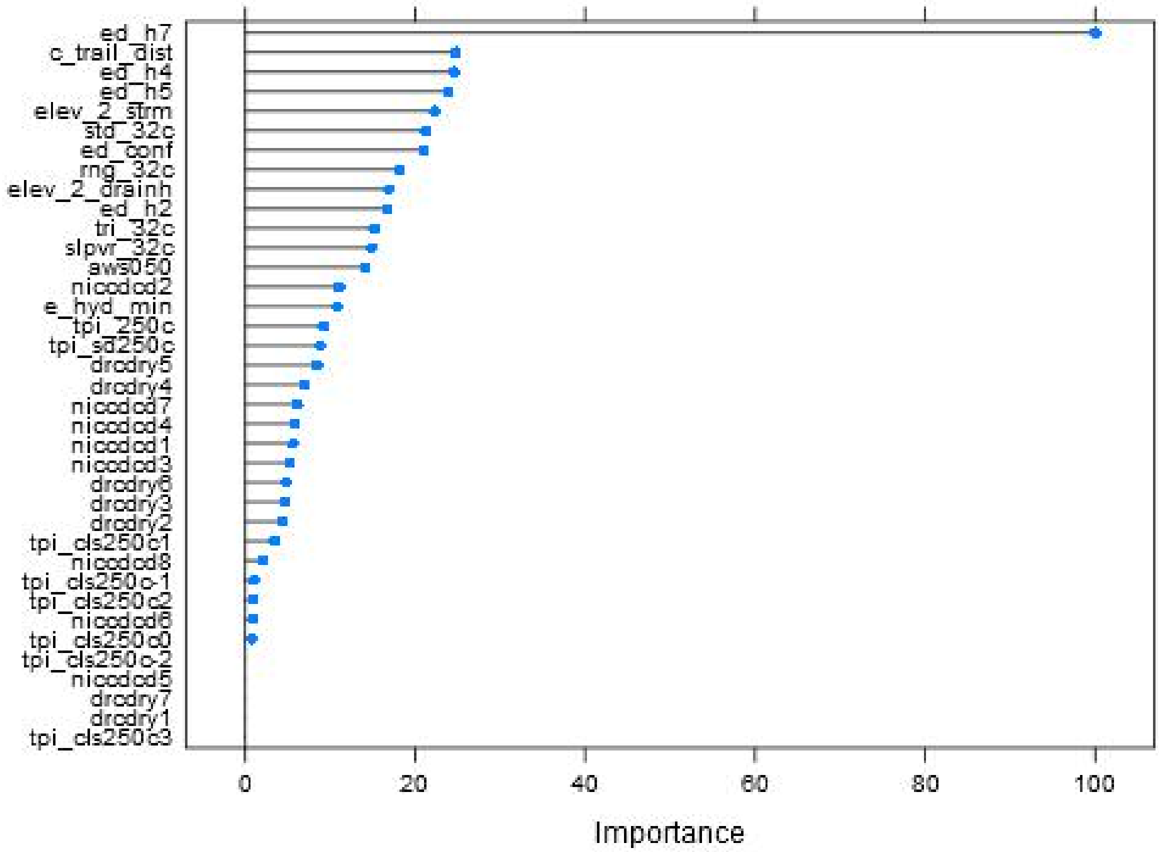


Chart 59. Region 9/10 All – Upland Section 8

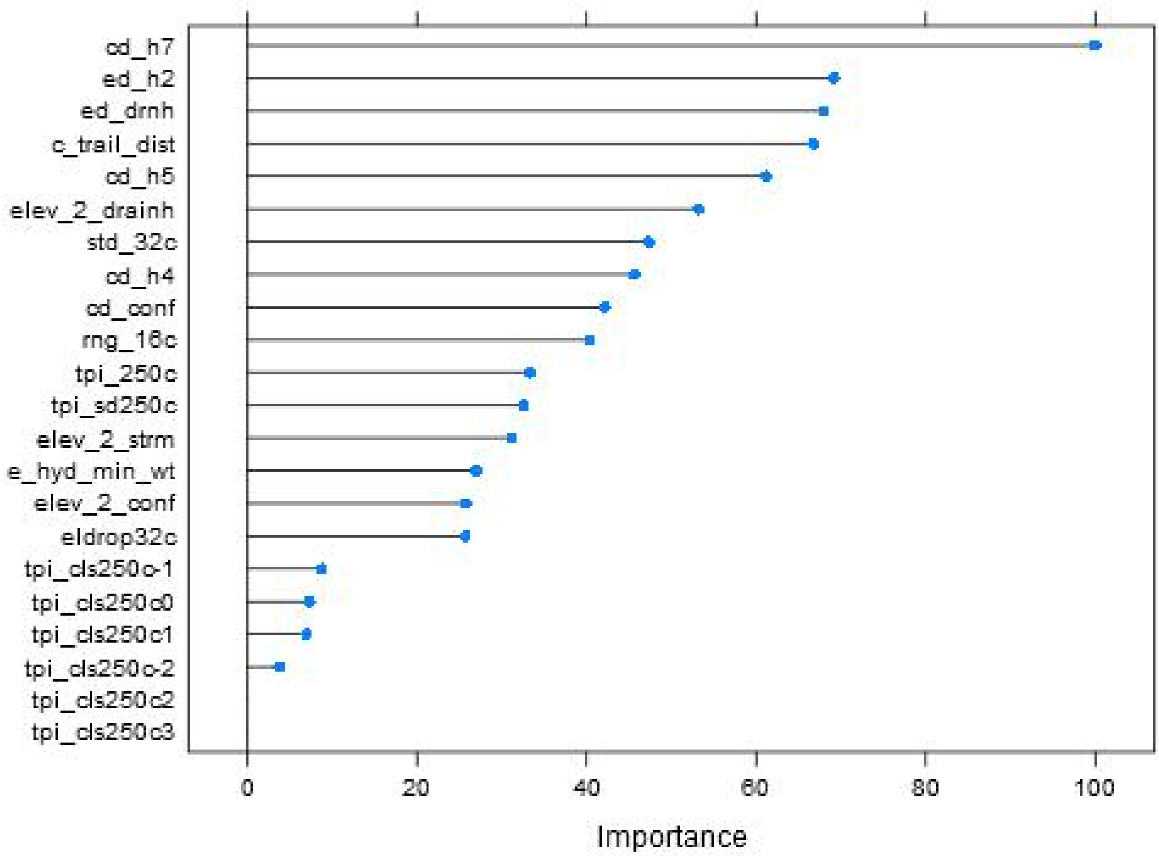


Chart 60. Region 9/10 All – Upland Section 9

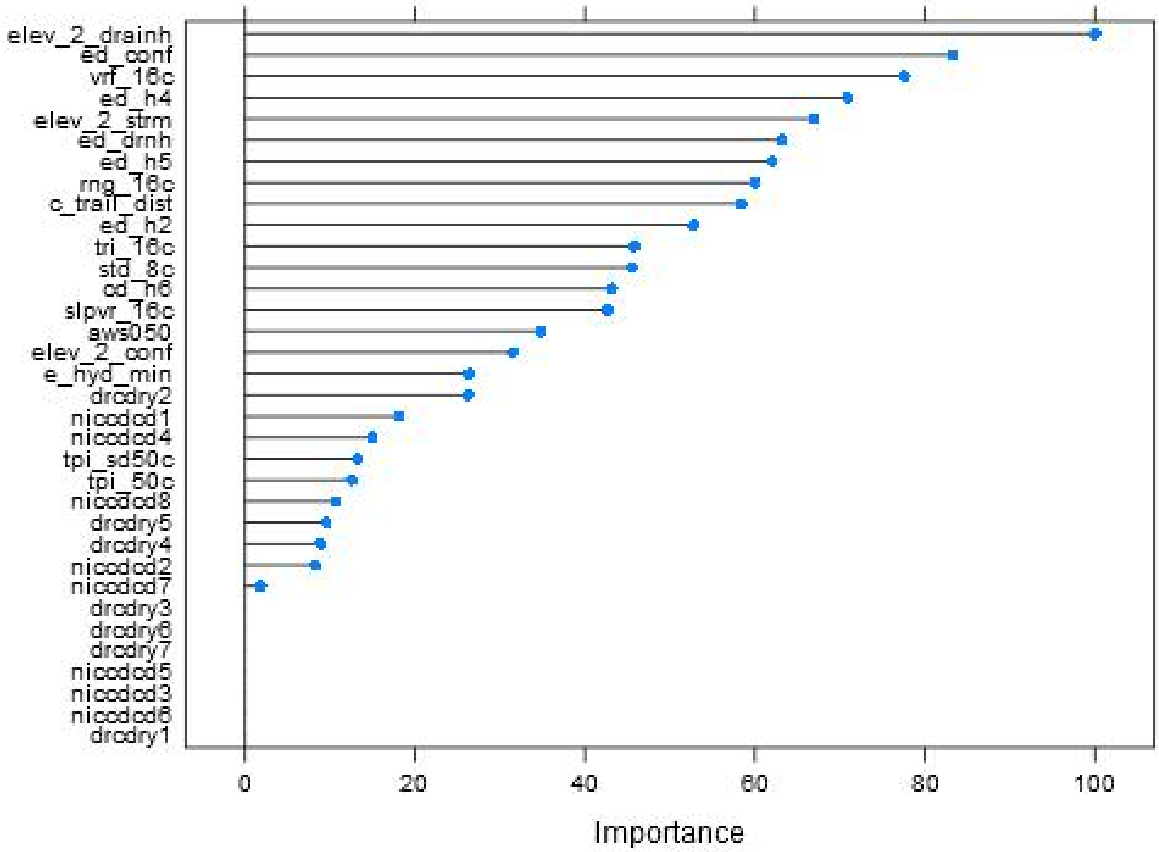


Chart 61. Region 9/10 All – Upland Section 10

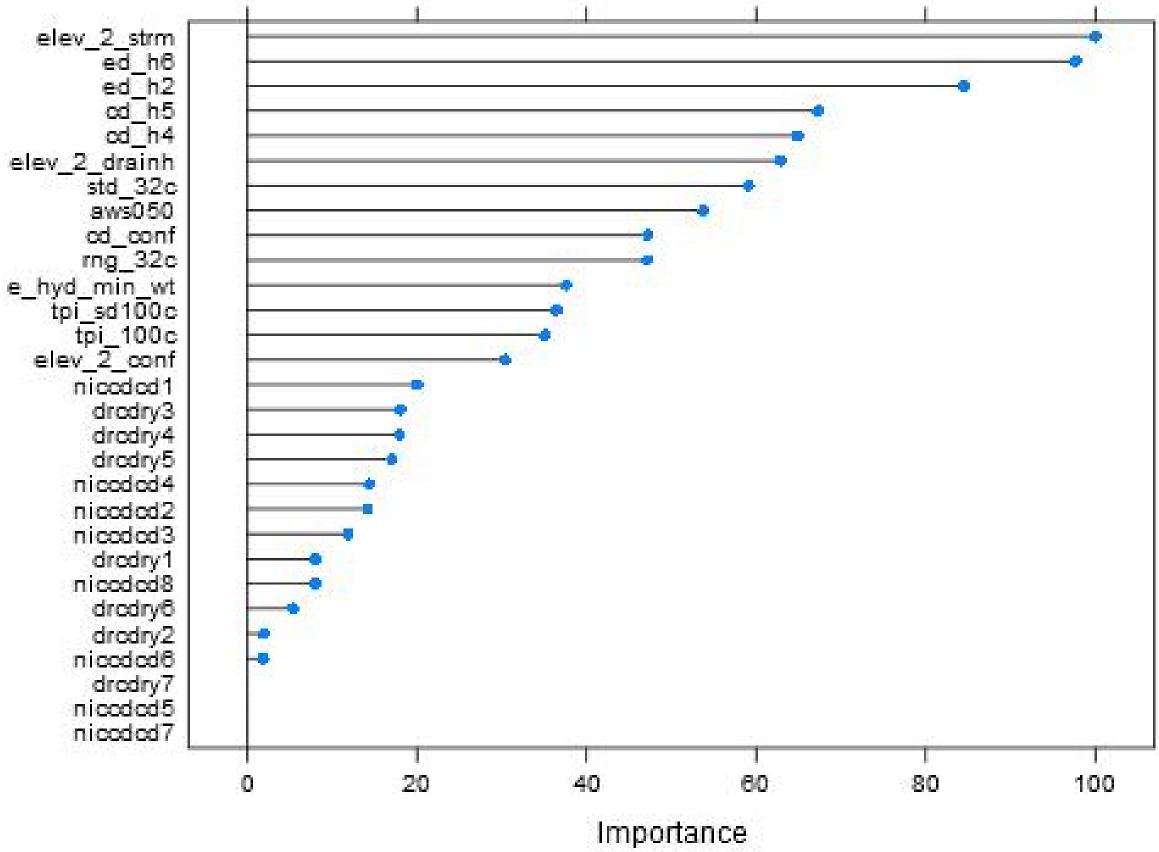


Chart 62. Region 9/10 All – Upland Section 11

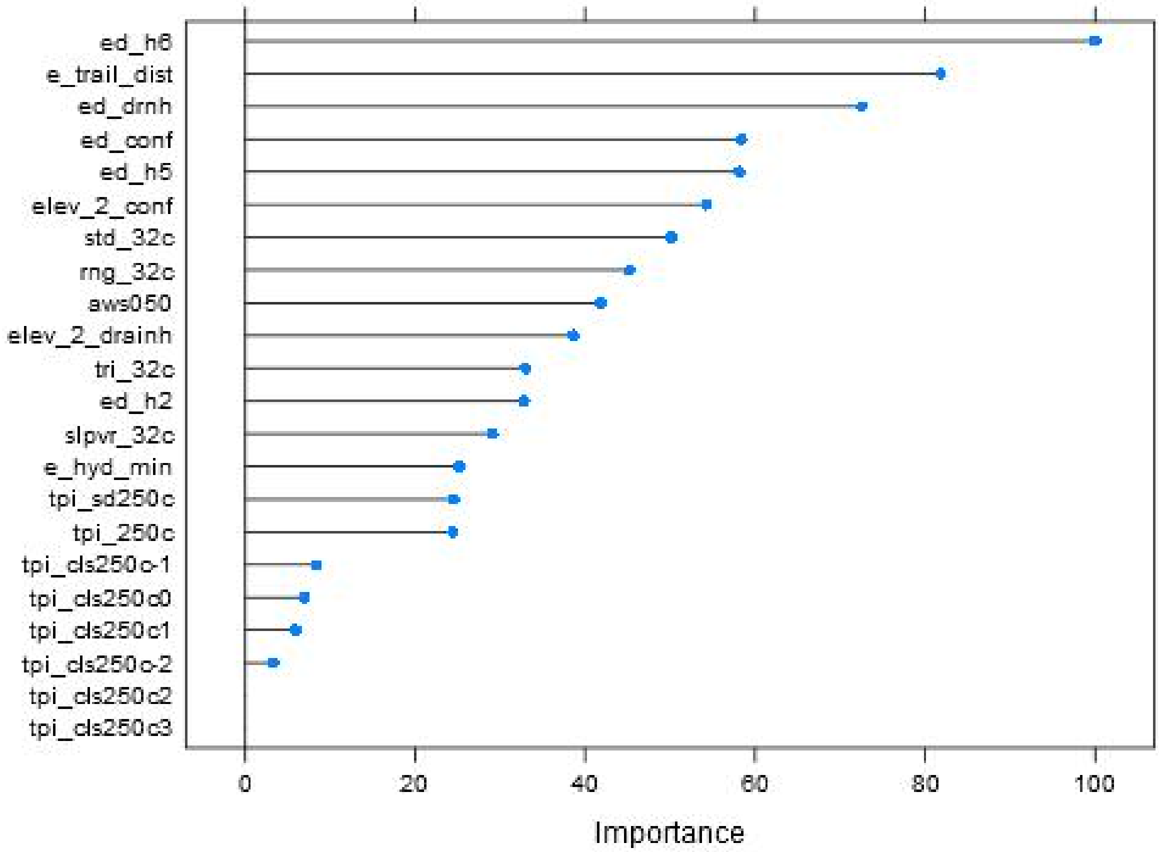


Chart 63. Region 9/10 All – Upland Section 12

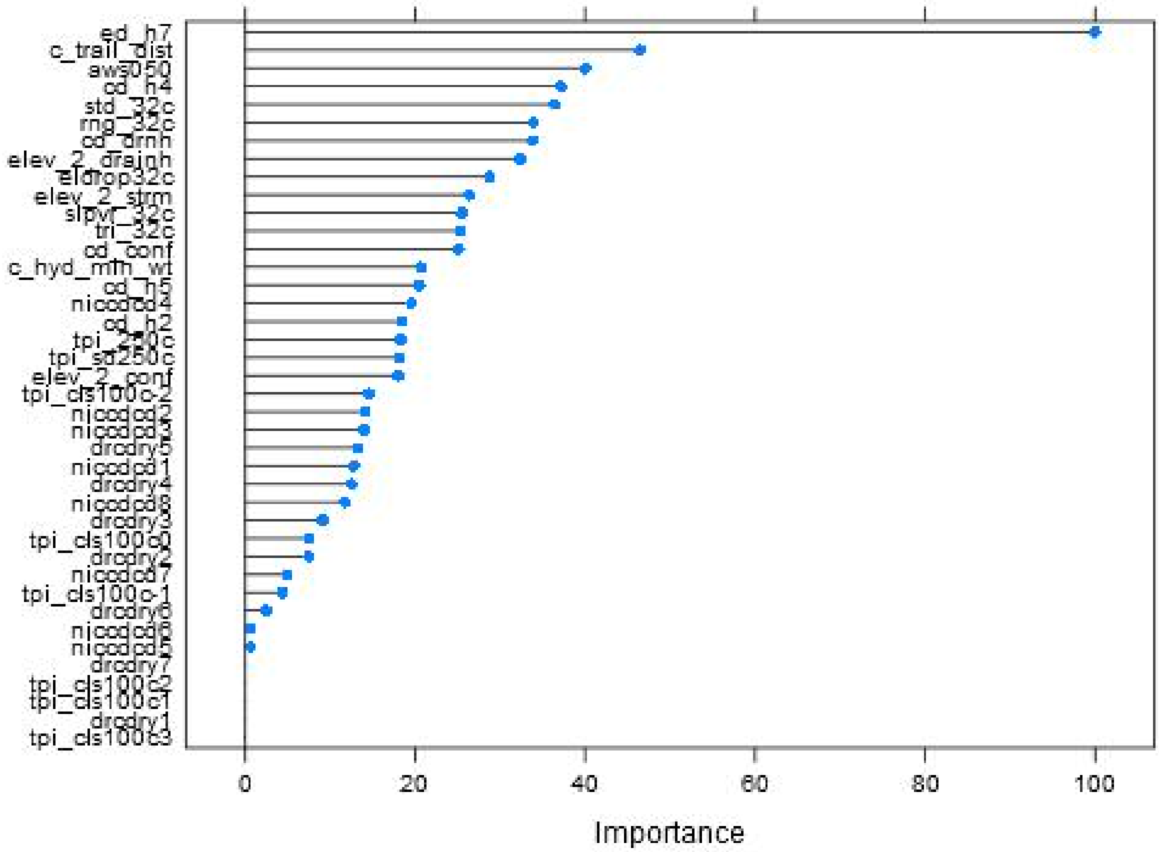


Chart 64. Region 9/10 All – Upland Section 13

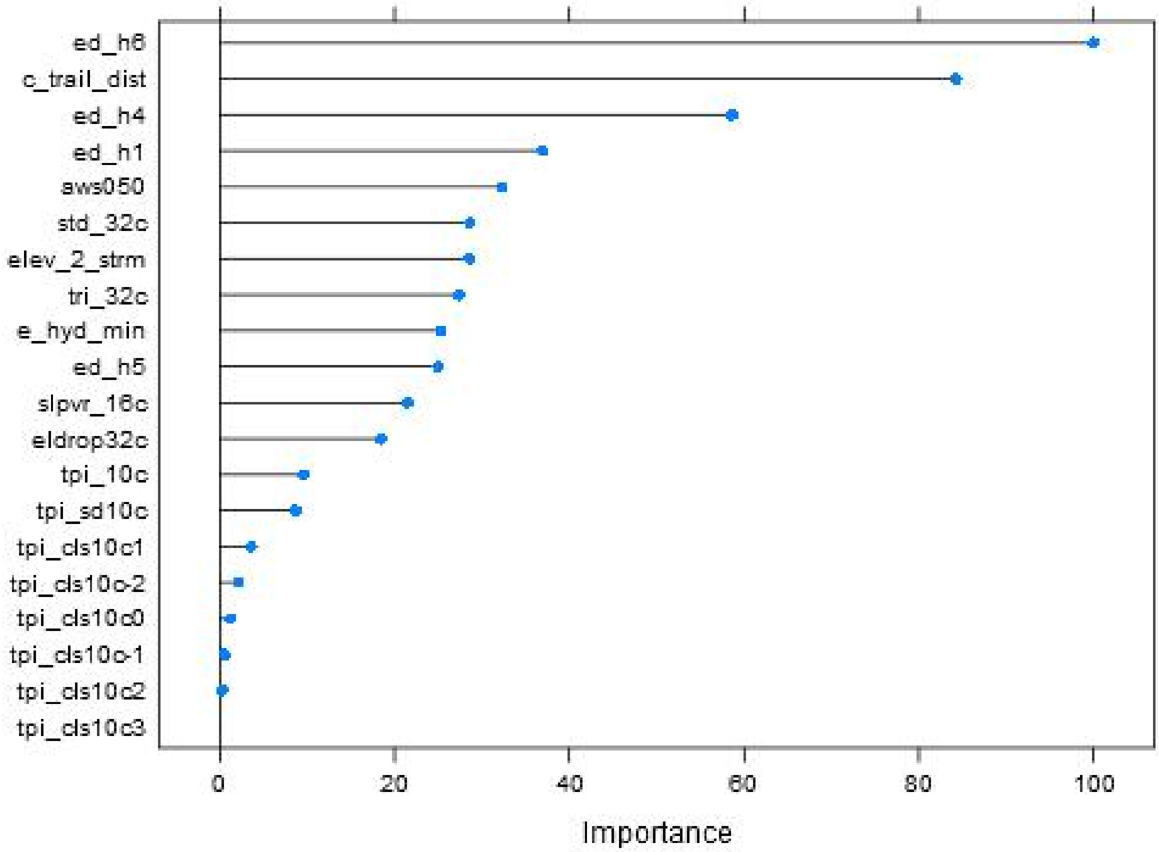


Chart 65. Region 9/10 All – Upland Section 14

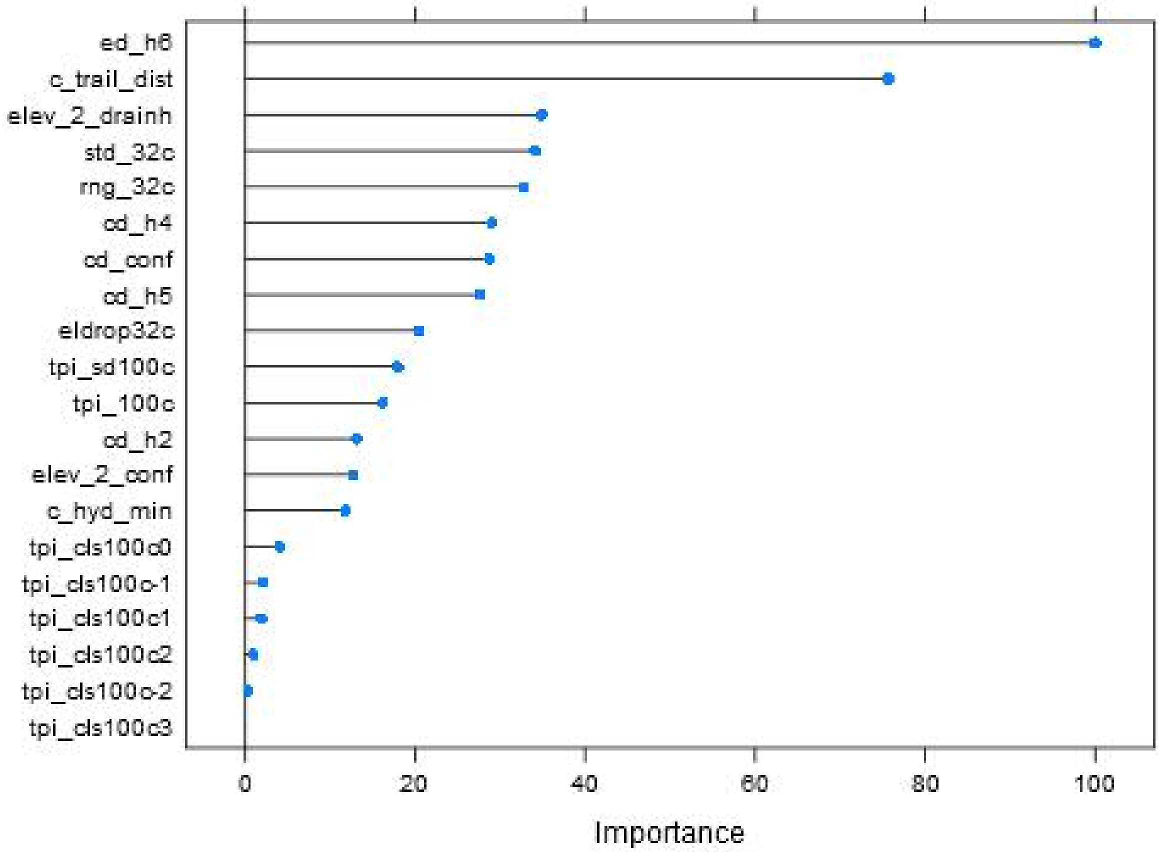
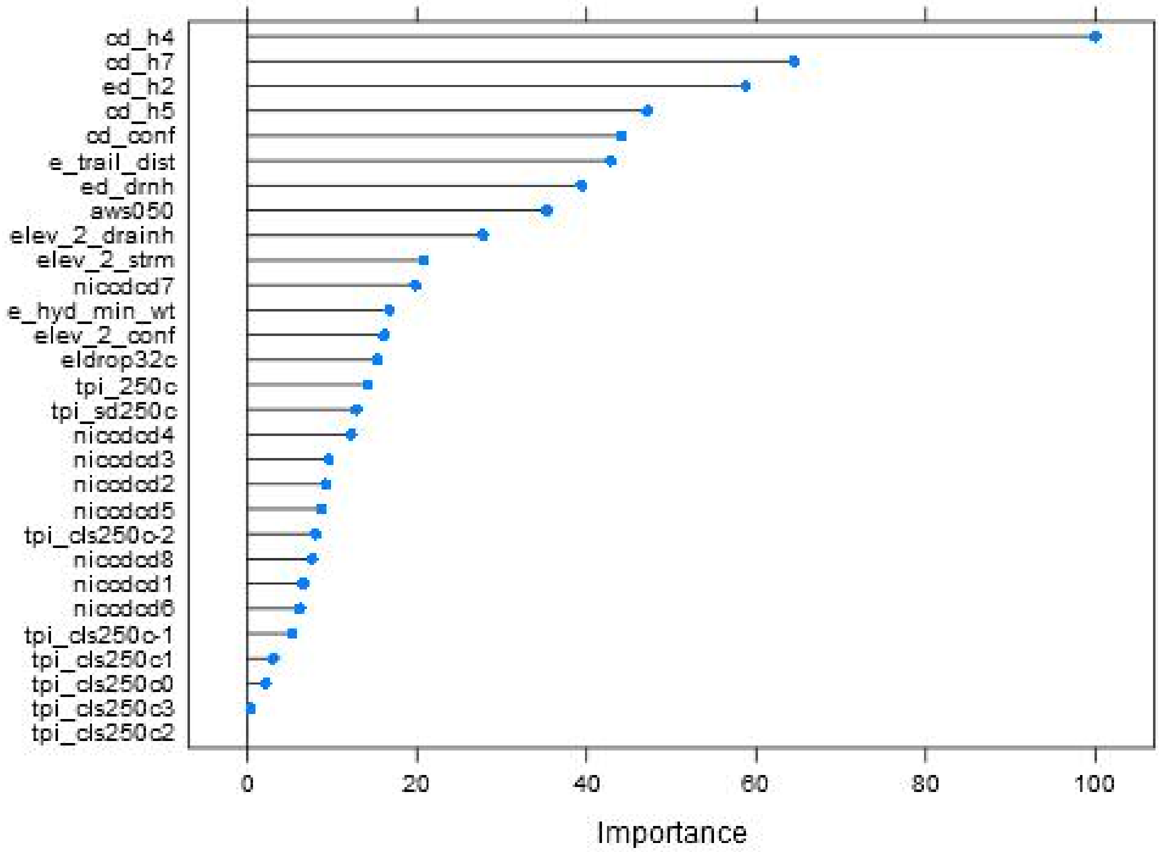


Chart 66. Region 9/10 All – Upland Section 15



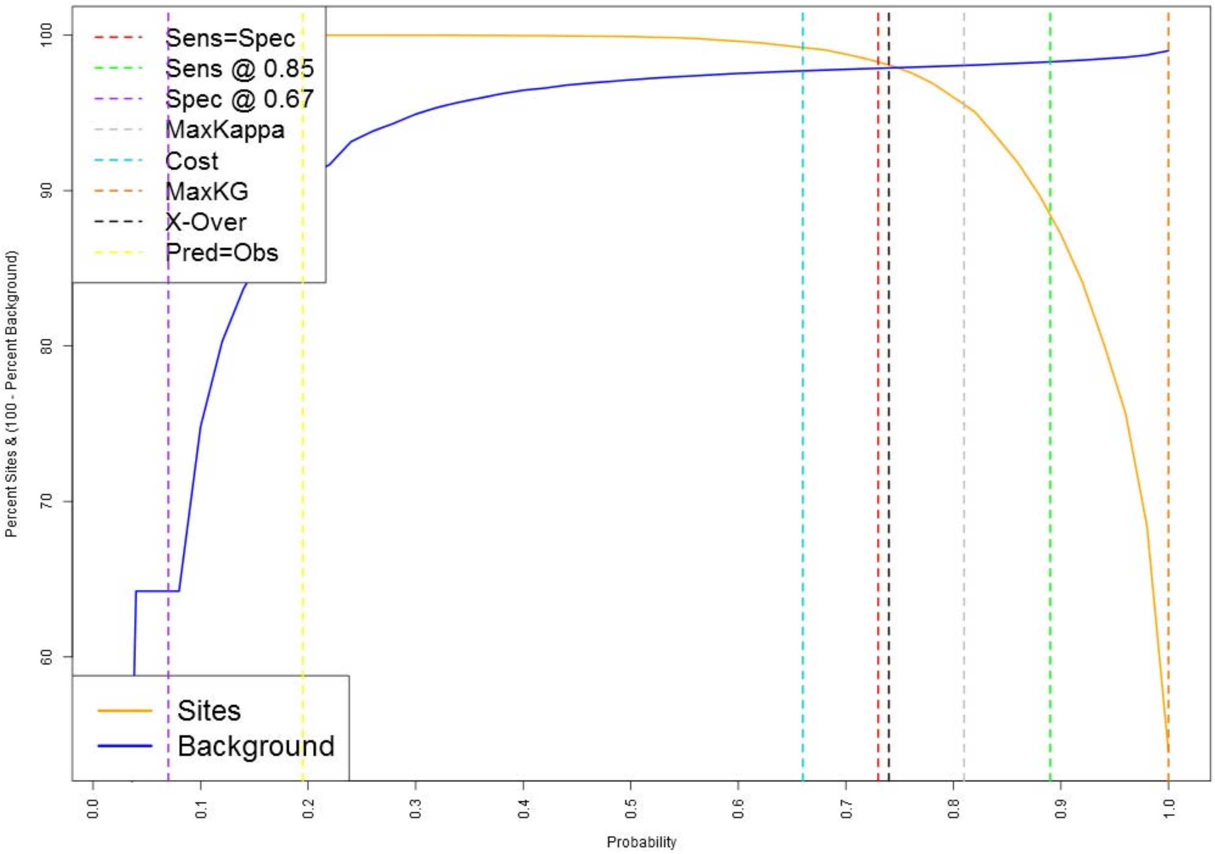
APPENDIX F

POTENTIAL THRESHOLDS

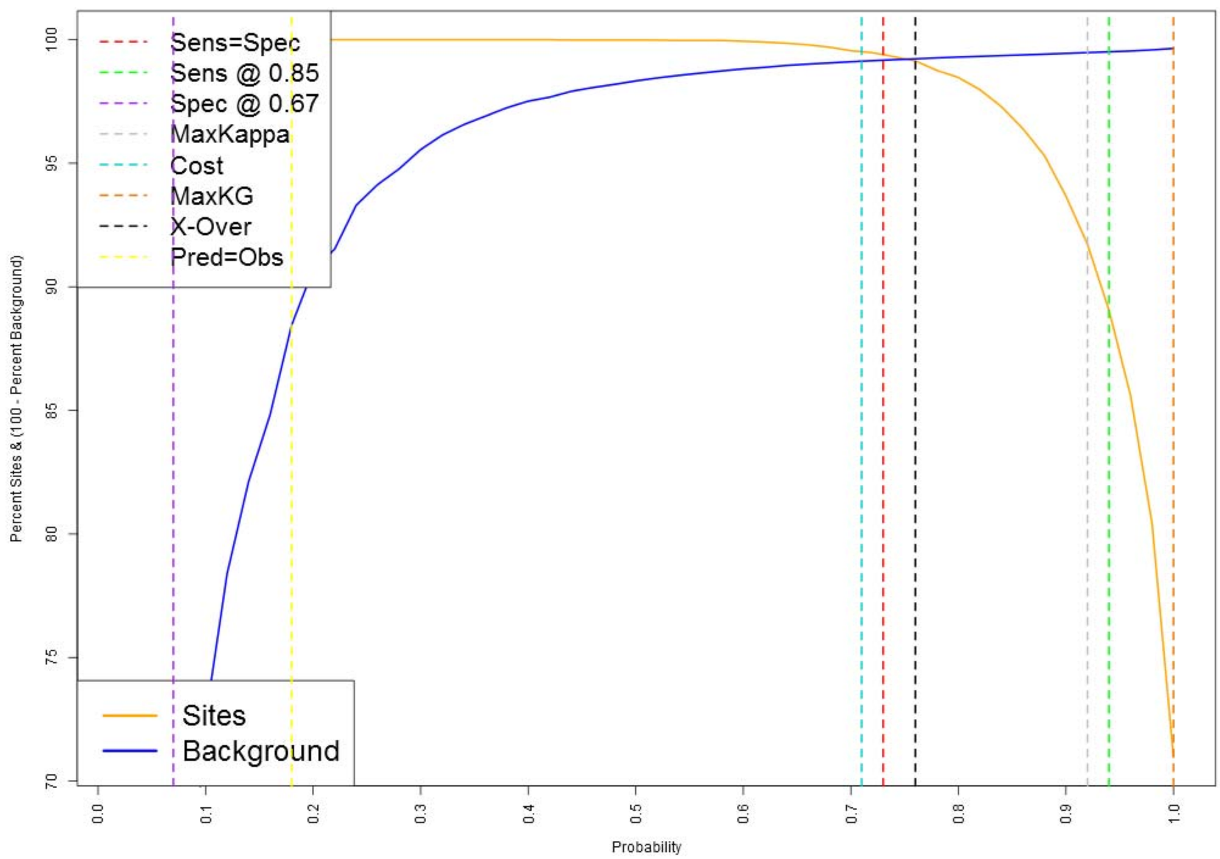
FOR EACH OF 66 MODELS

WITHIN REGIONS 7, 8, AND 9/10

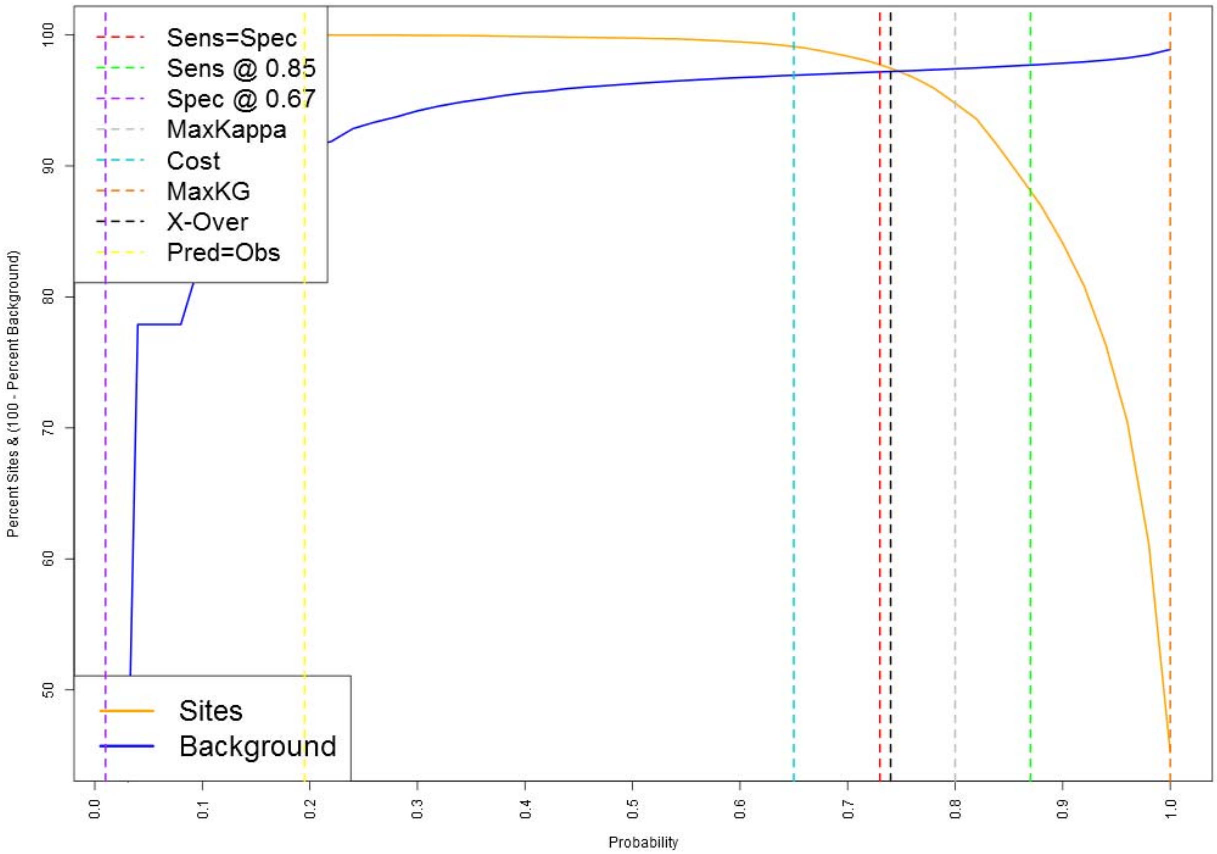
Region 7 All – Riverine Section 1



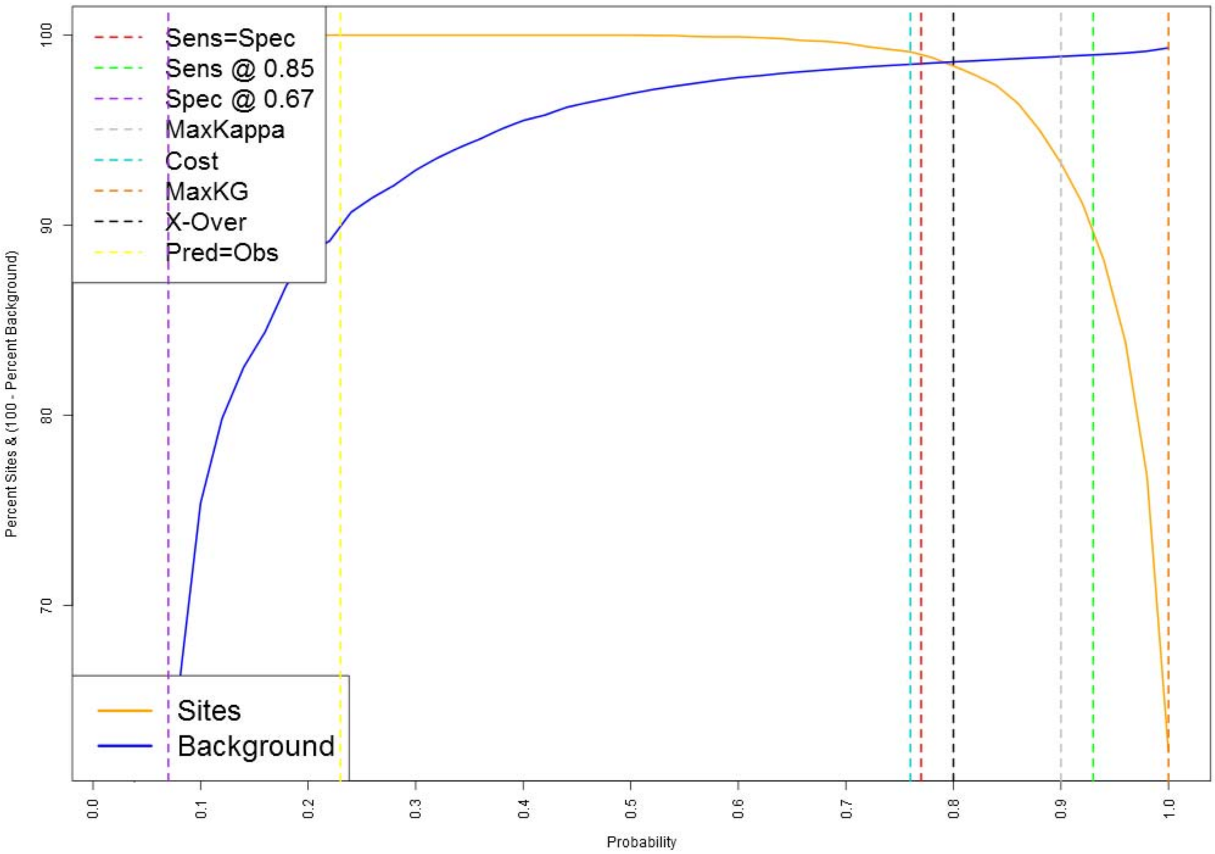
Region 7 All – Riverine Section 2



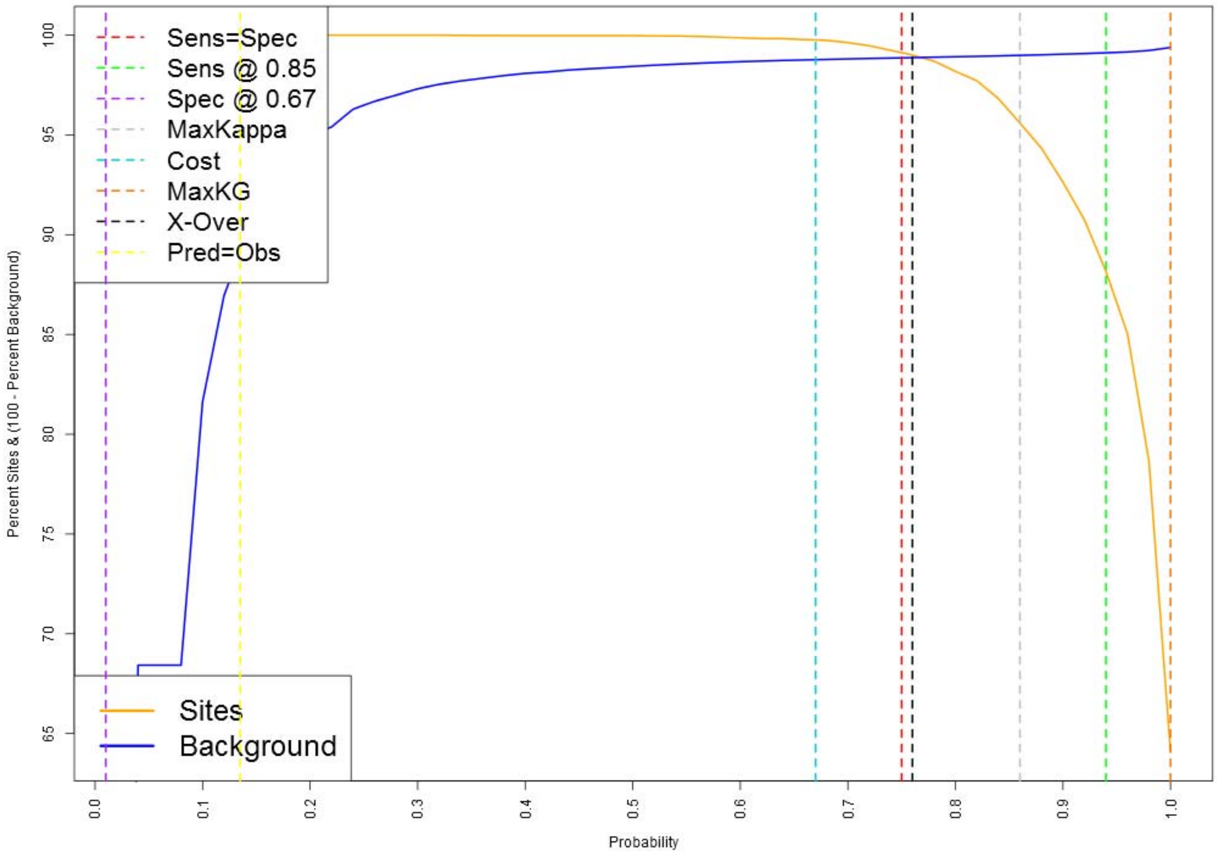
Region 7 All – Riverine Section 3



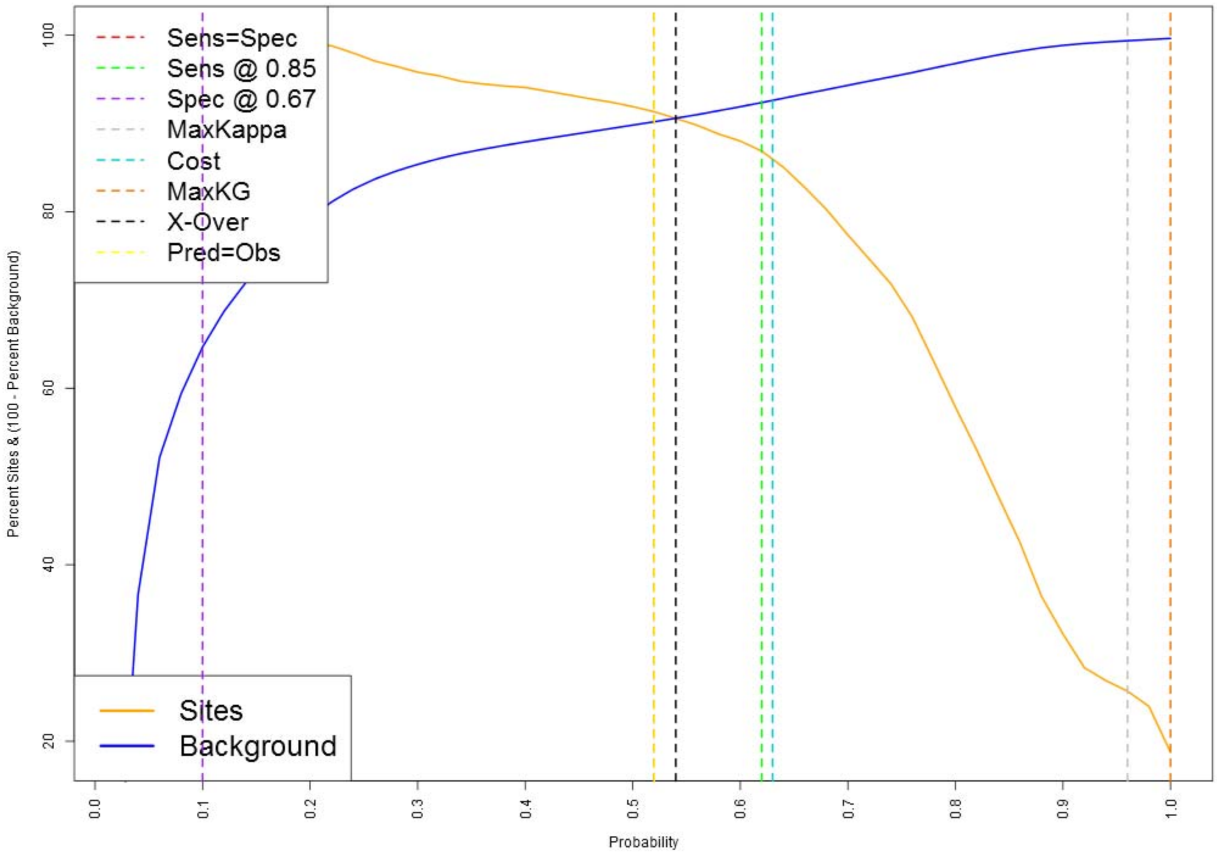
Region 7 All – Riverine Section 4



Region 7 All – Riverine Section 5

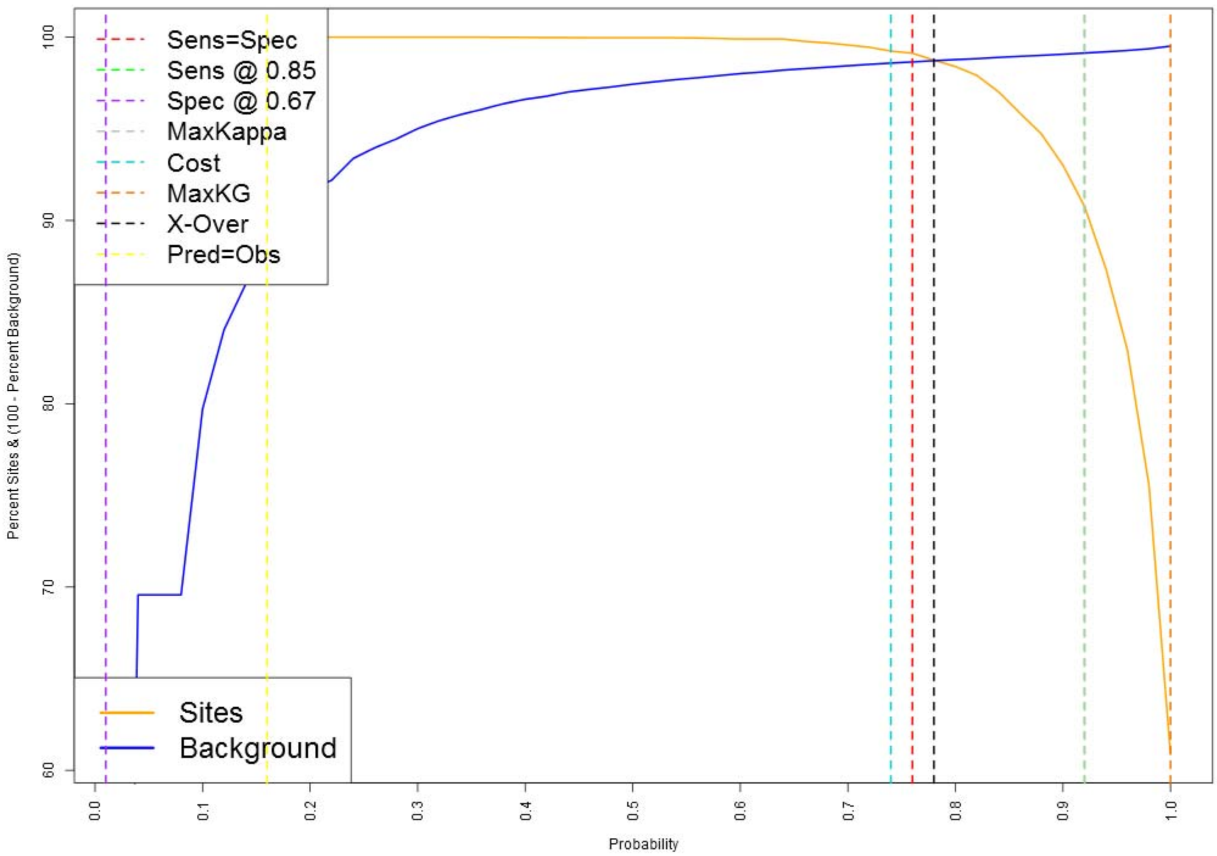


Region 7 All – Riverine Section 6



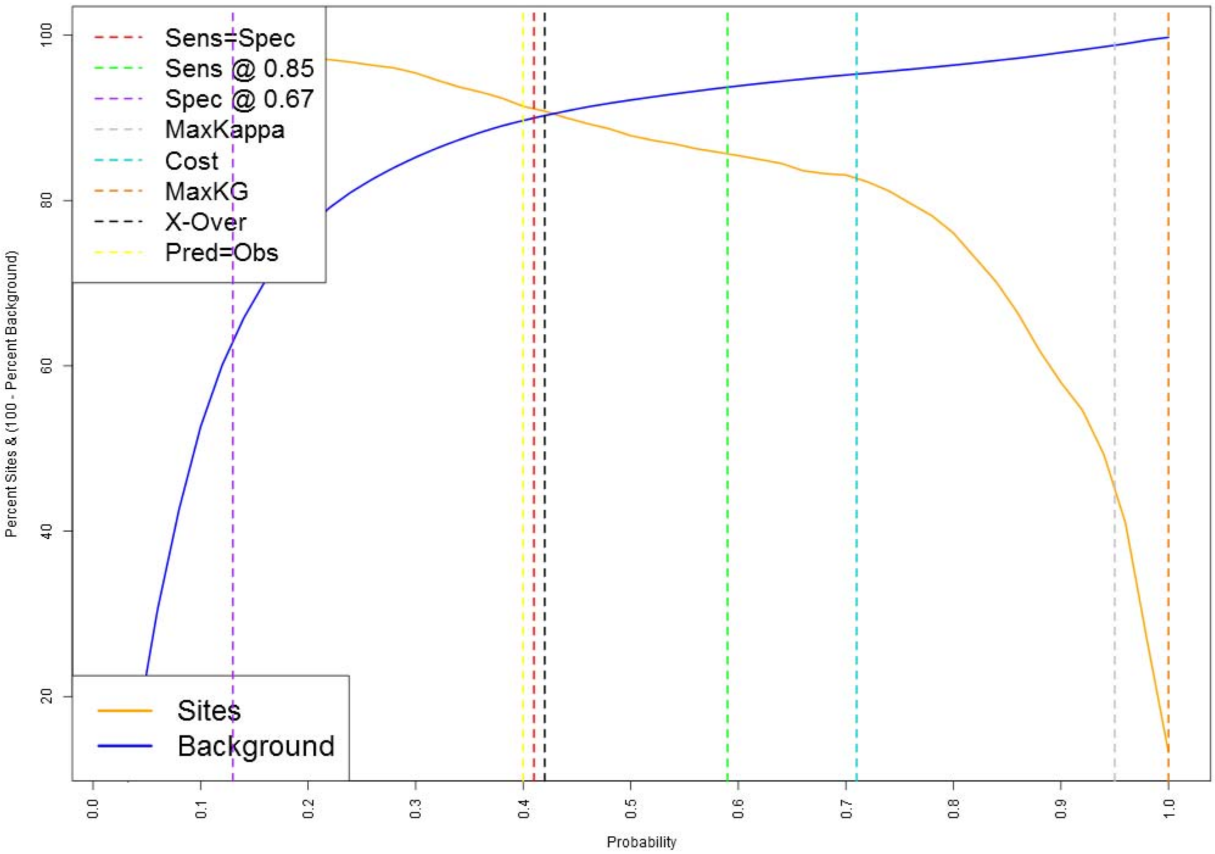
Note: The line for Sens=Spec is underneath the line for Pred=Obs and is not visible because the values are identical (0.52).

Region 7 All – Riverine Section 7

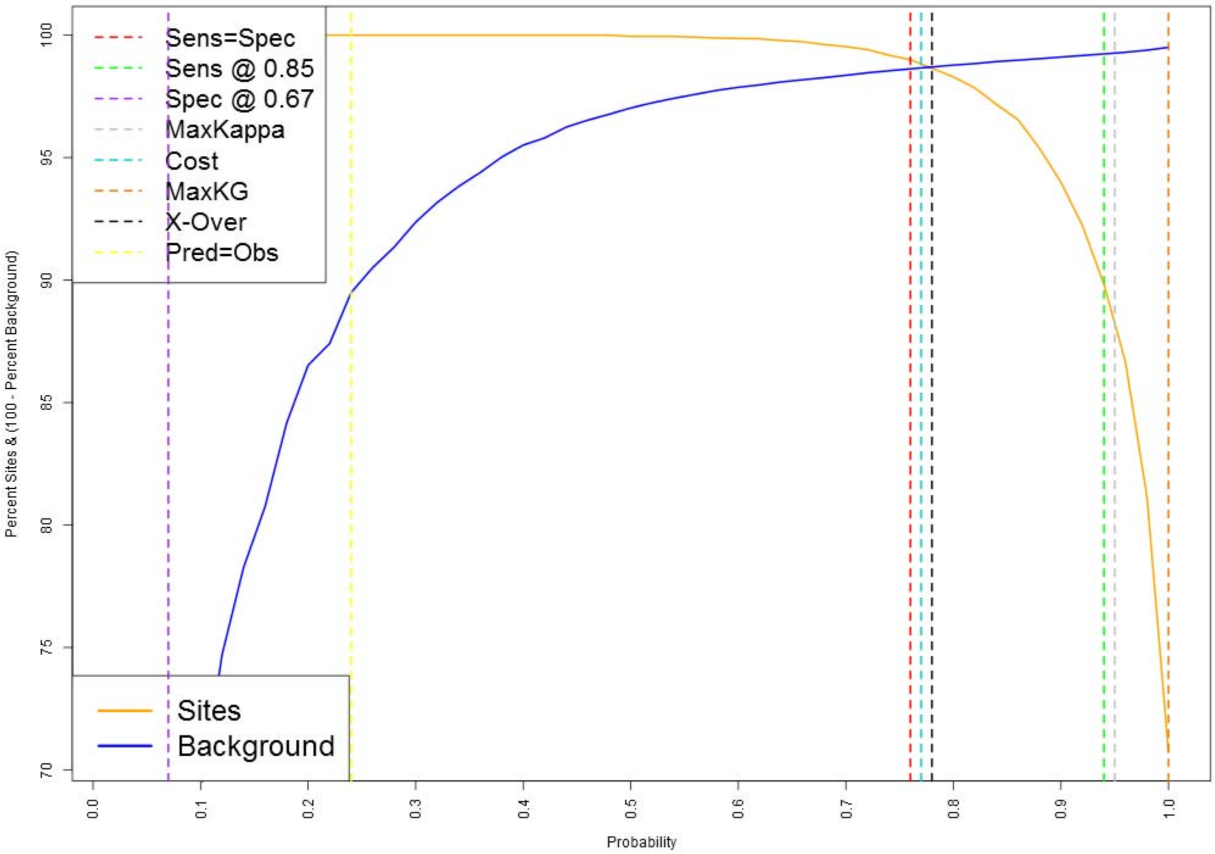


Note: The line for Sens @ 0.85 is underneath the line for MaxKappa and is not visible because the values are identical (0.92).

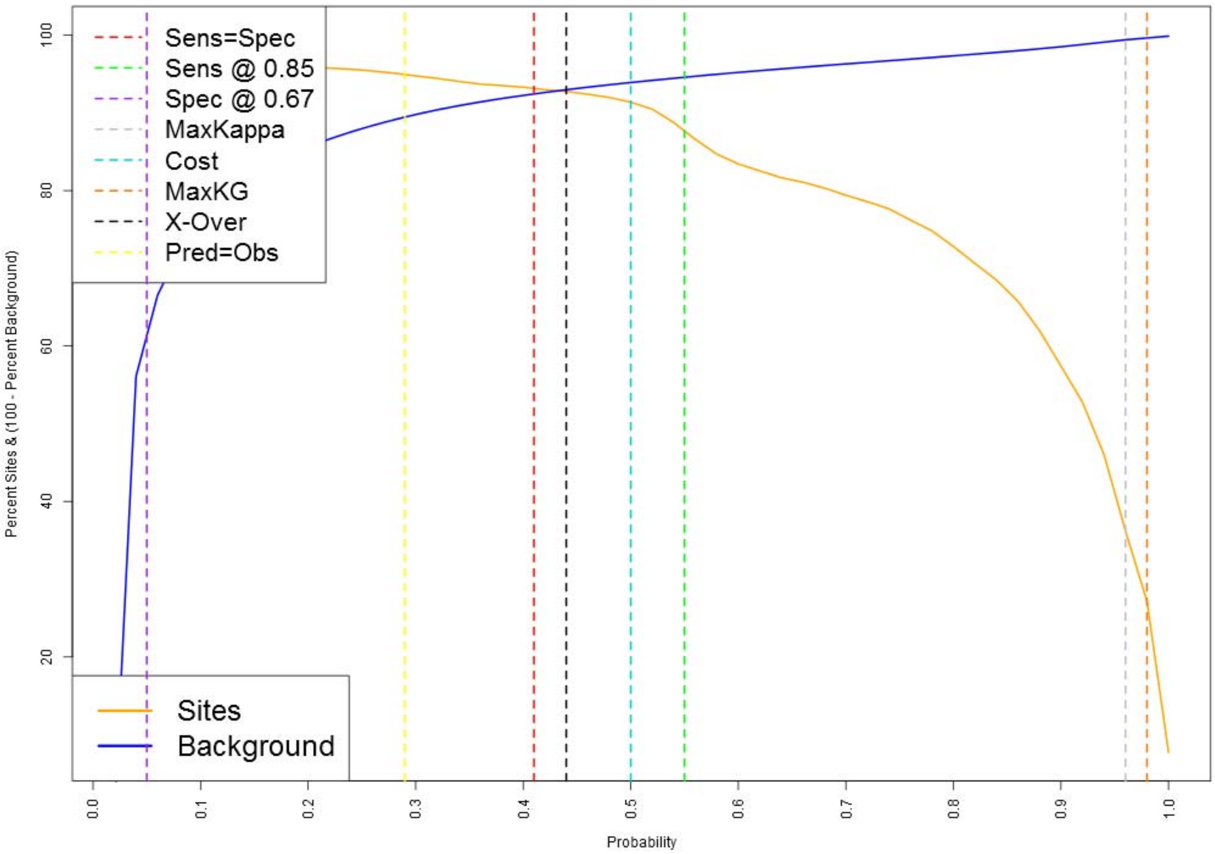
Region 7 All – Riverine Section 8



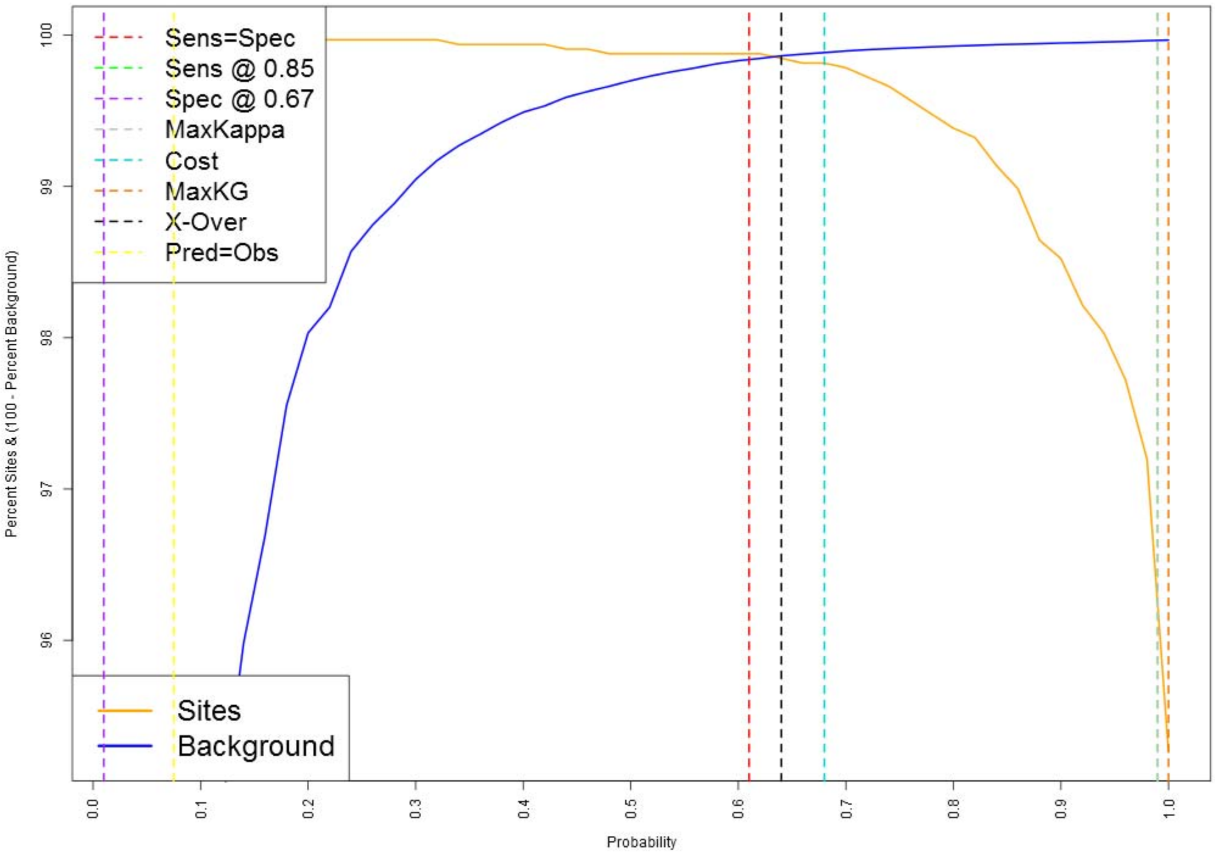
Region 7 All – Riverine Section 9



Region 7 All – Upland Section 1

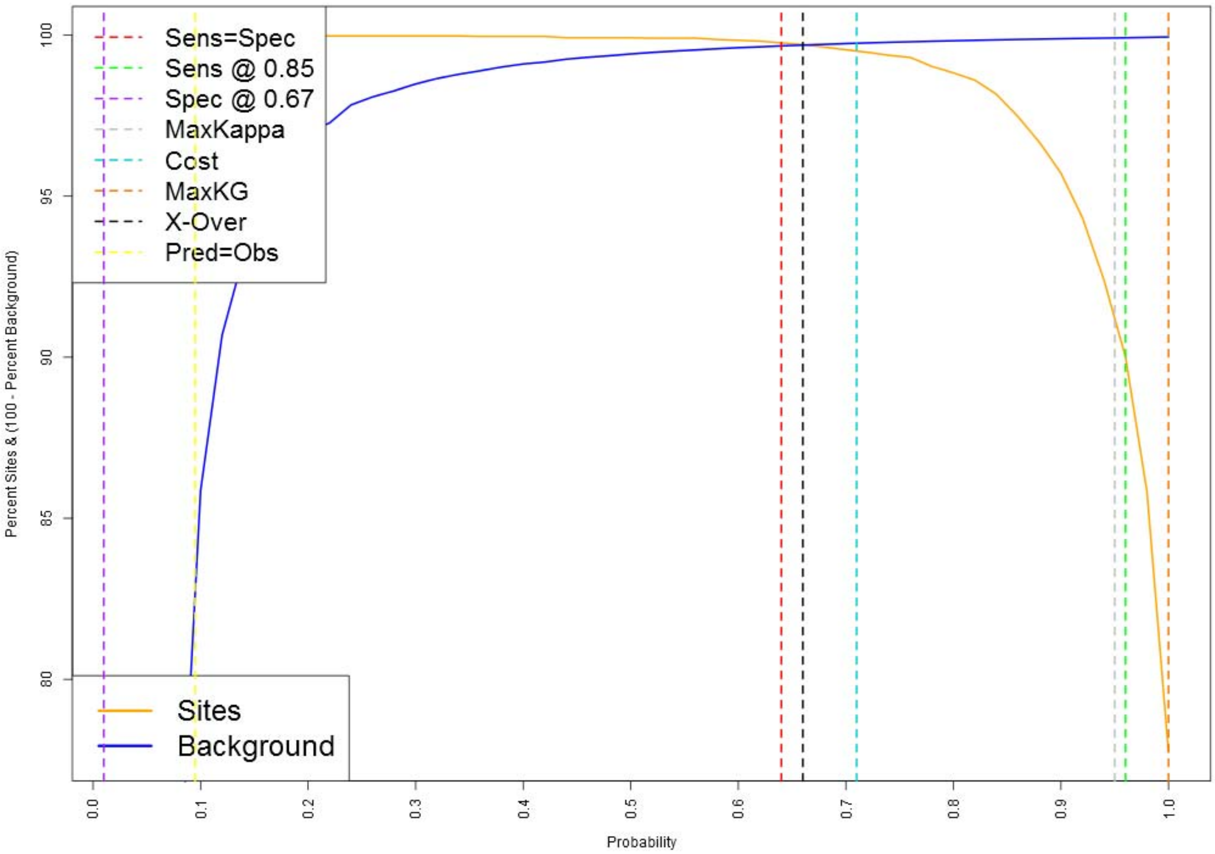


Region 7 All – Upland Section 2

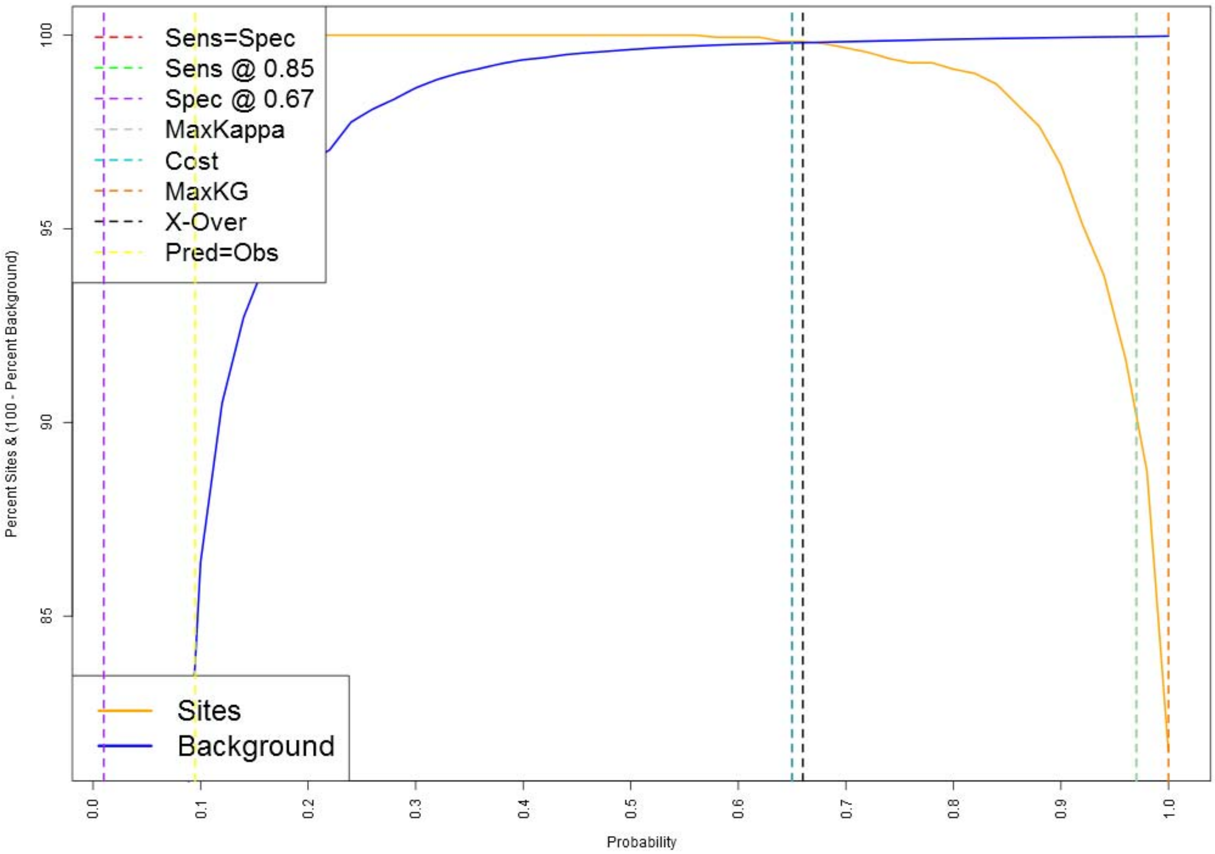


Note: The line for Sens @ 0.85 is underneath the line for MaxKappa and is not visible because the values are identical (0.99).

Region 7 All – Upland Section 3

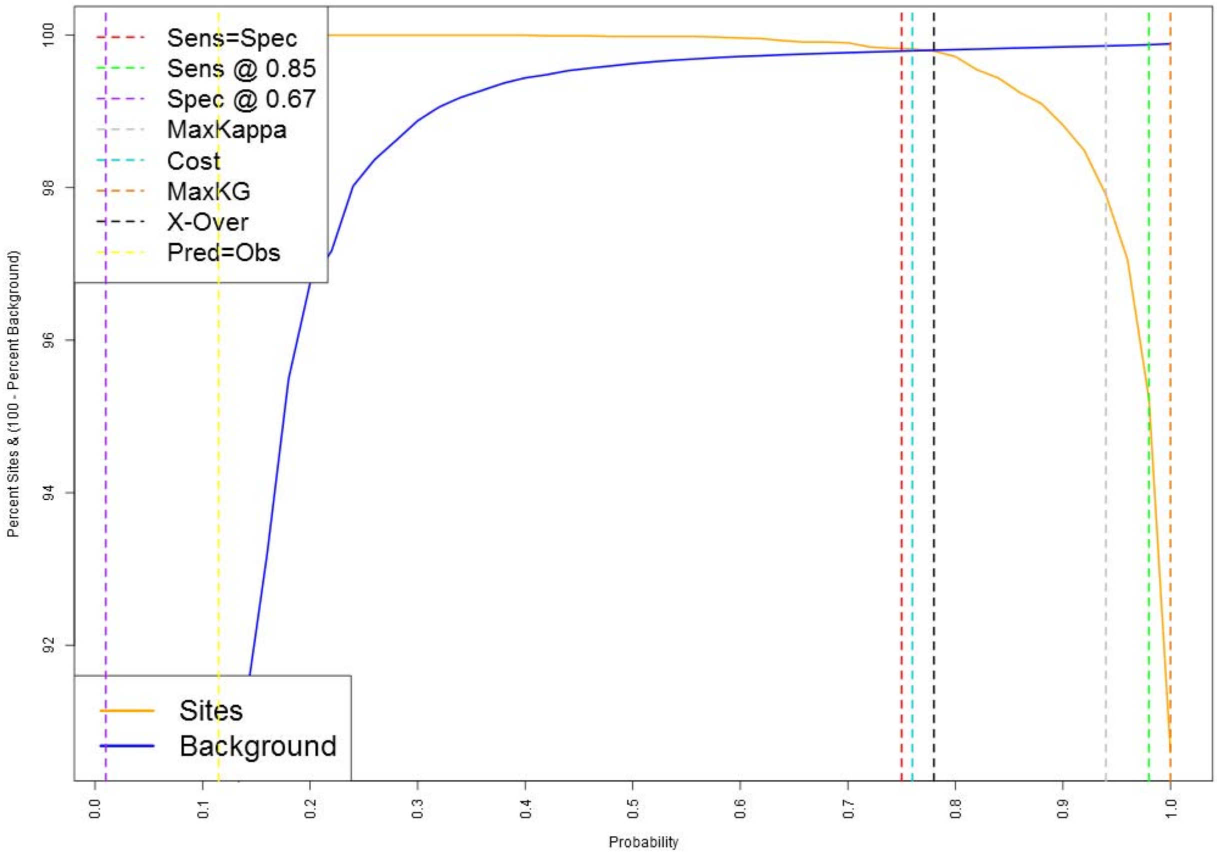


Region 7 All – Upland Section 4

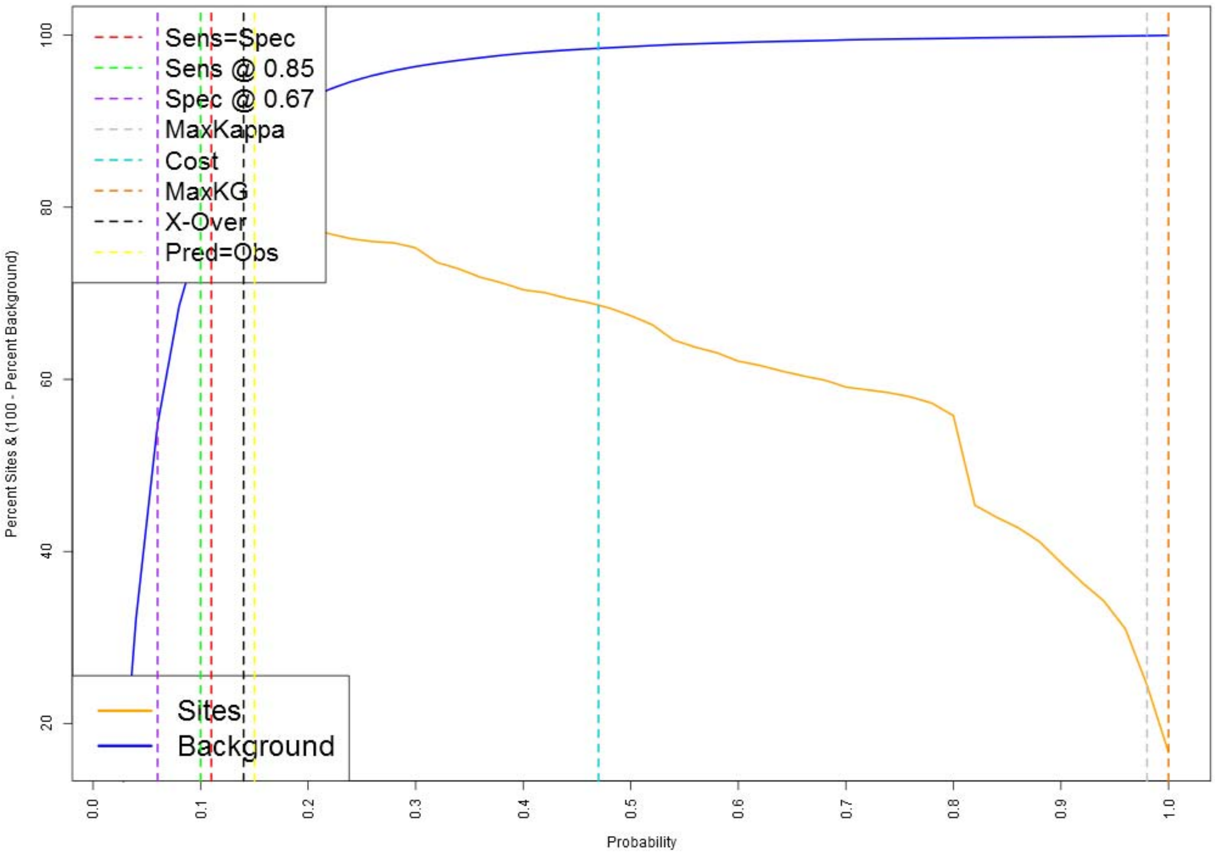


Note: The line for Sens @ 0.85 is underneath the line for MaxKappa and is not visible because the values are identical (0.97); similarly, the line for Sens=Spec is obscured by the line for Cost (0.65).

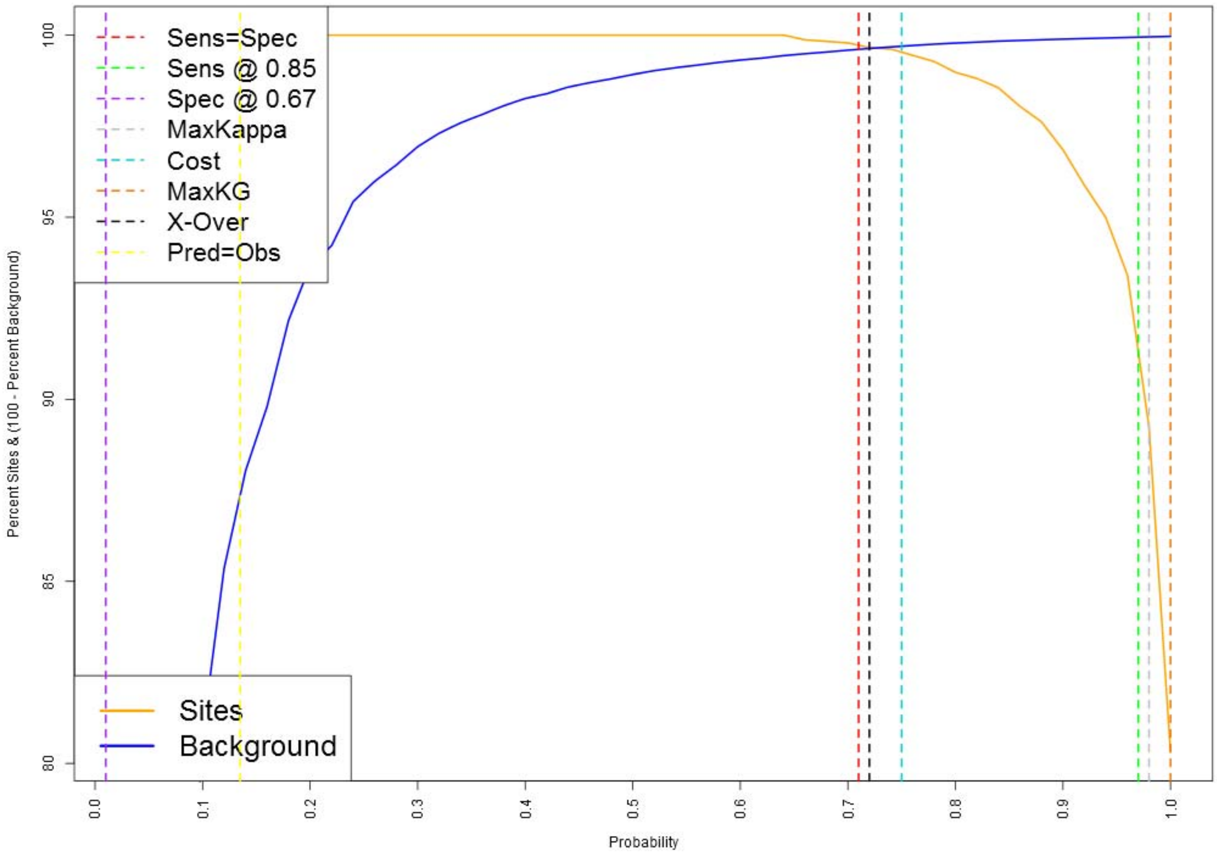
Region 7 All – Upland Section 5



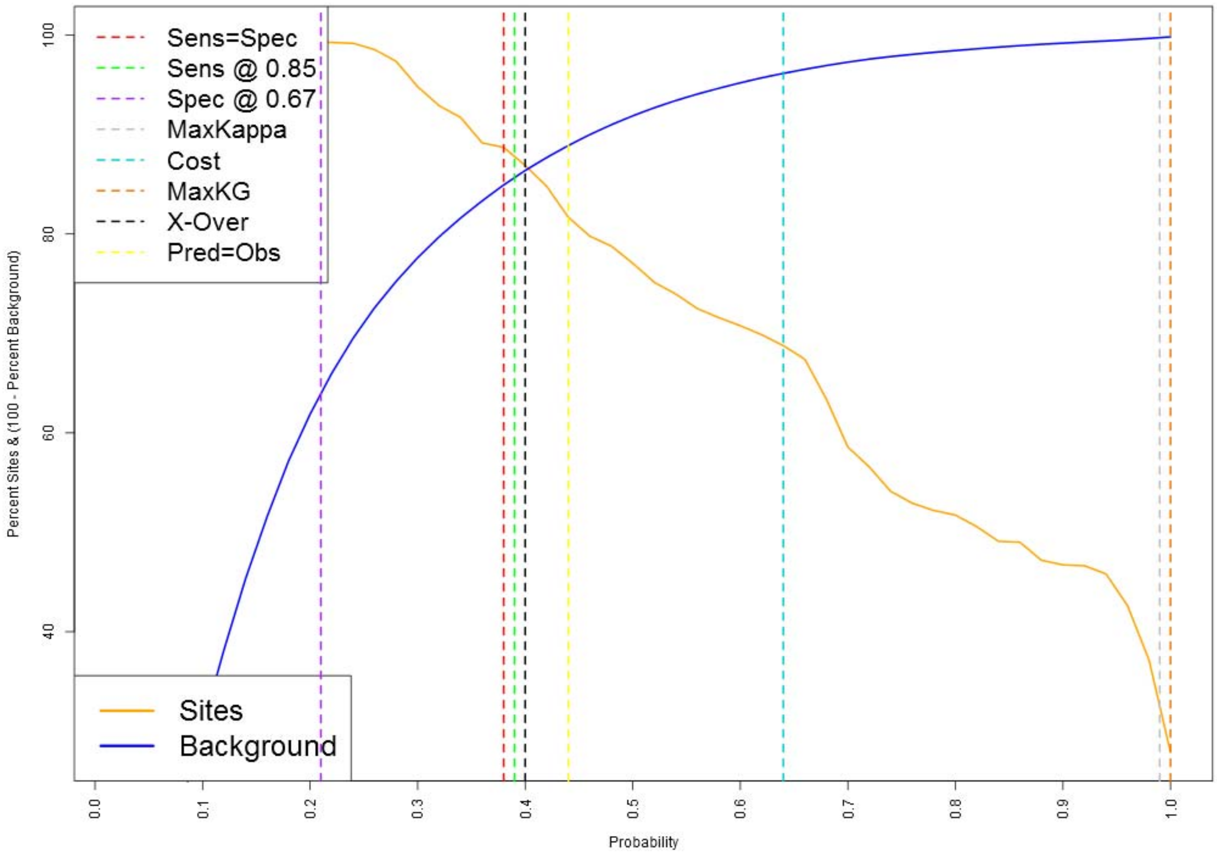
Region 7 All – Upland Section 6



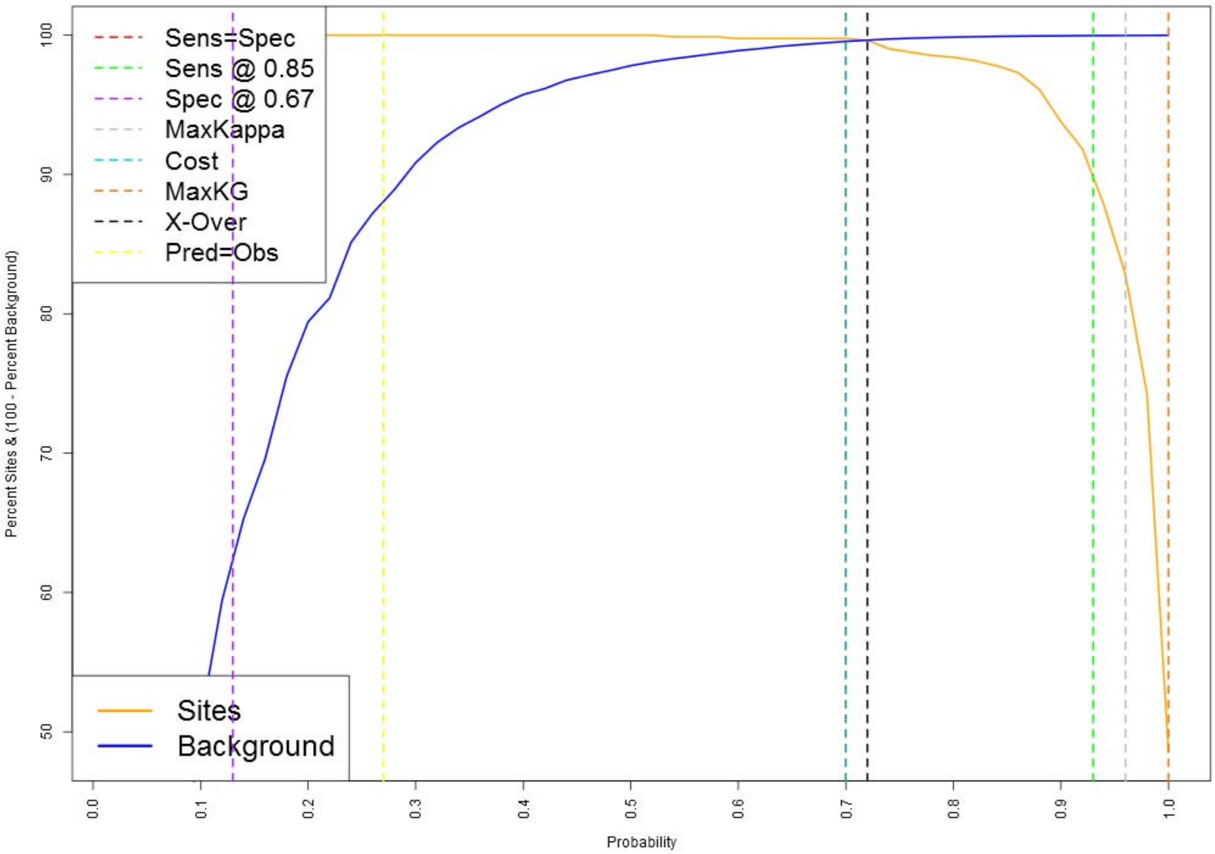
Region 7 All – Upland Section 7



Region 7 All – Upland Section 8

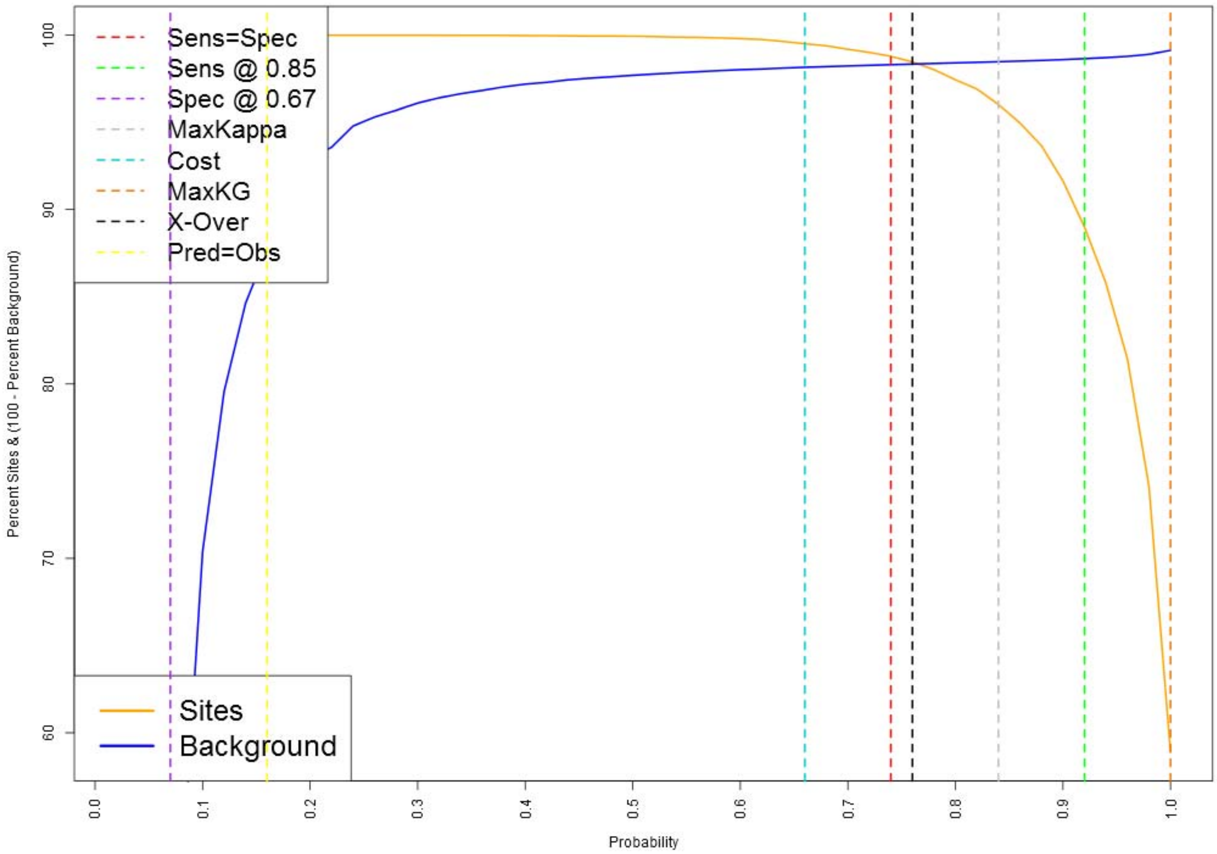


Region 7 All – Upland Section 9

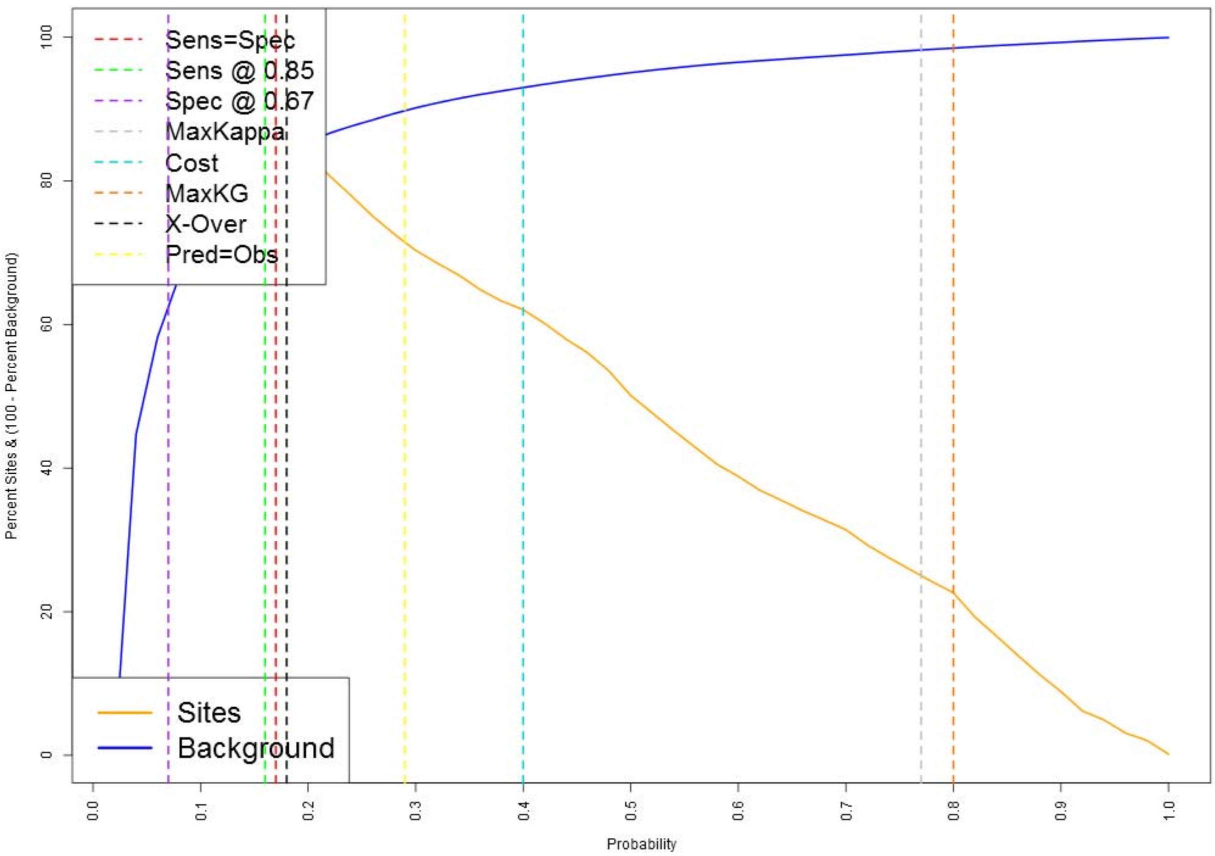


Note: The line for Sens=Spec is underneath the line for Cost and is not visible because the values are identical (0.70).

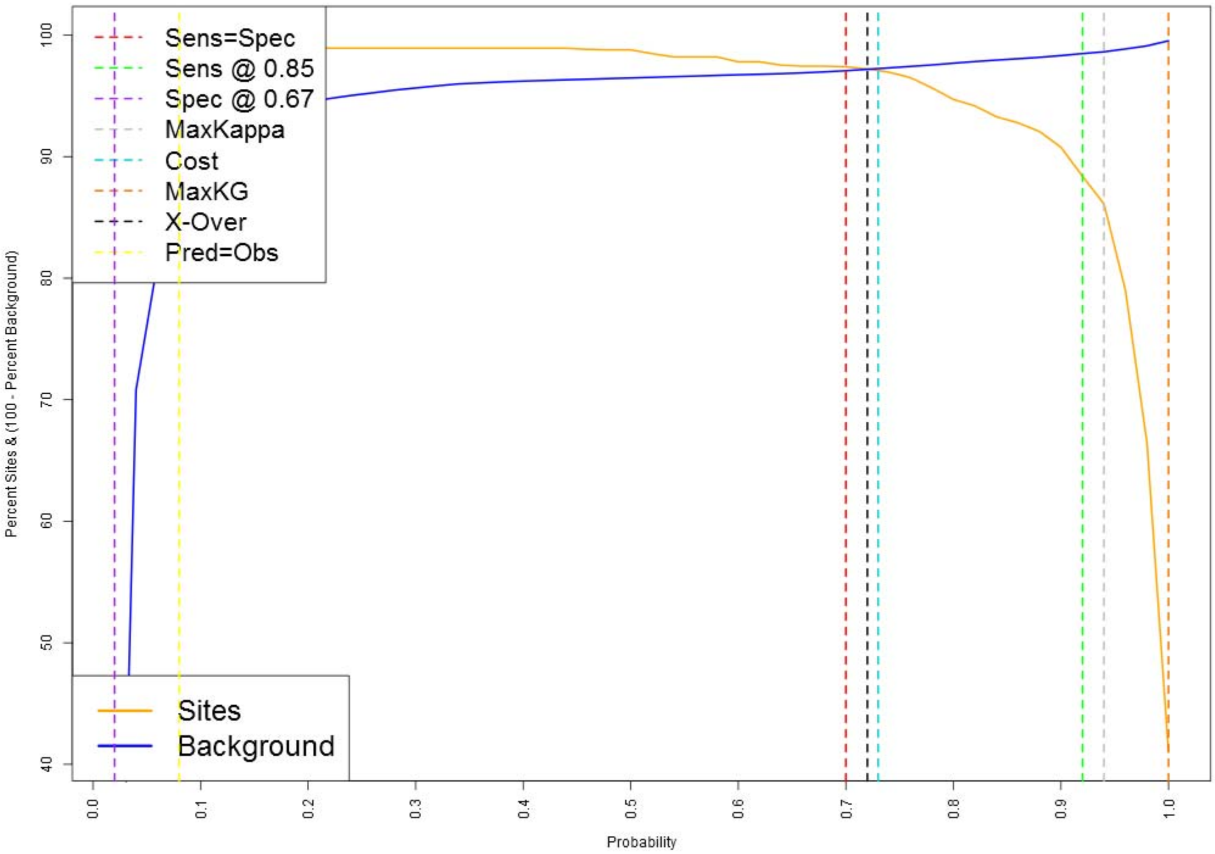
Region 8 All – Riverine Section 1



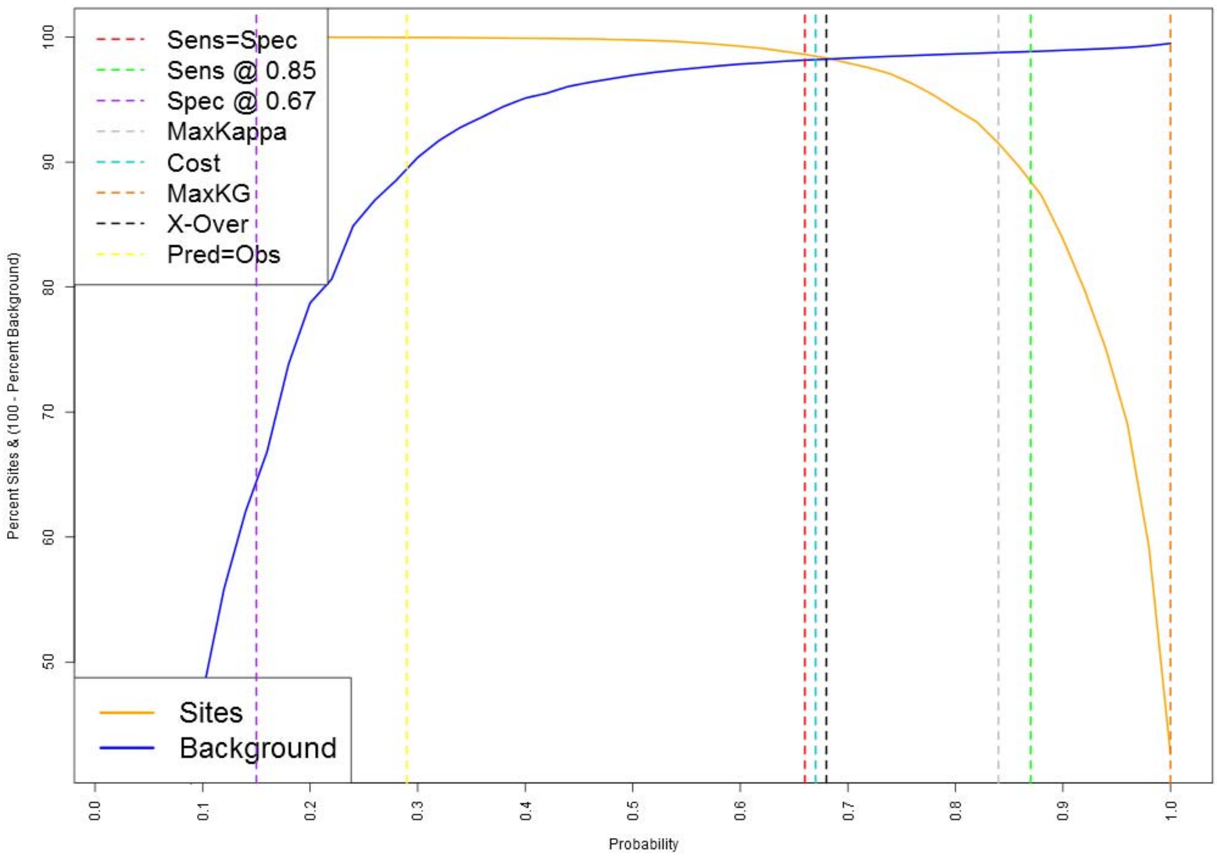
Region 8 All – Riverine Section 2



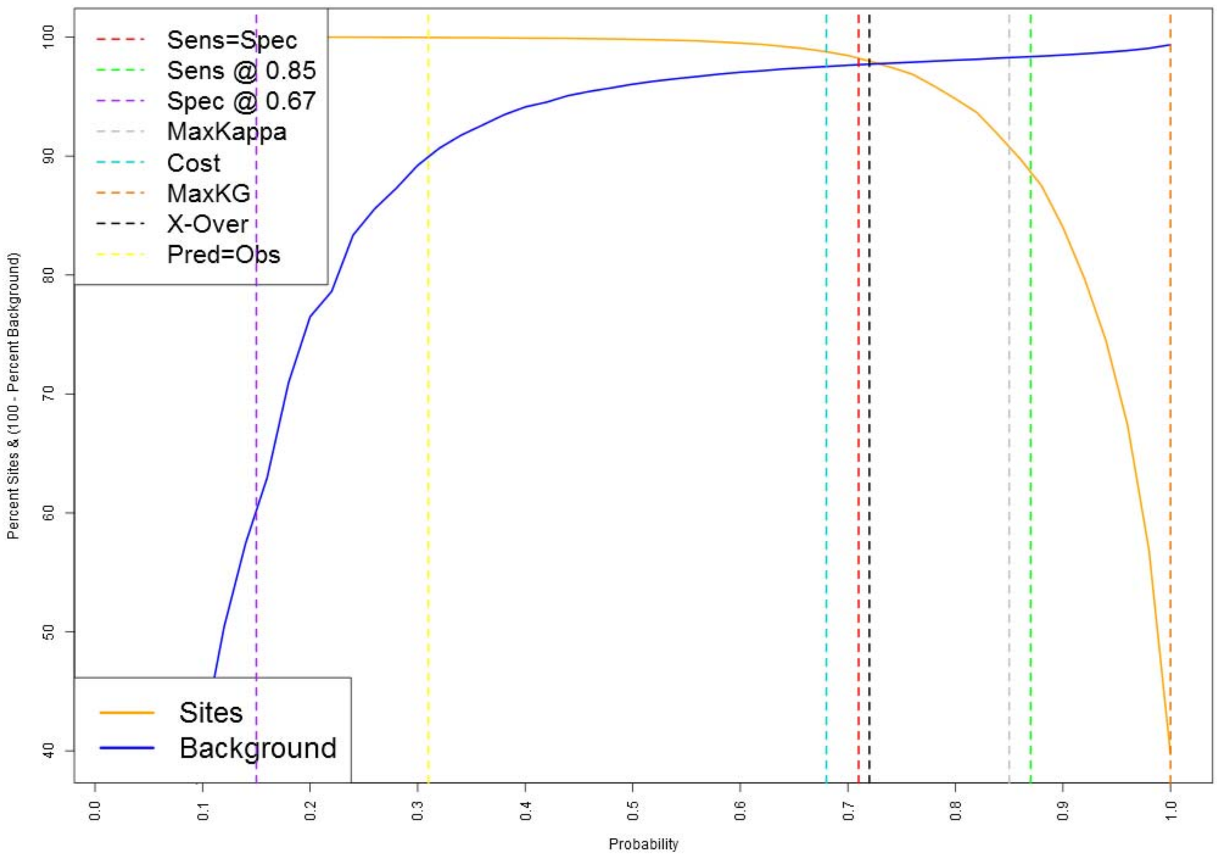
Region 8 All – Riverine Section 3



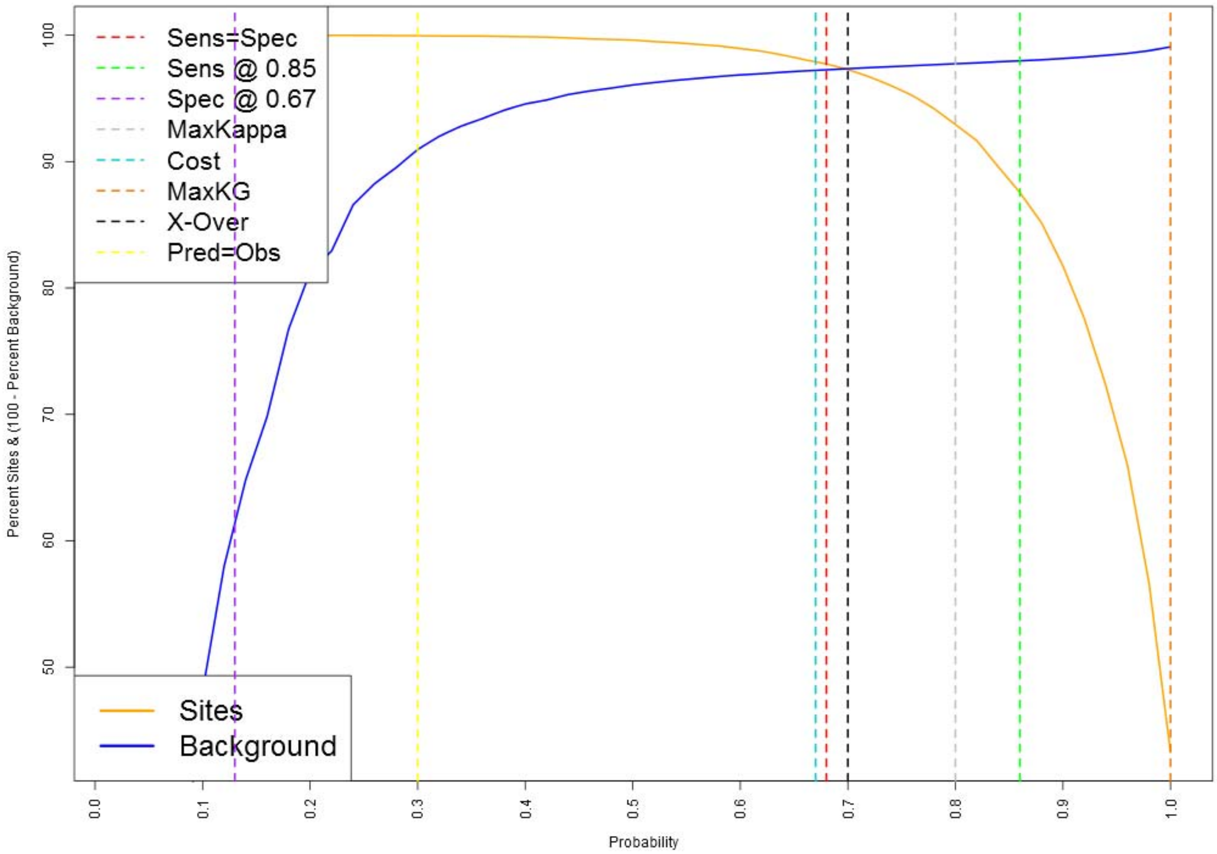
Region 8 All – Riverine Section 4



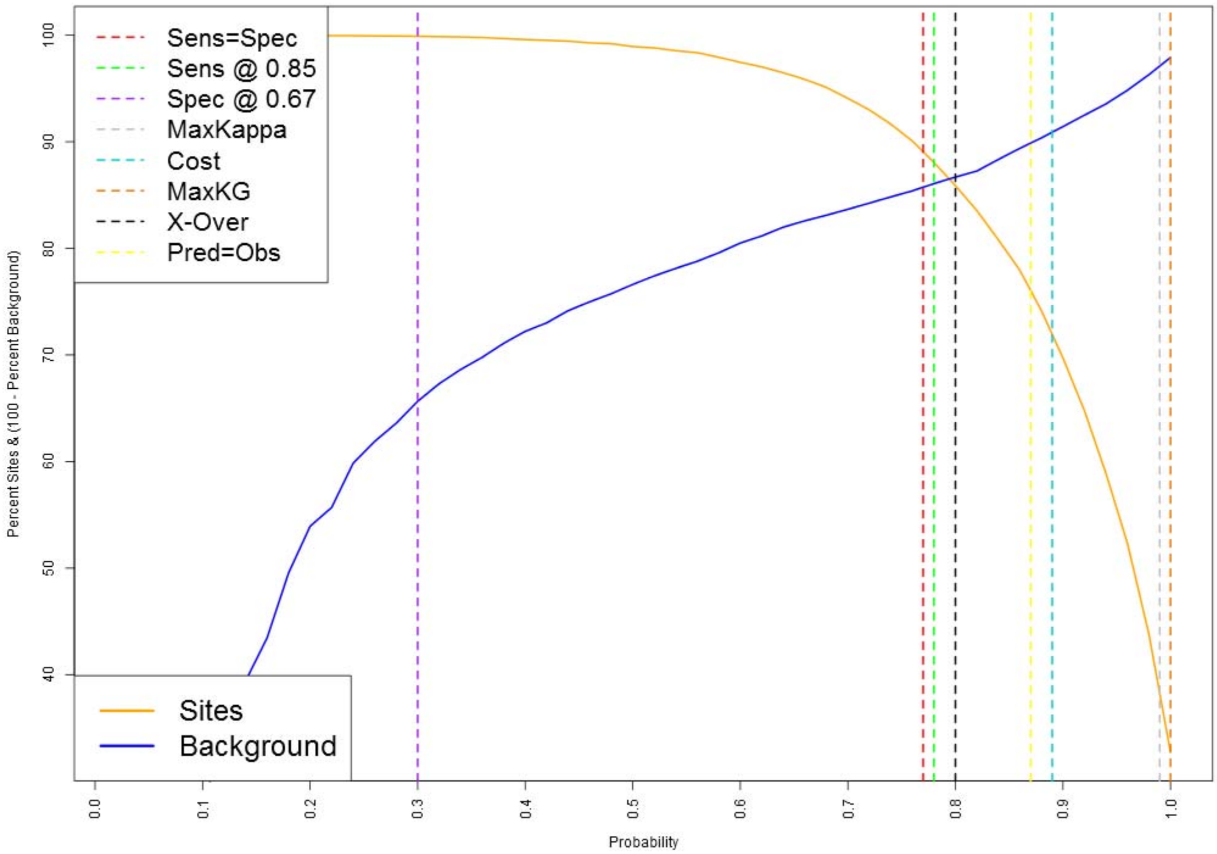
Region 8 All – Riverine Section 5



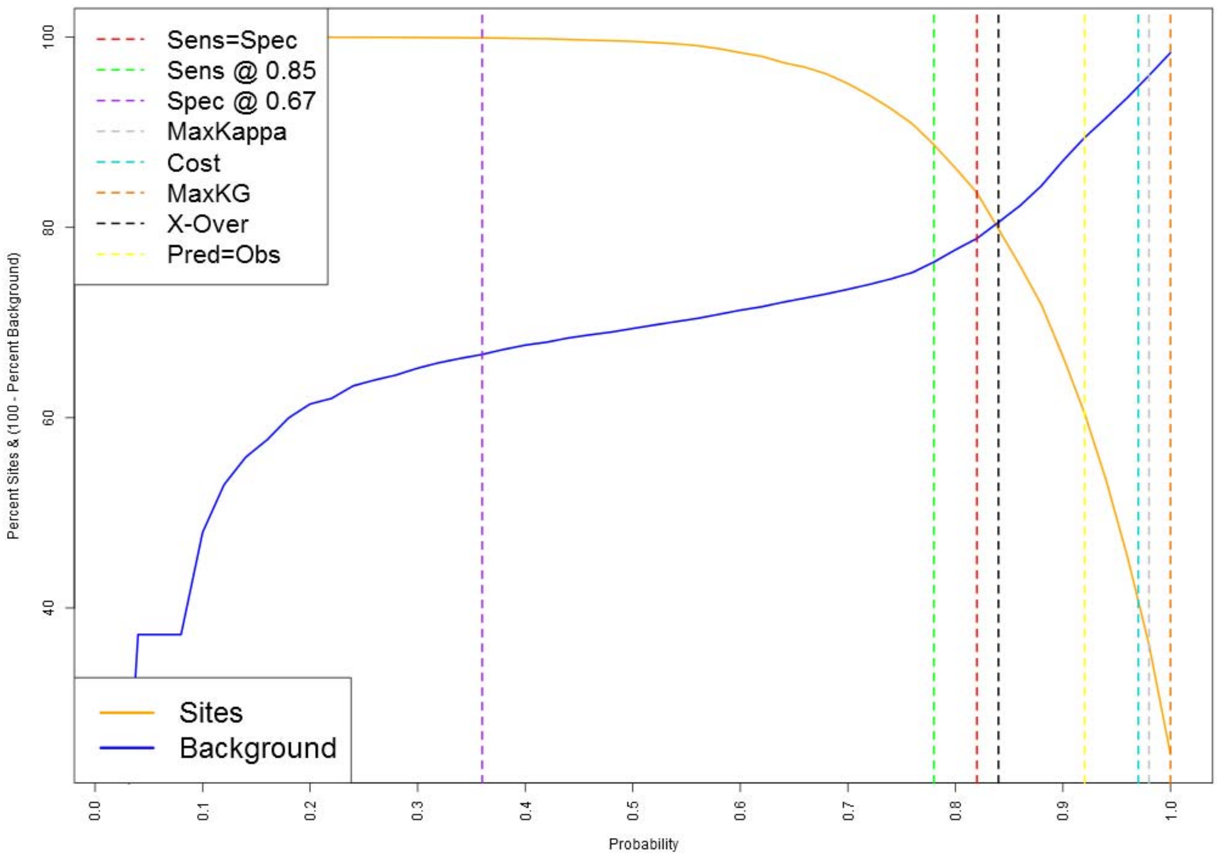
Region 8 All – Riverine Section 6



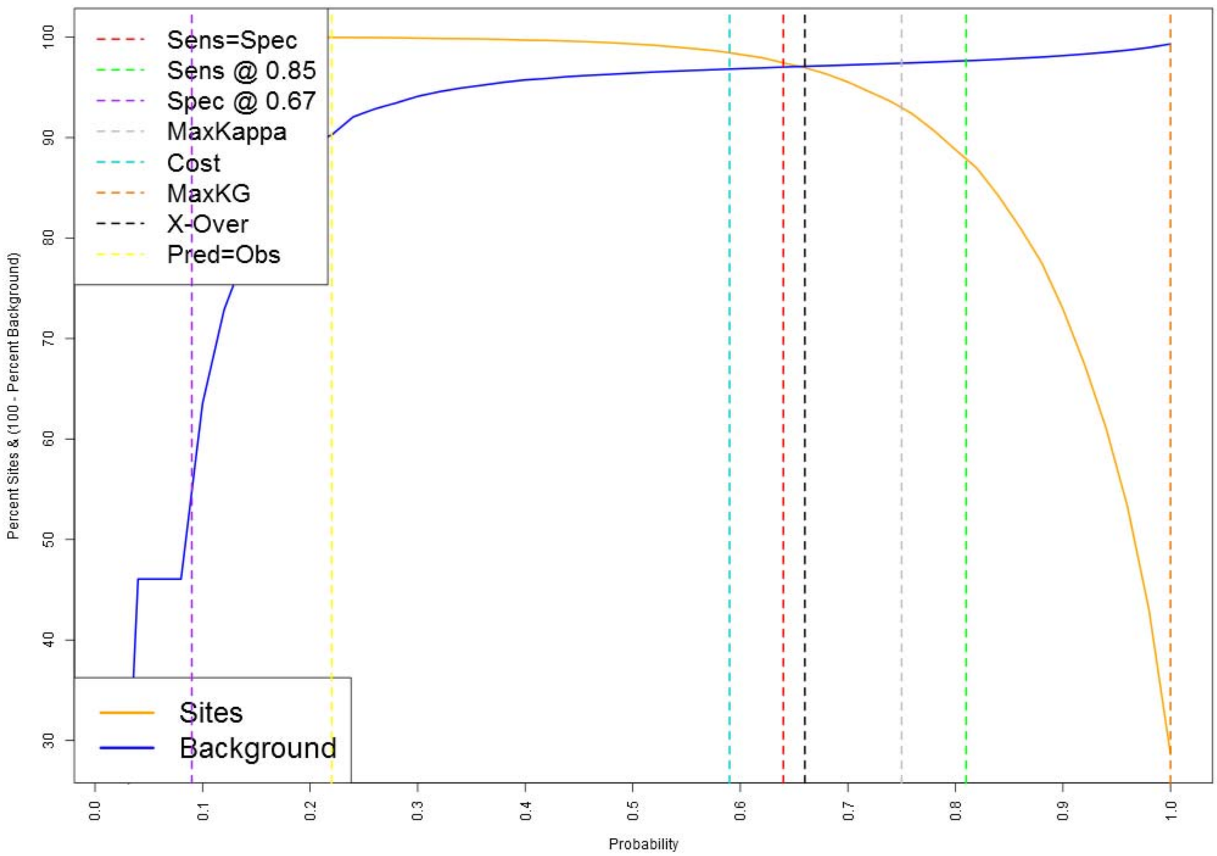
Region 8 All – Riverine Section 7



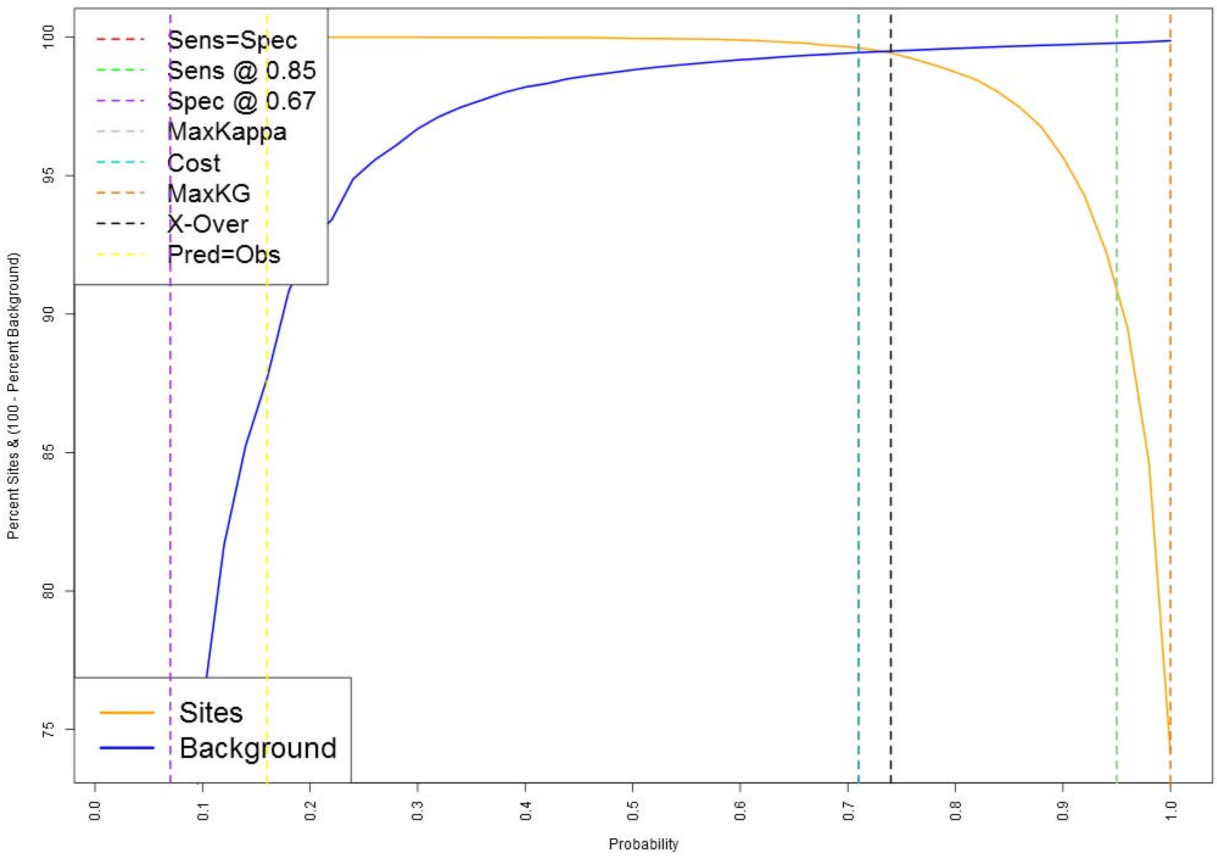
Region 8 All – Riverine Section 8



Region 8 All – Riverine Section 9

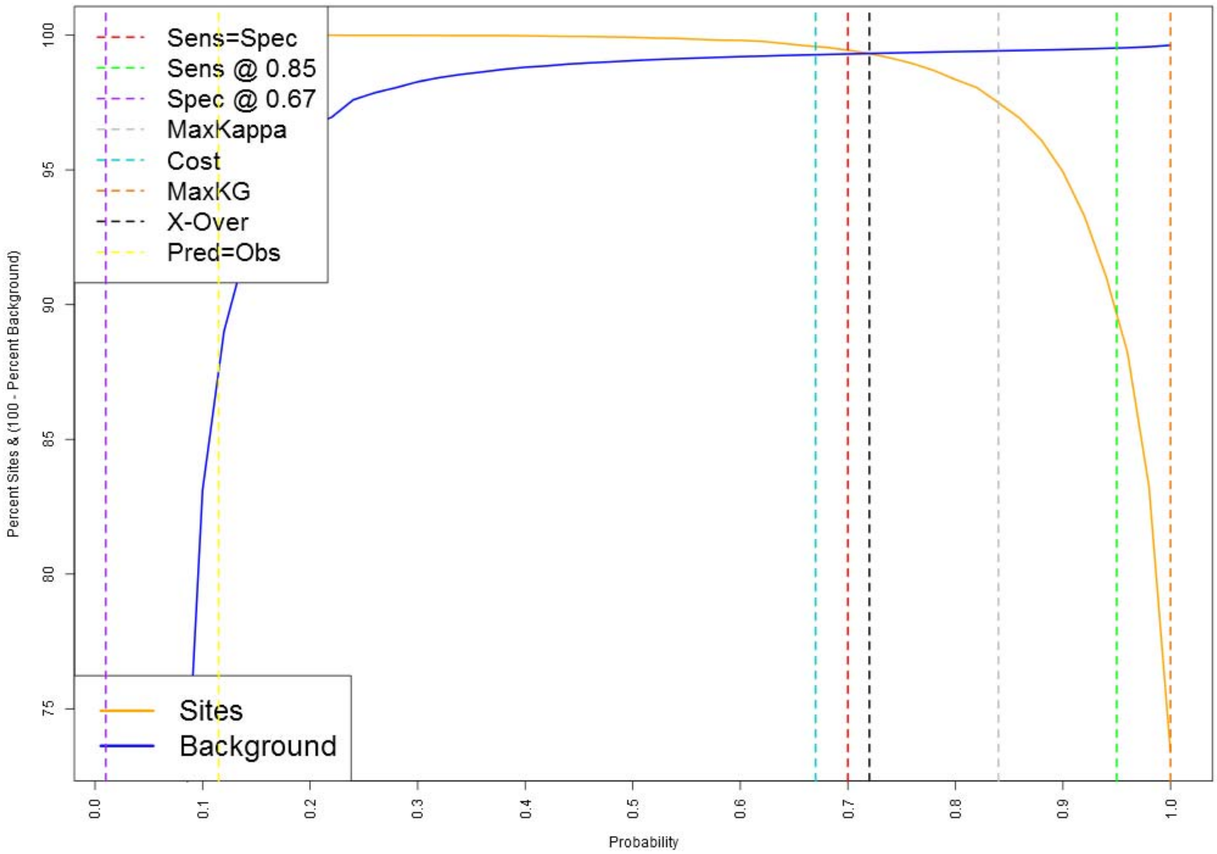


Region 8 All – Upland Section 1

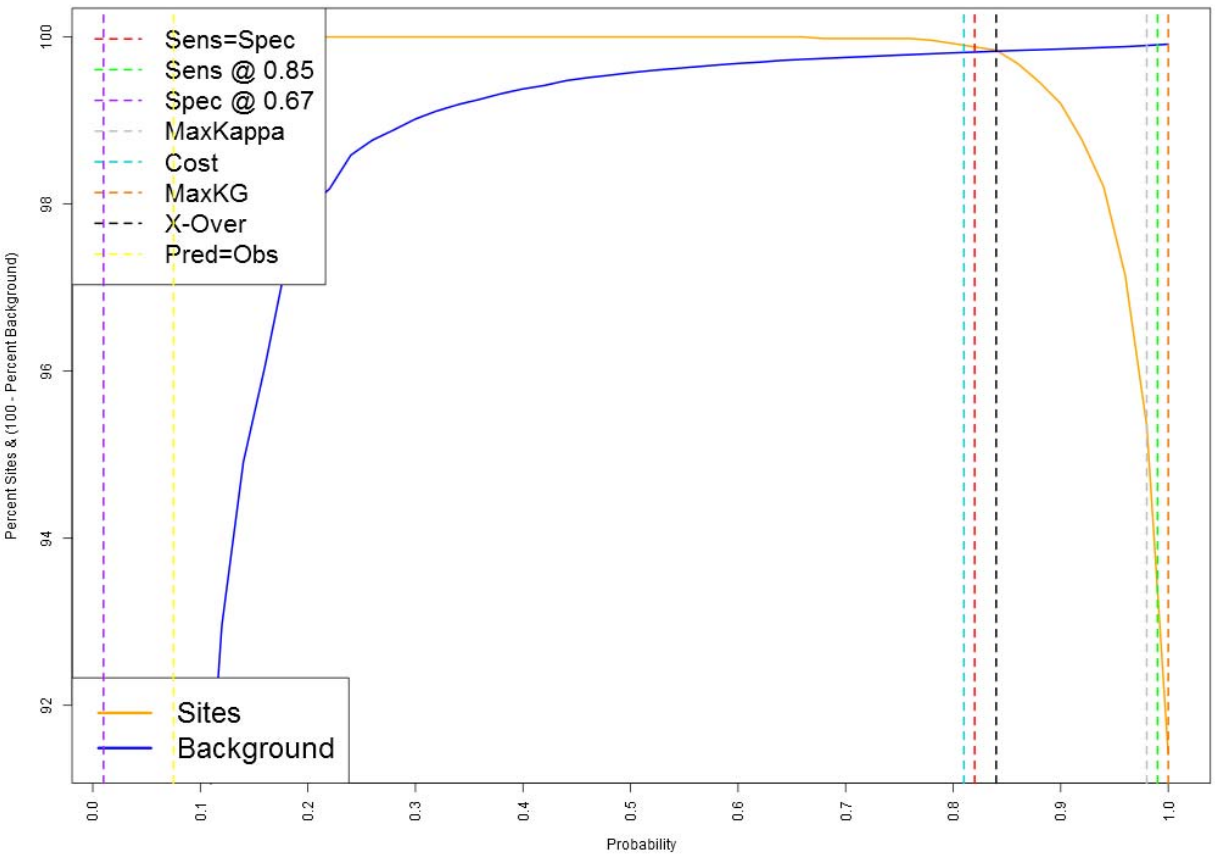


Note: The line for Sens @ 0.85 is underneath the line for MaxKappa and is not visible because the values are identical (0.95); similarly, the line for Sens=Spec is obscured by the line for Cost (0.71).

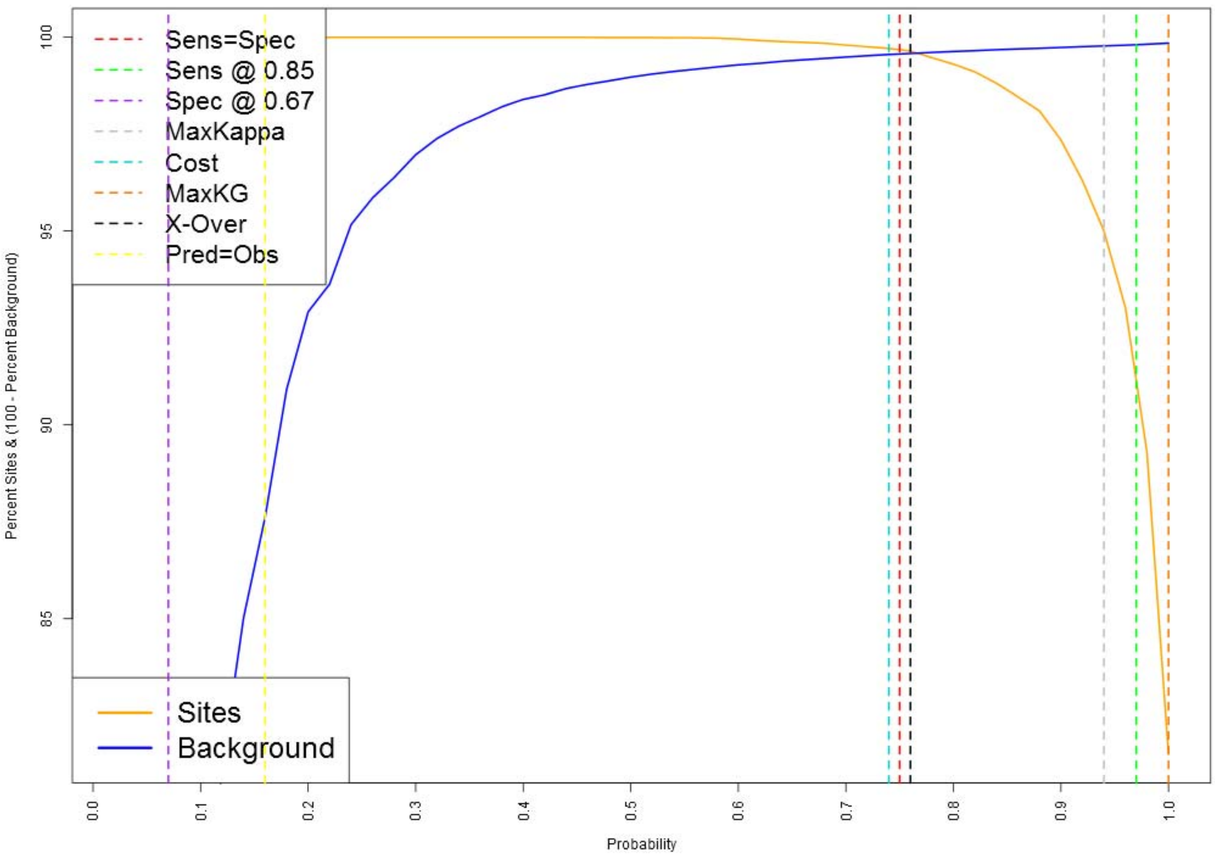
Region 8 All – Upland Section 2



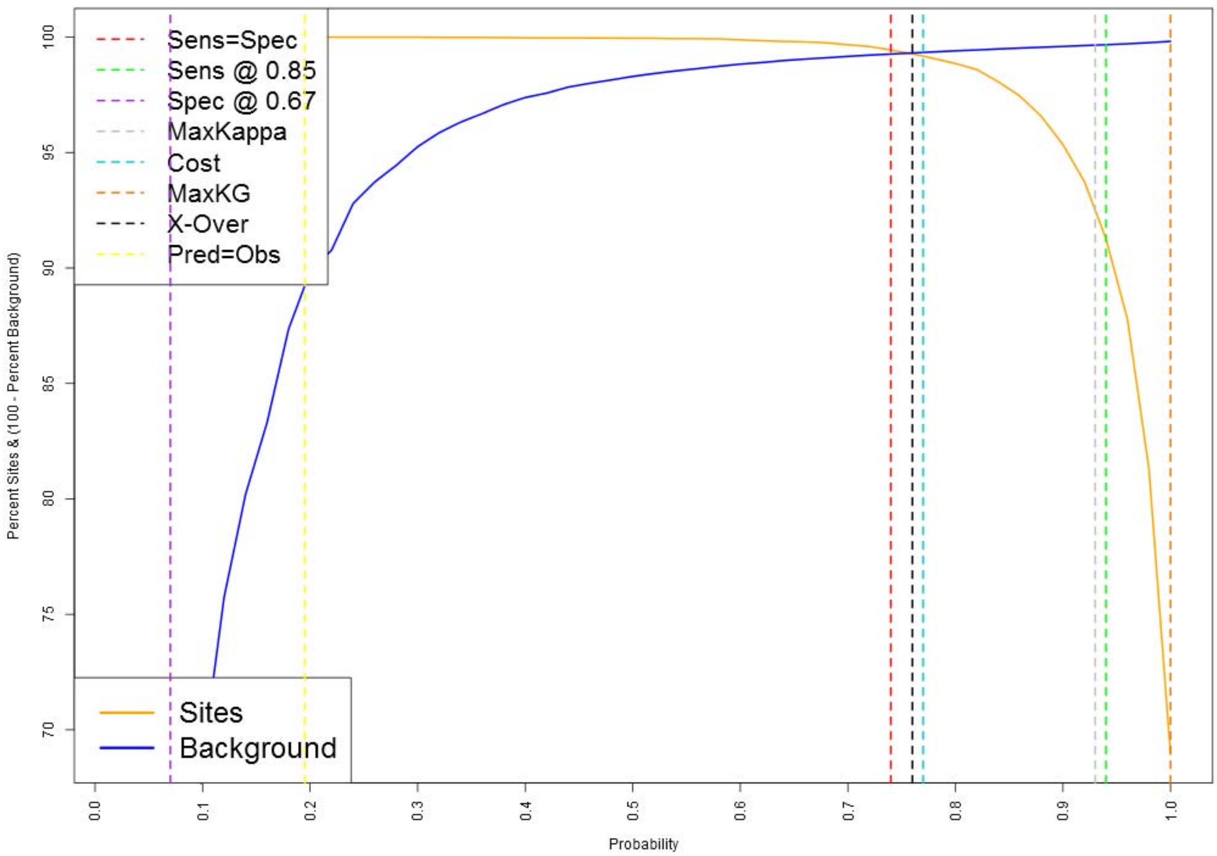
Region 8 All – Upland Section 3



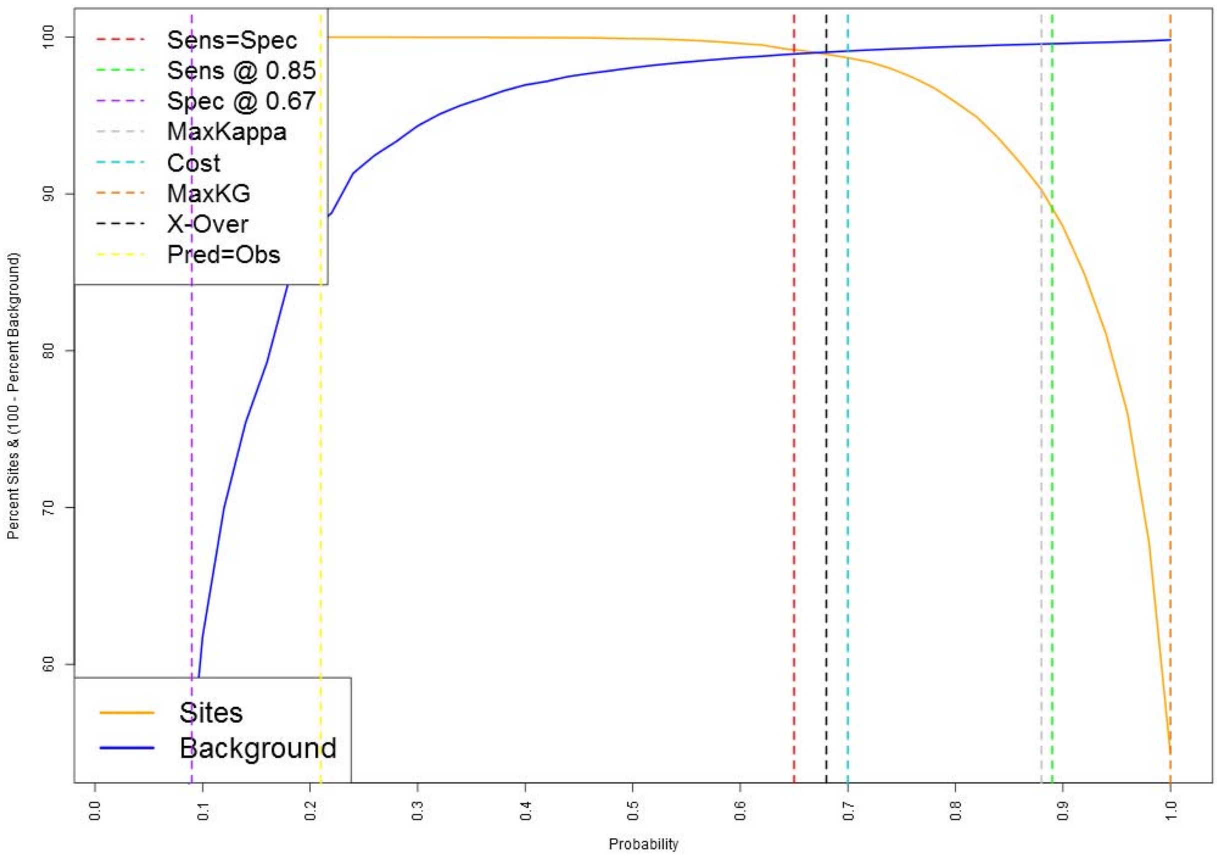
Region 8 All – Upland Section 4



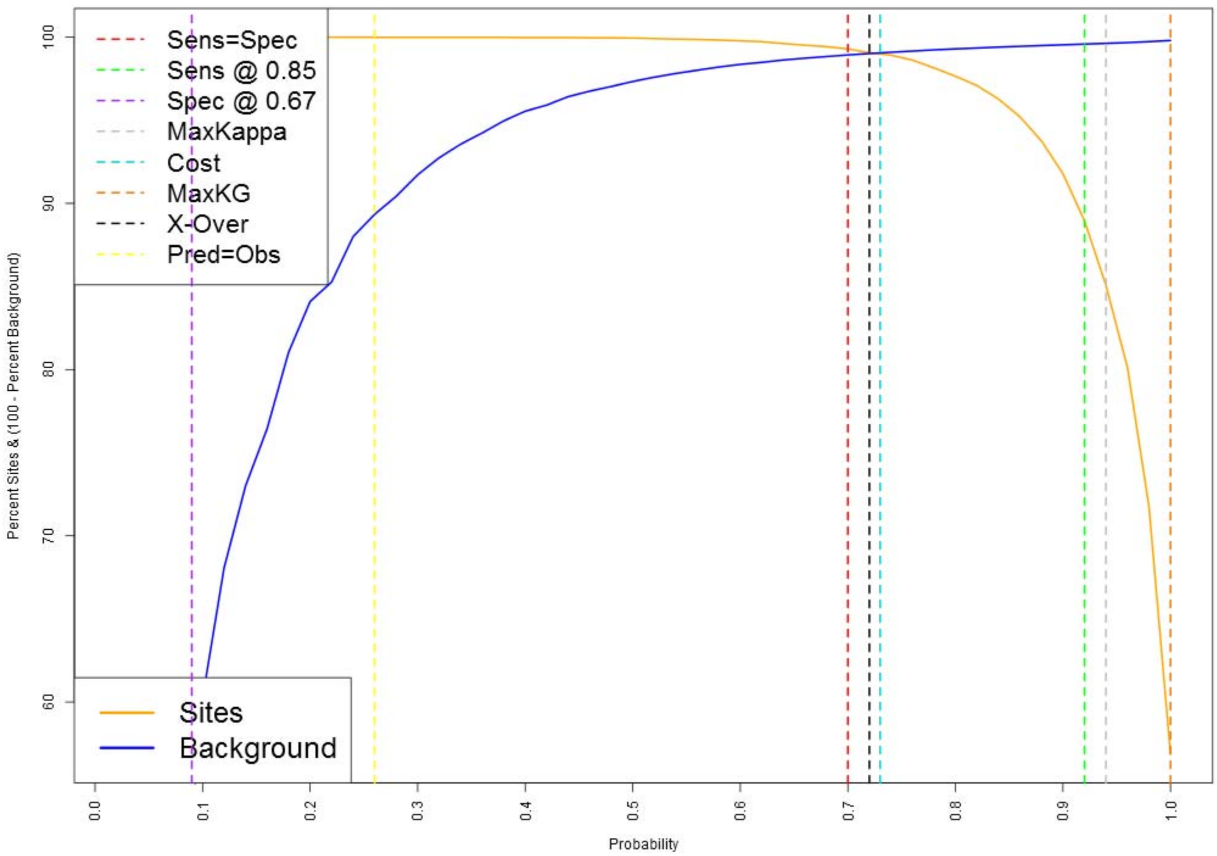
Region 8 All – Upland Section 5



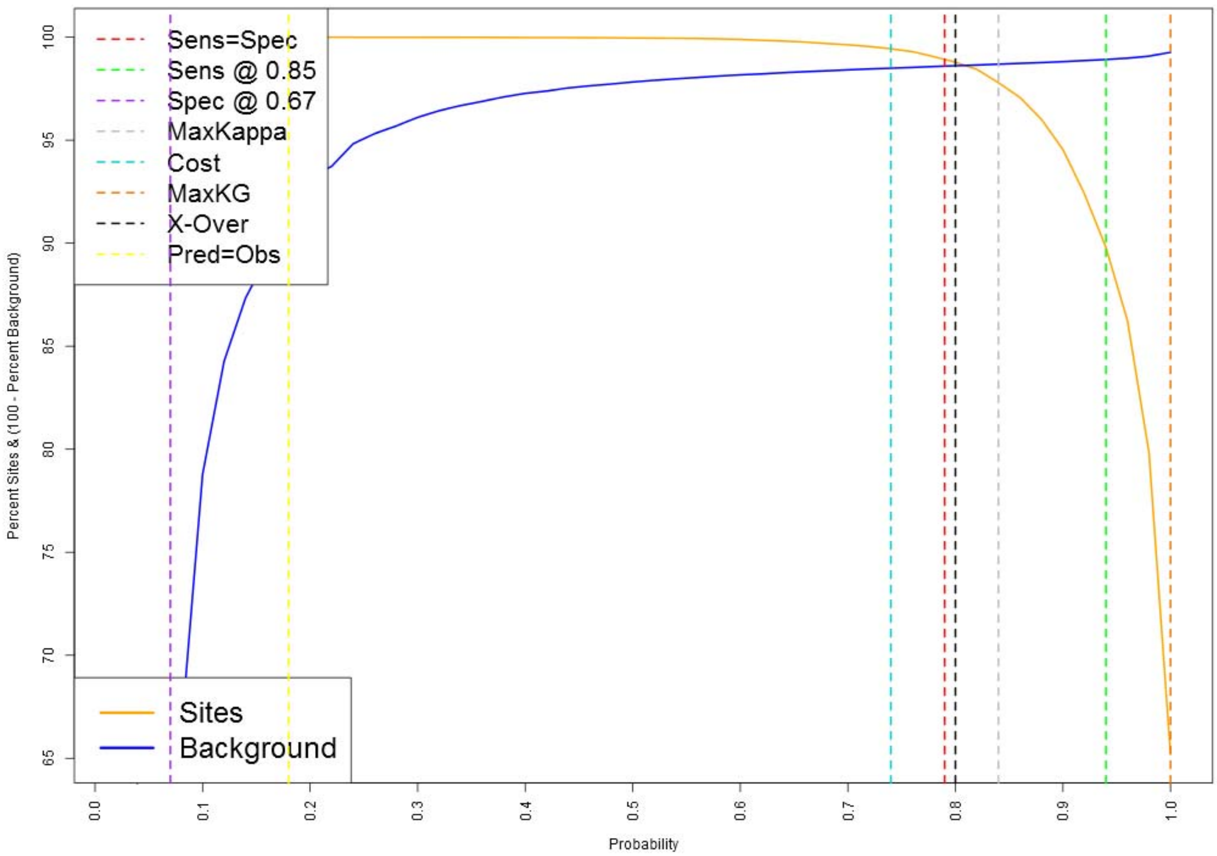
Region 8 All – Upland Section 6



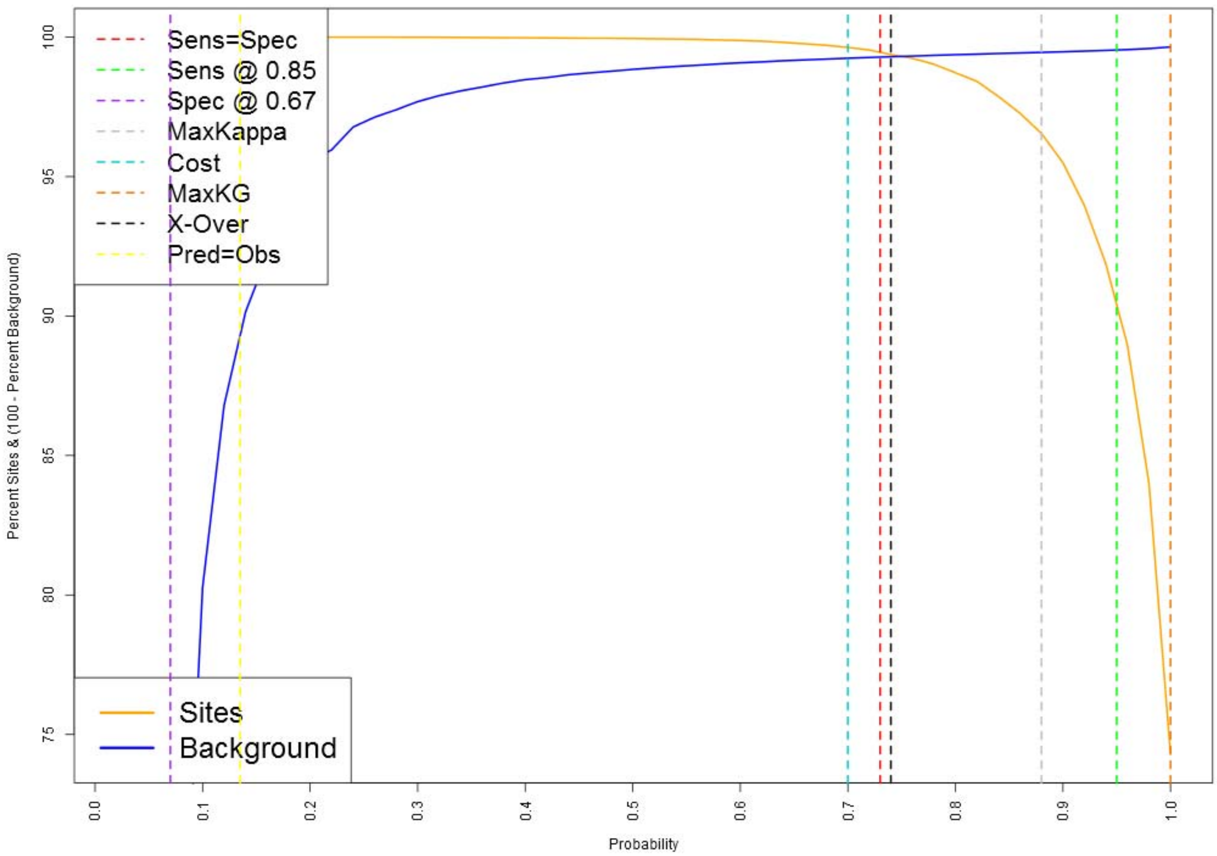
Region 8 All – Upland Section 7



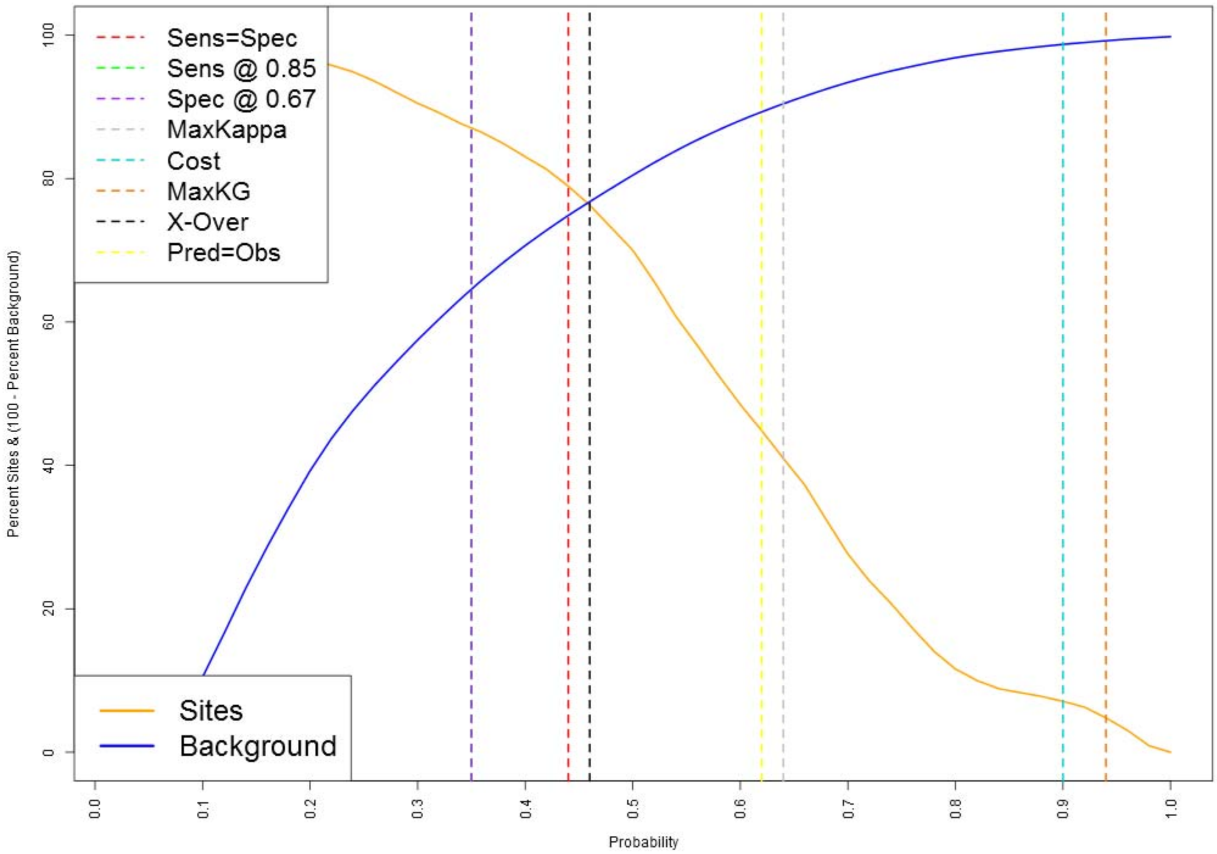
Region 8 All – Upland Section 8



Region 8 All – Upland Section 9

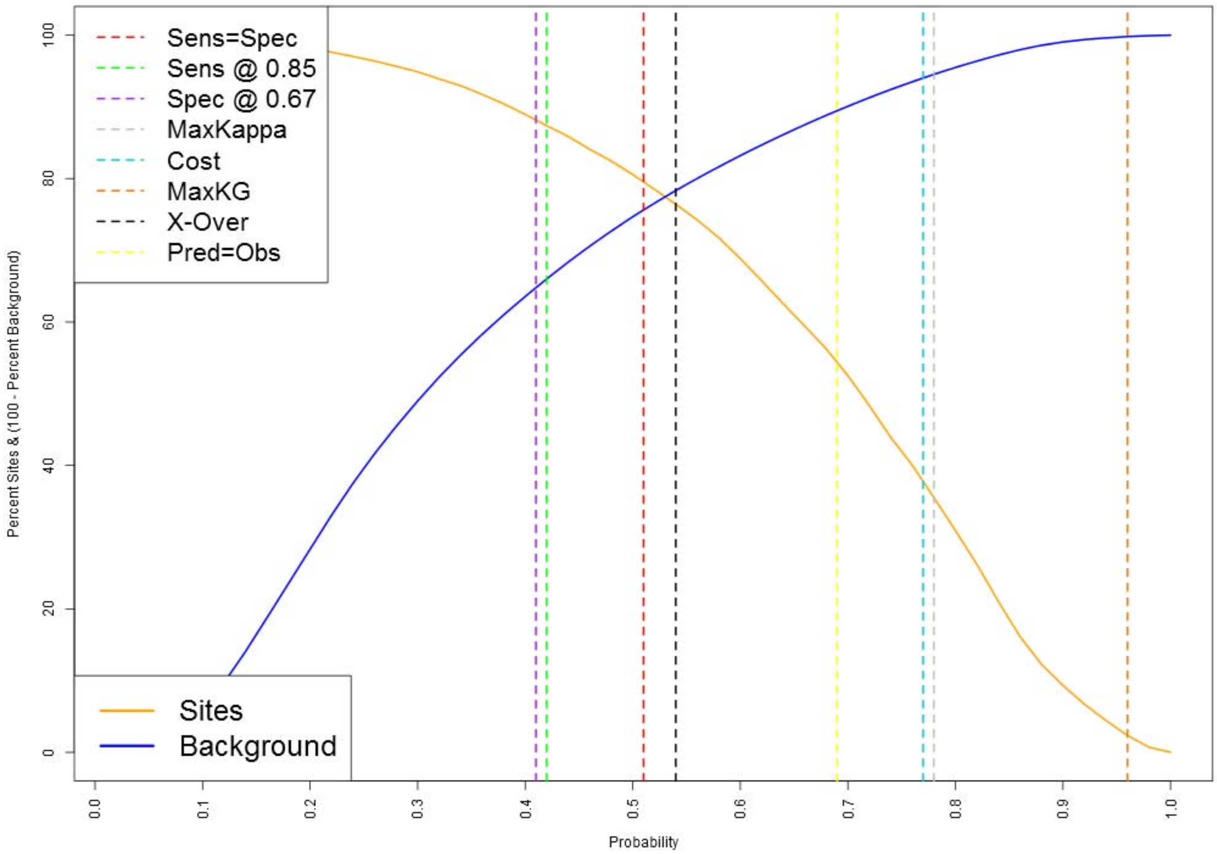


Region 9-10 All – Riverine Section 1

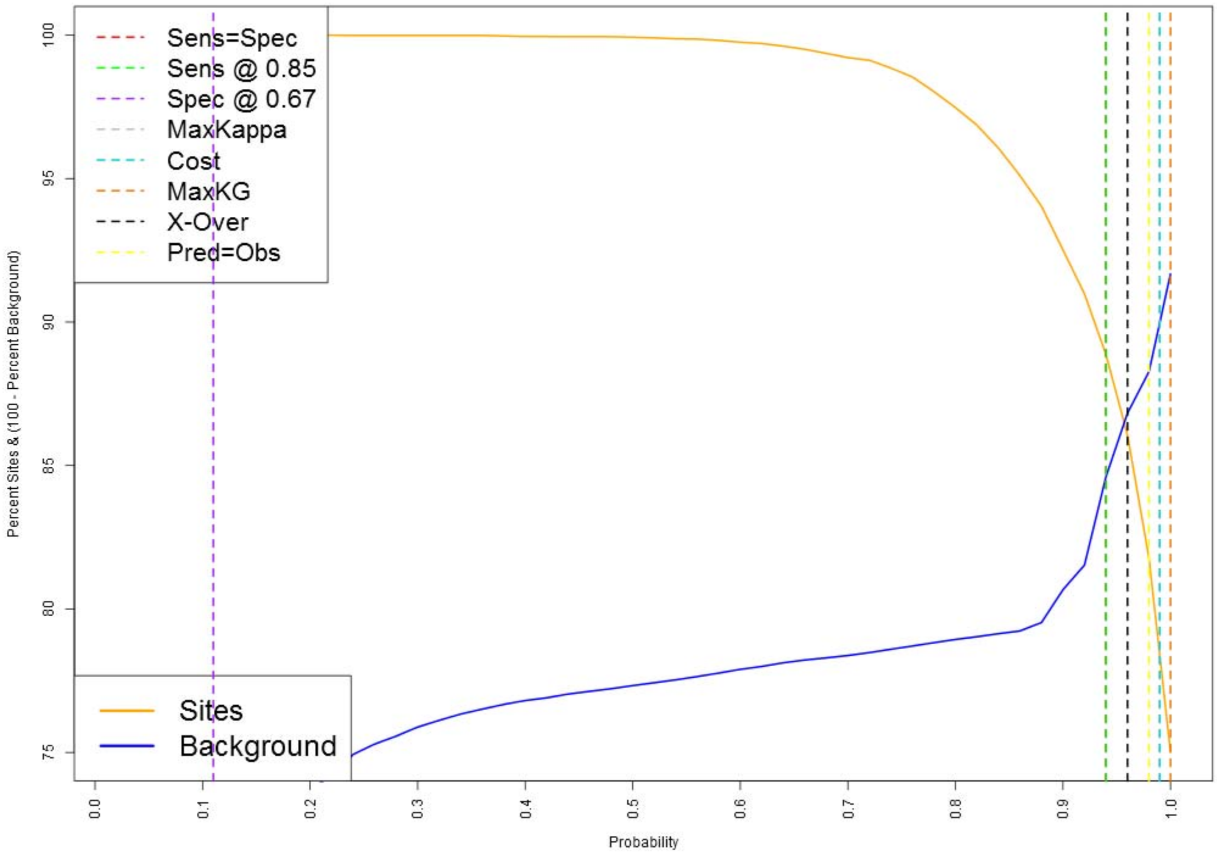


Note: The line for Sens @ 0.85 is underneath the line for Spec @ 0.67 and is not visible because the values are identical (0.35).

Region 9-10 All – Riverine Section 2

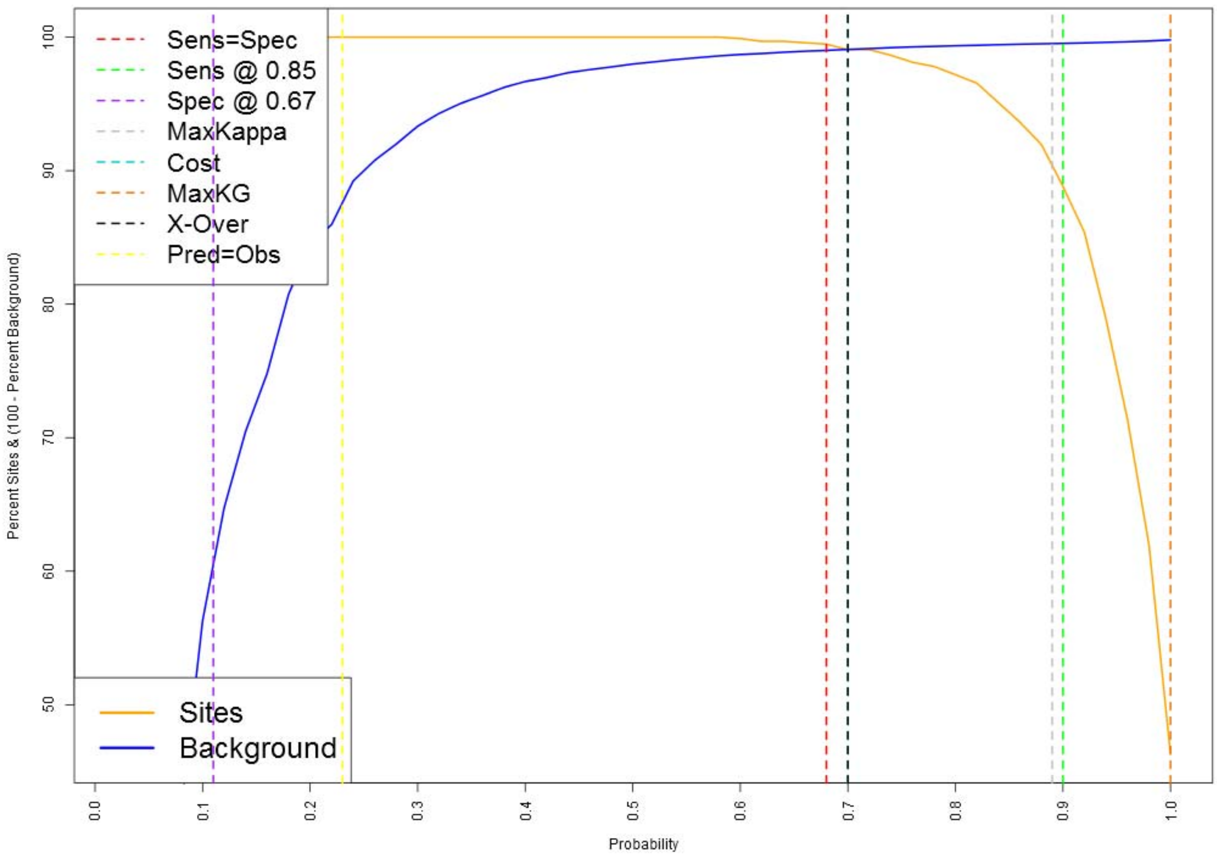


Region 9-10 All – Riverine Section 3



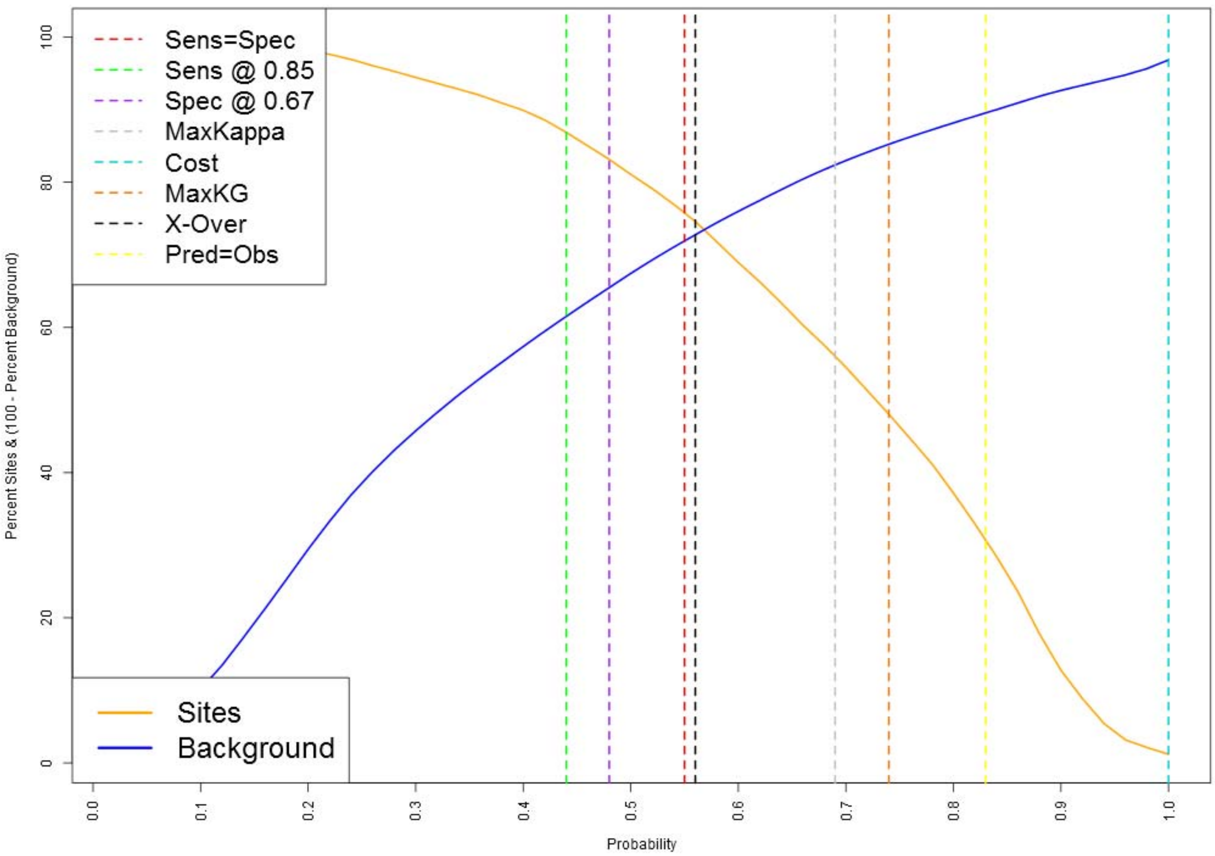
Note: The line for Sens=Spec is underneath the line for Sens @ 0.85 and is not visible because the values are identical (0.94); similarly, the line for MaxKappa is obscured by the line for Cost (0.99).

Region 9-10 All – Riverine Section 4

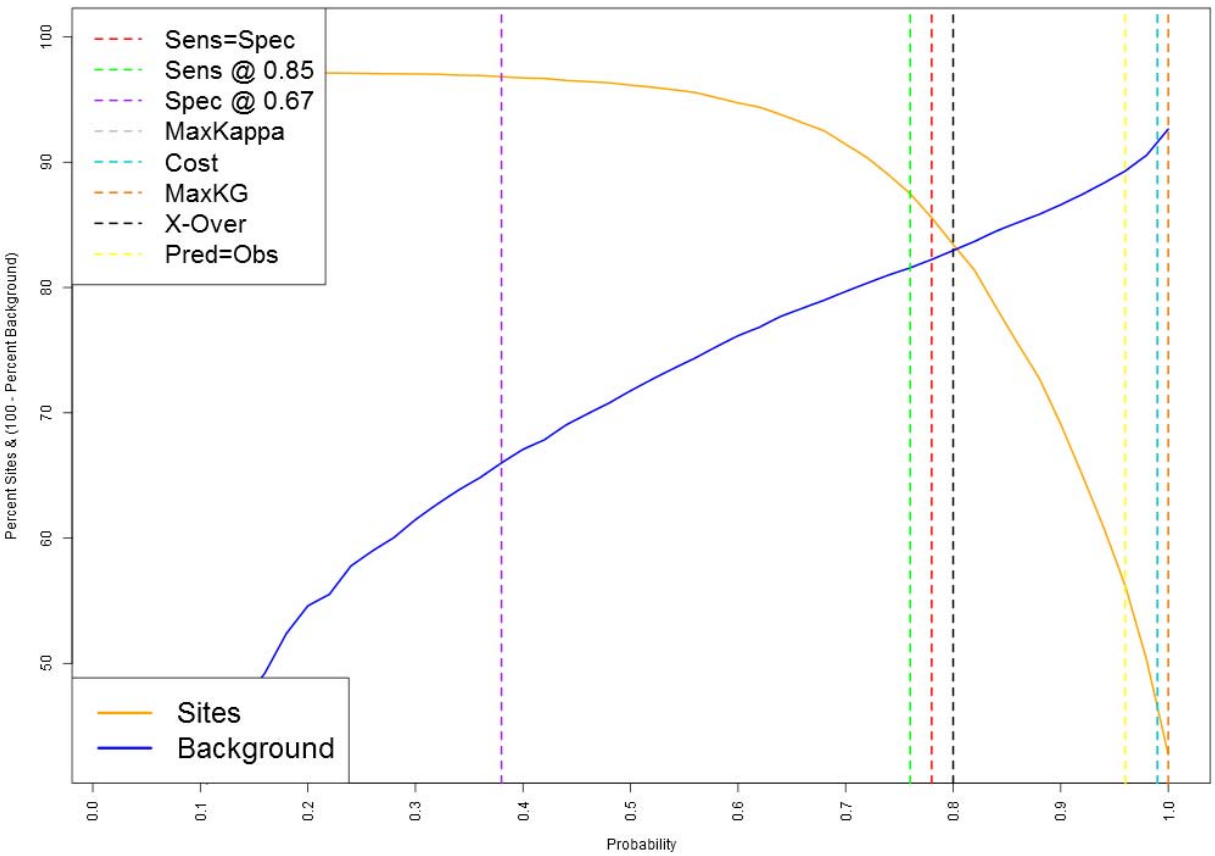


Note: The line for Cost is underneath the line for Sens @ 0.85 and is not visible because the values are identical (0.90).

Region 9-10 All – Riverine Section 5

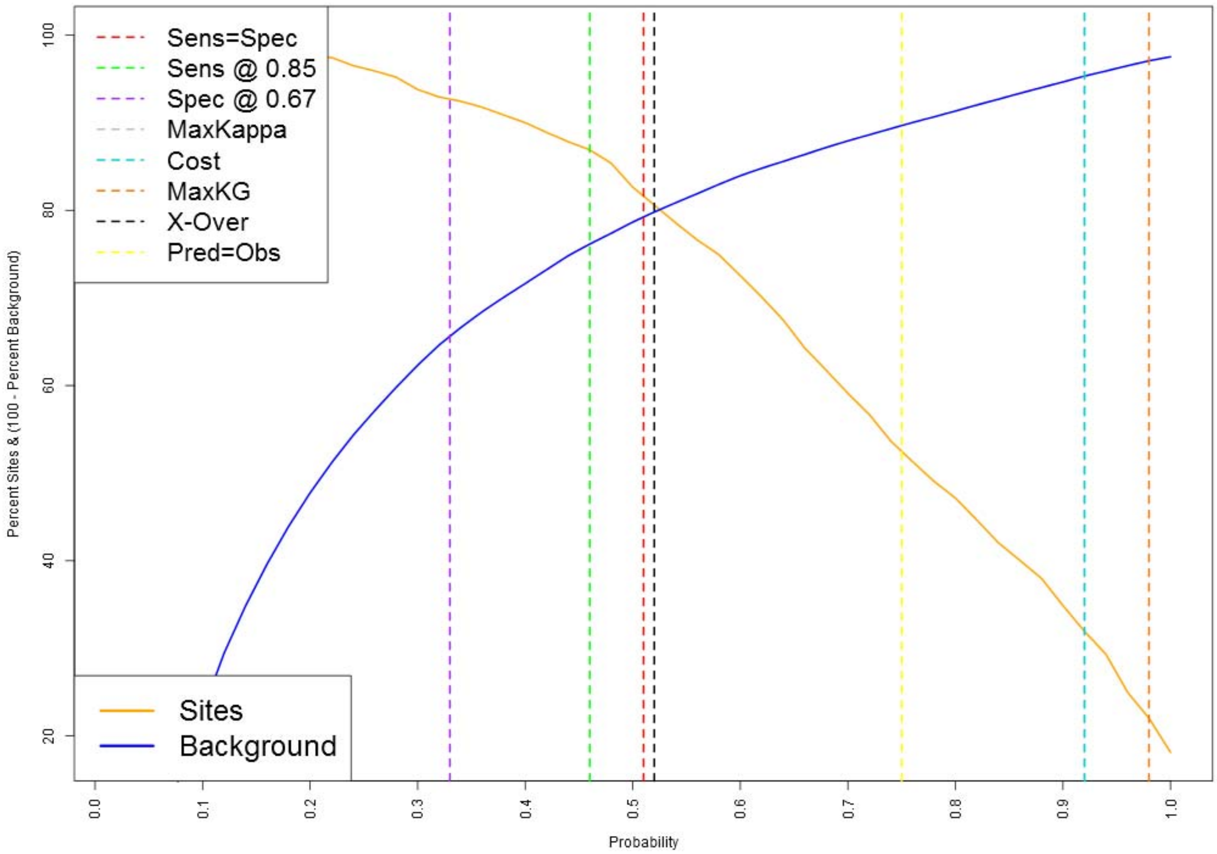


Region 9-10 All – Riverine Section 6



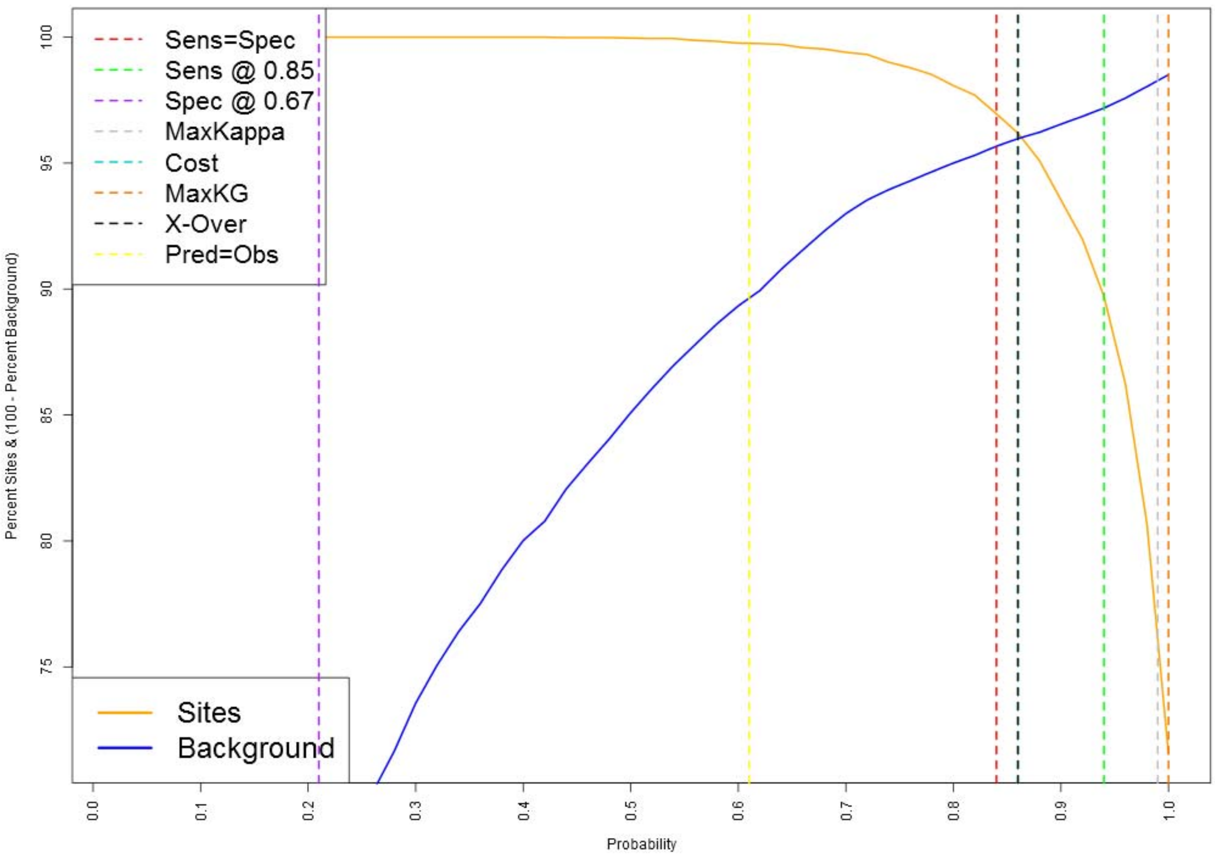
Note: The line for MaxKappa is underneath the line for Cost and is not visible because the values are identical (0.99).

Region 9-10 All – Riverine Section 7



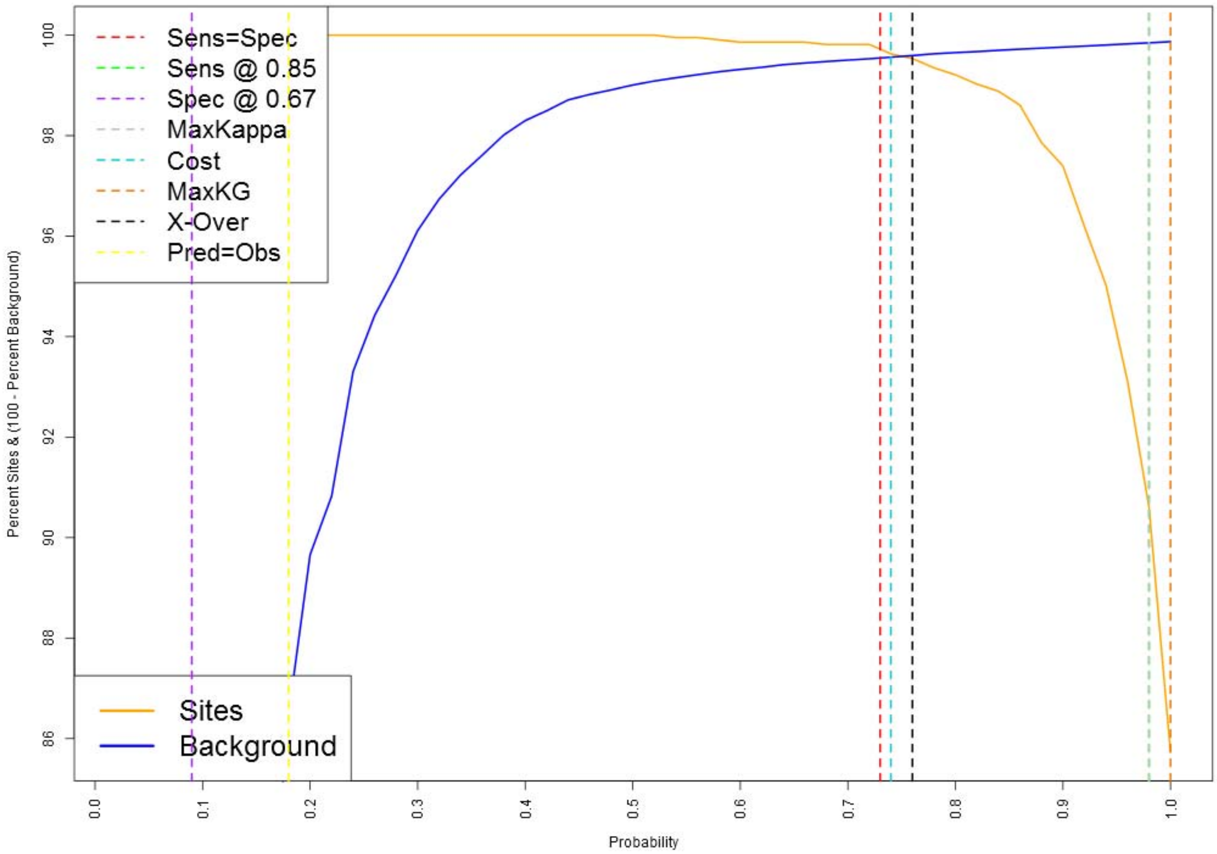
Note: The line for MaxKappa is underneath the line for Cost and is not visible because the values are identical (0.92).

Region 9-10 All – Riverine Section 8



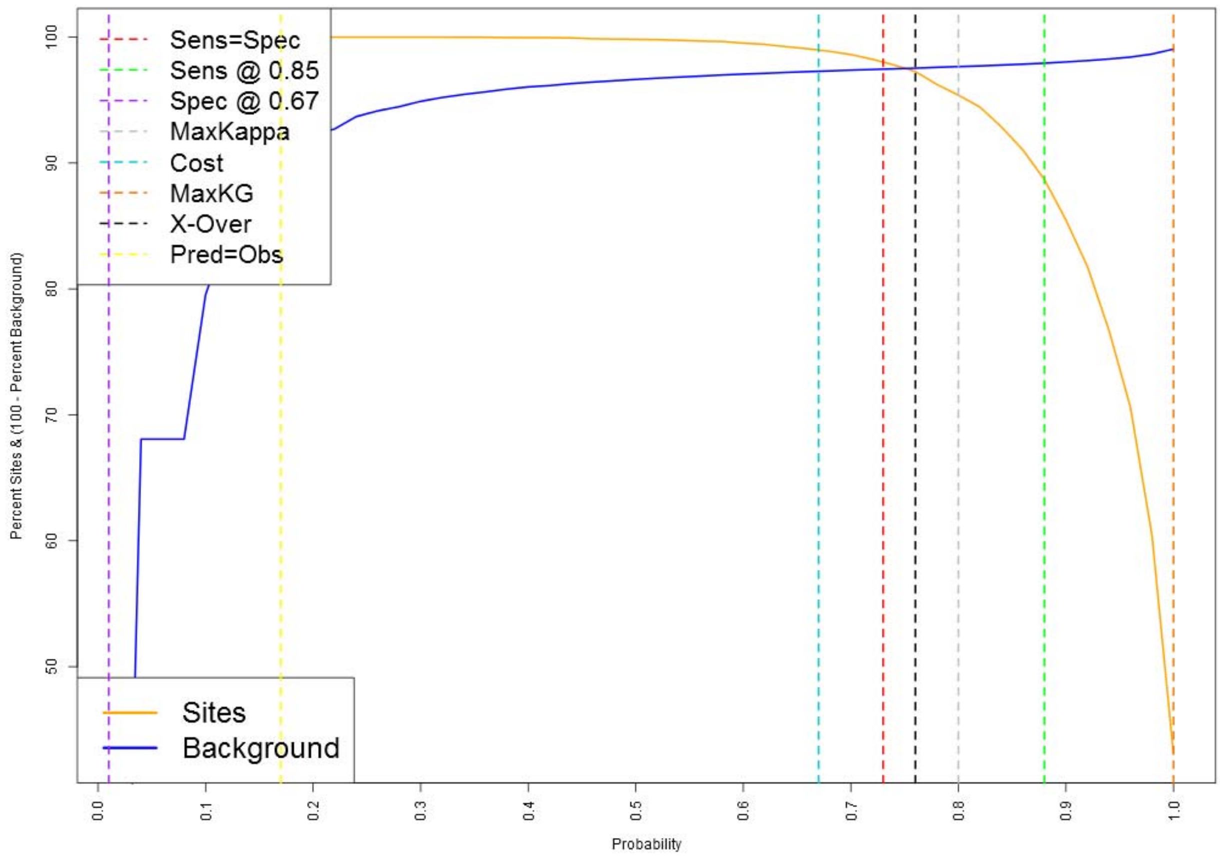
Note: The line for Cost is underneath the line for X-Over and is not visible because the values are identical (0.86).

Region 9-10 All – Riverine Section 9

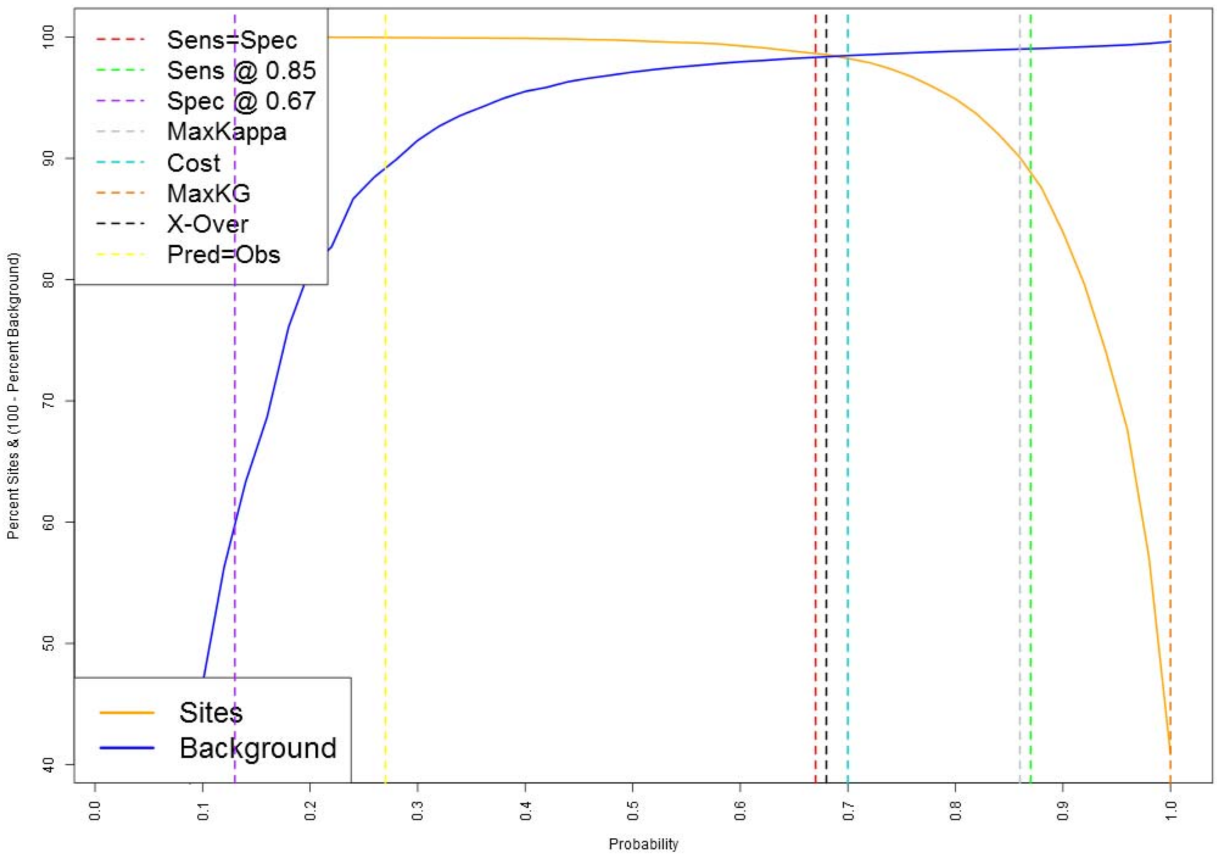


Note: The line for Sens @ 0.85 is underneath the line for MaxKappa and is not visible because the values are identical (0.98).

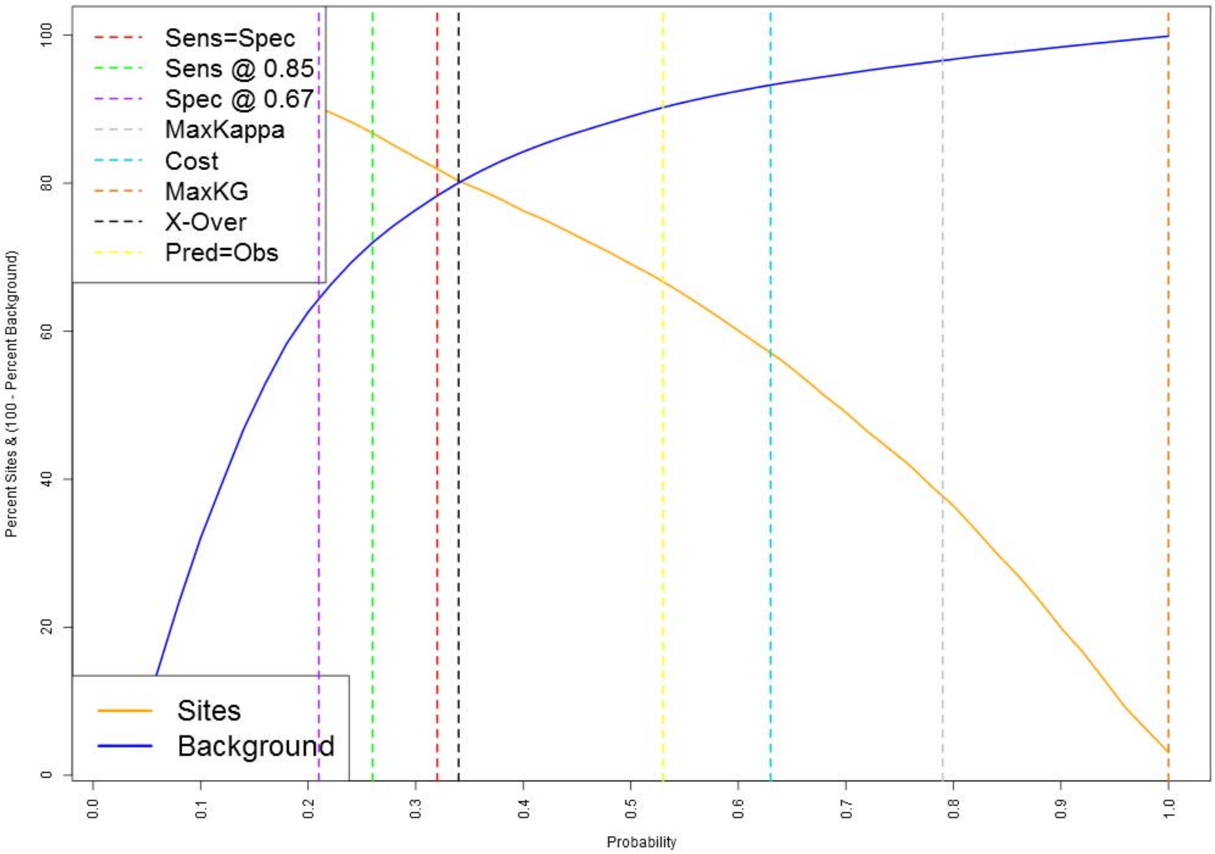
Region 9-10 All – Riverine Section 10



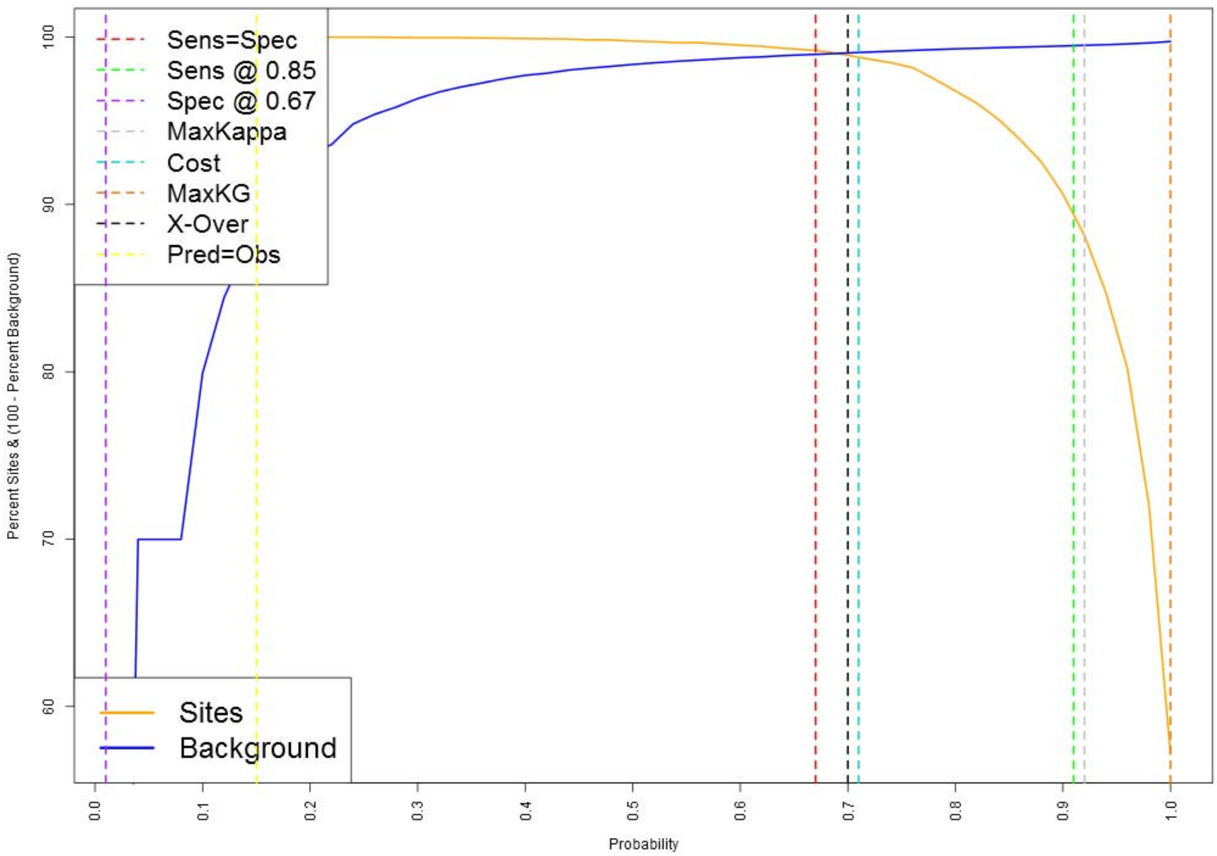
Region 9-10 All – Riverine Section 11



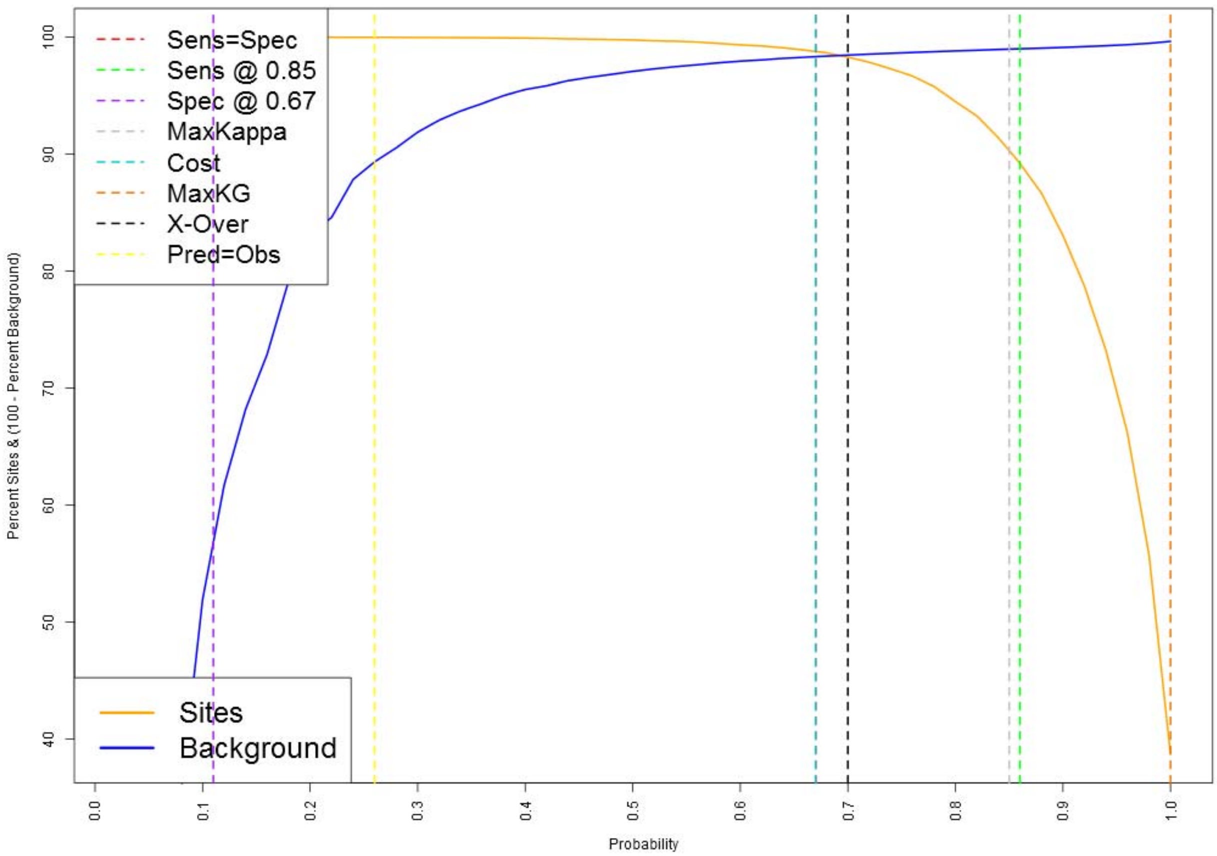
Region 9-10 All – Riverine Section 12



Region 9-10 All – Riverine Section 13

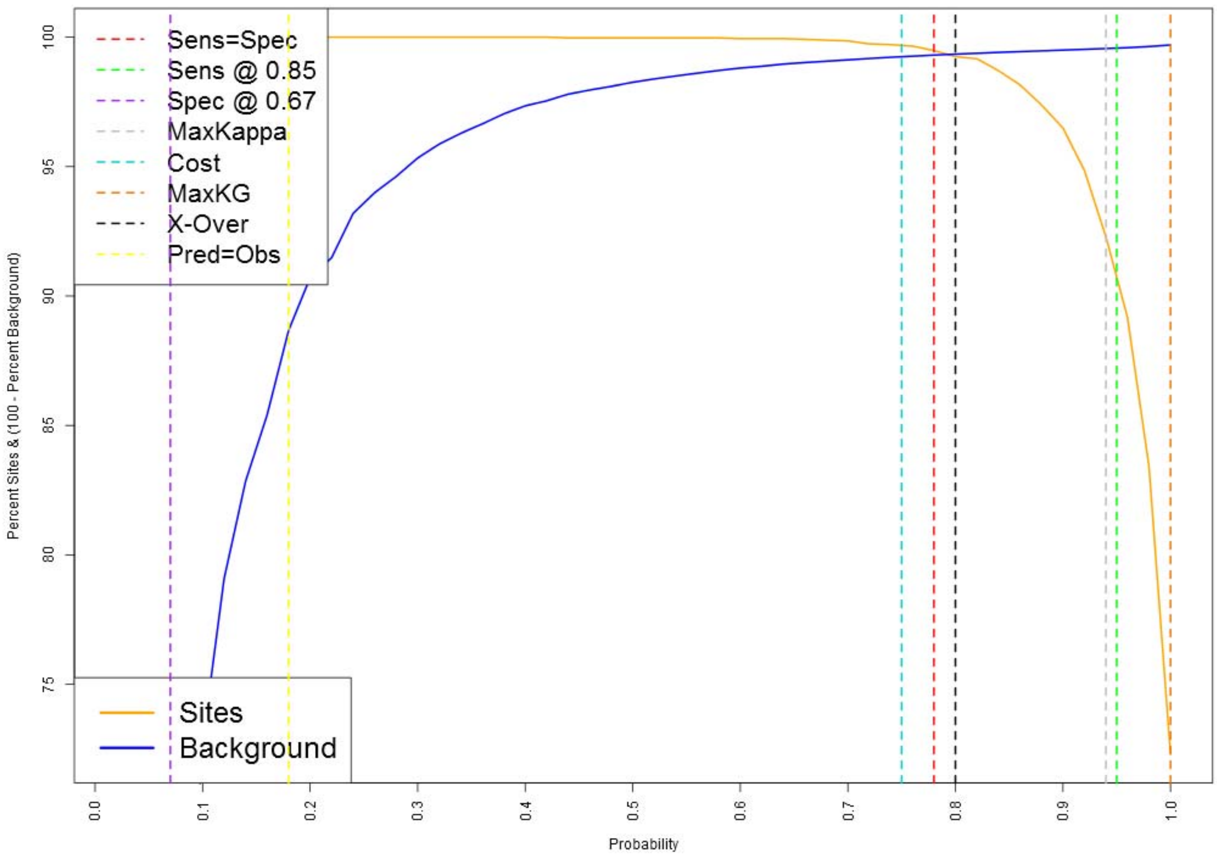


Region 9-10 All – Riverine Section 14

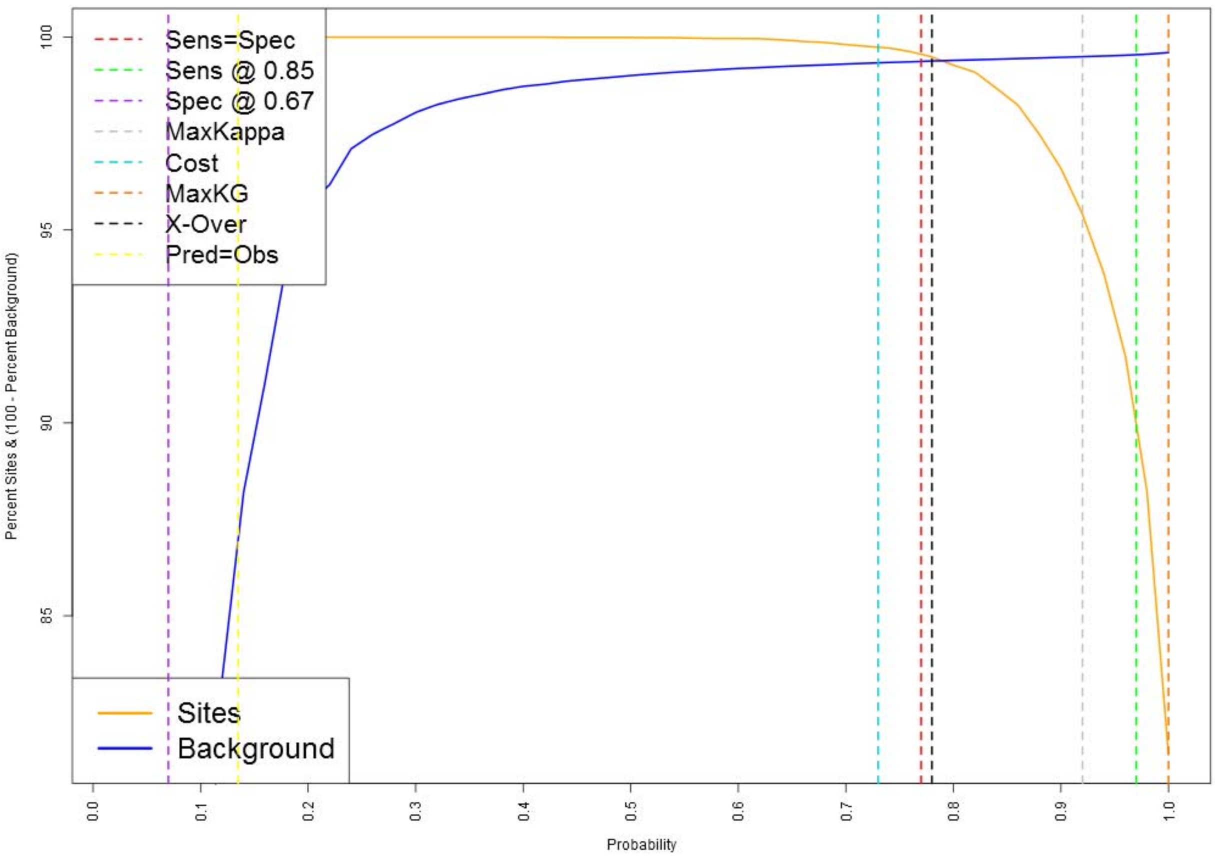


Note: The line for Sens=Spec is underneath the line for Cost and is not visible because the values are identical (0.67).

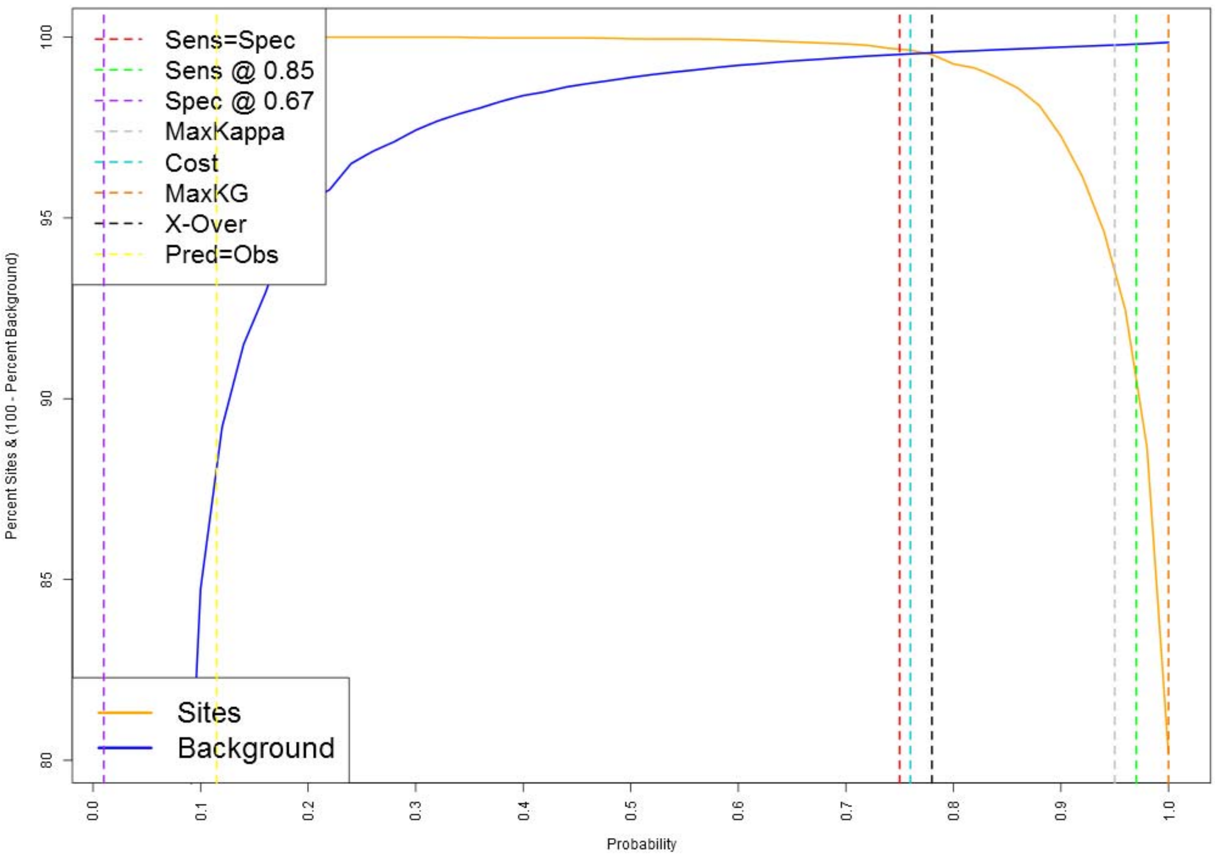
Region 9-10 All – Riverine Section 15



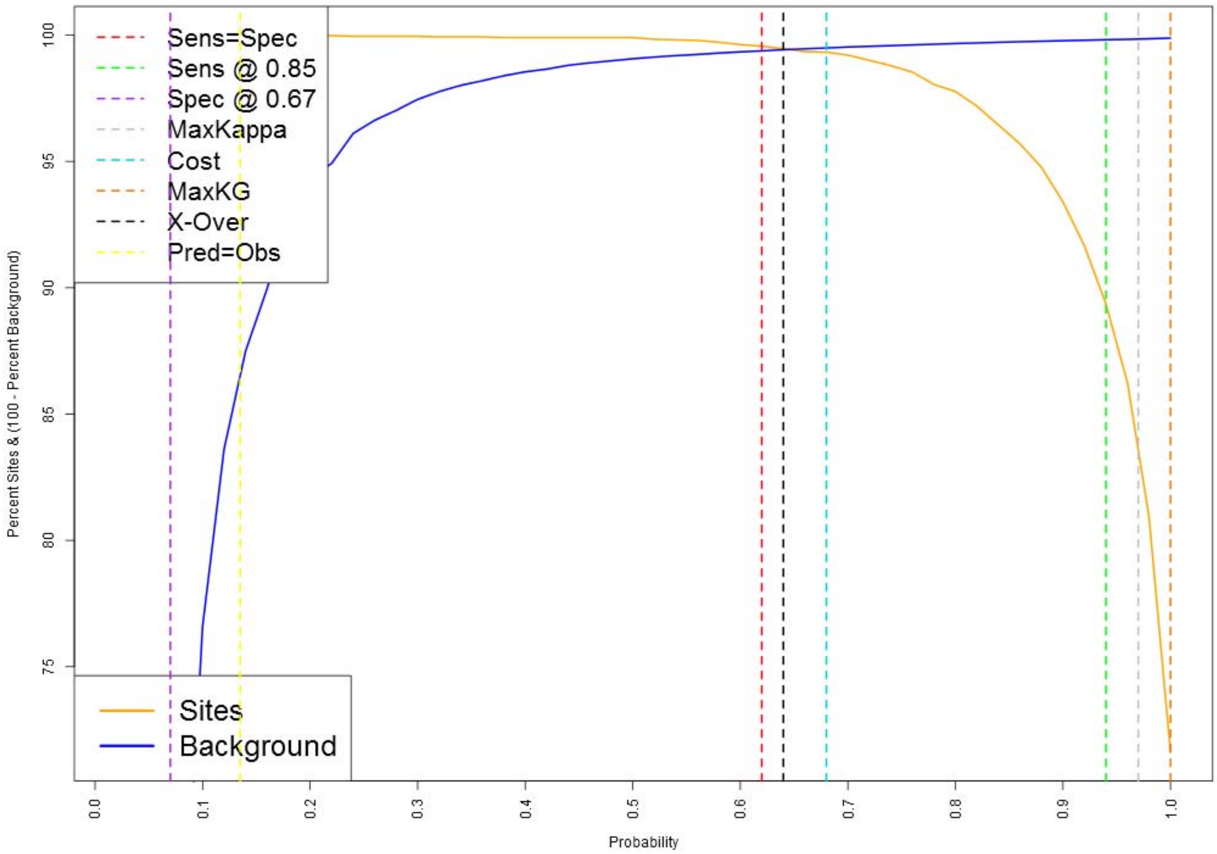
Region 9-10 All – Upland Section 1



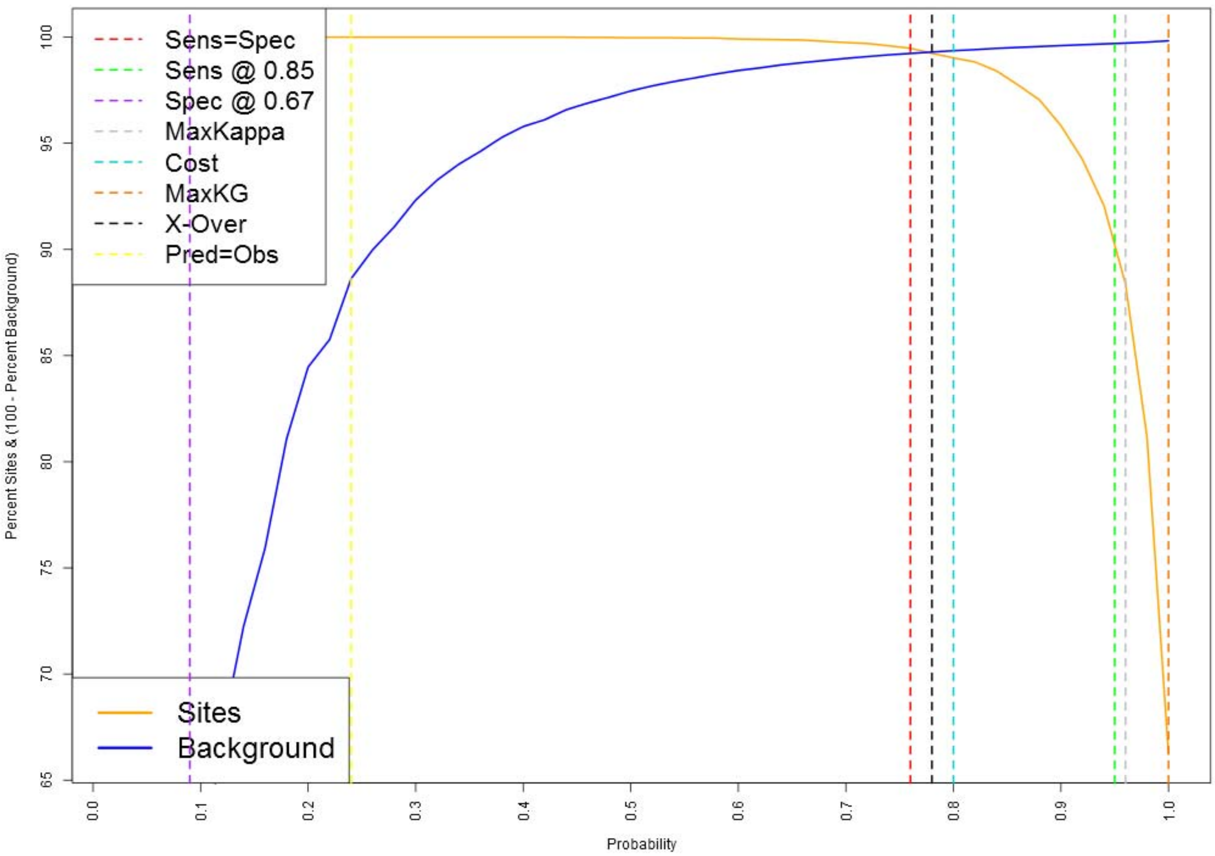
Region 9-10 All – Upland Section 2



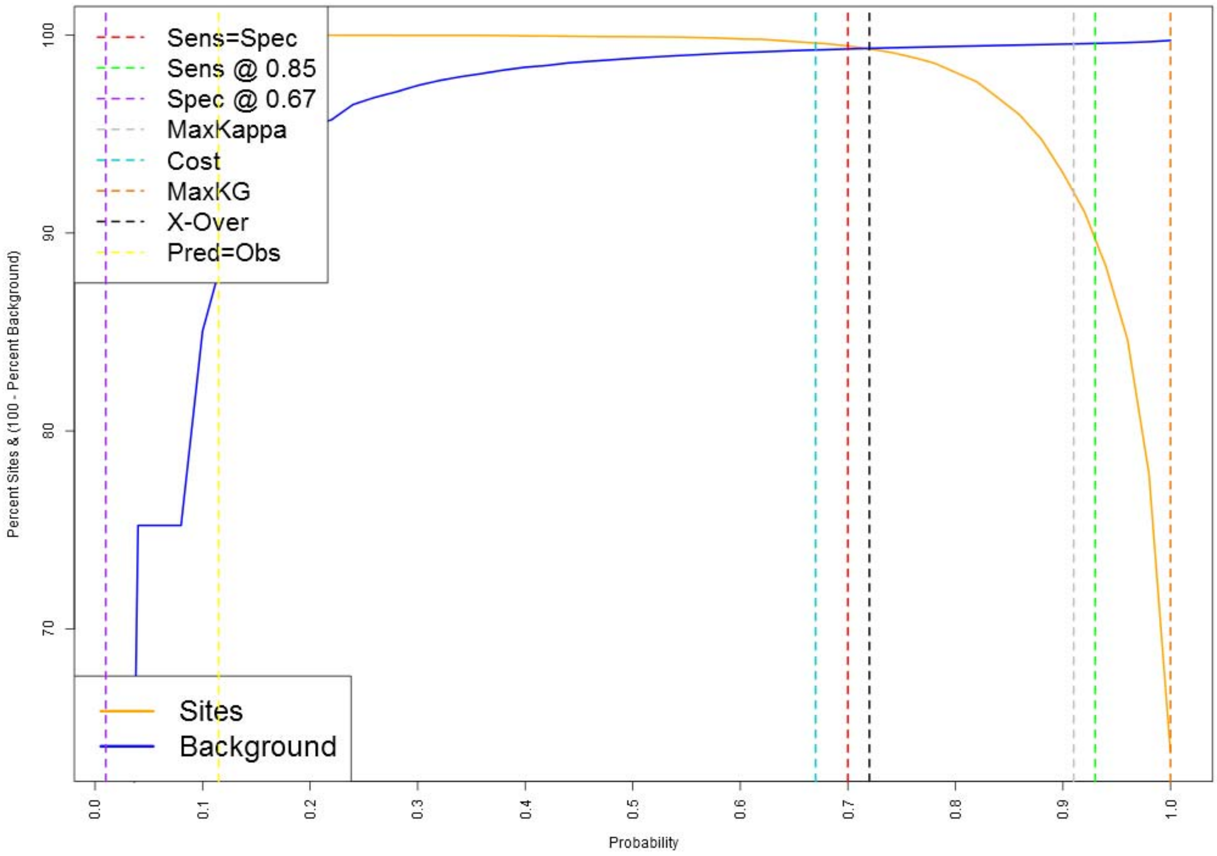
Region 9-10 All – Upland Section 3



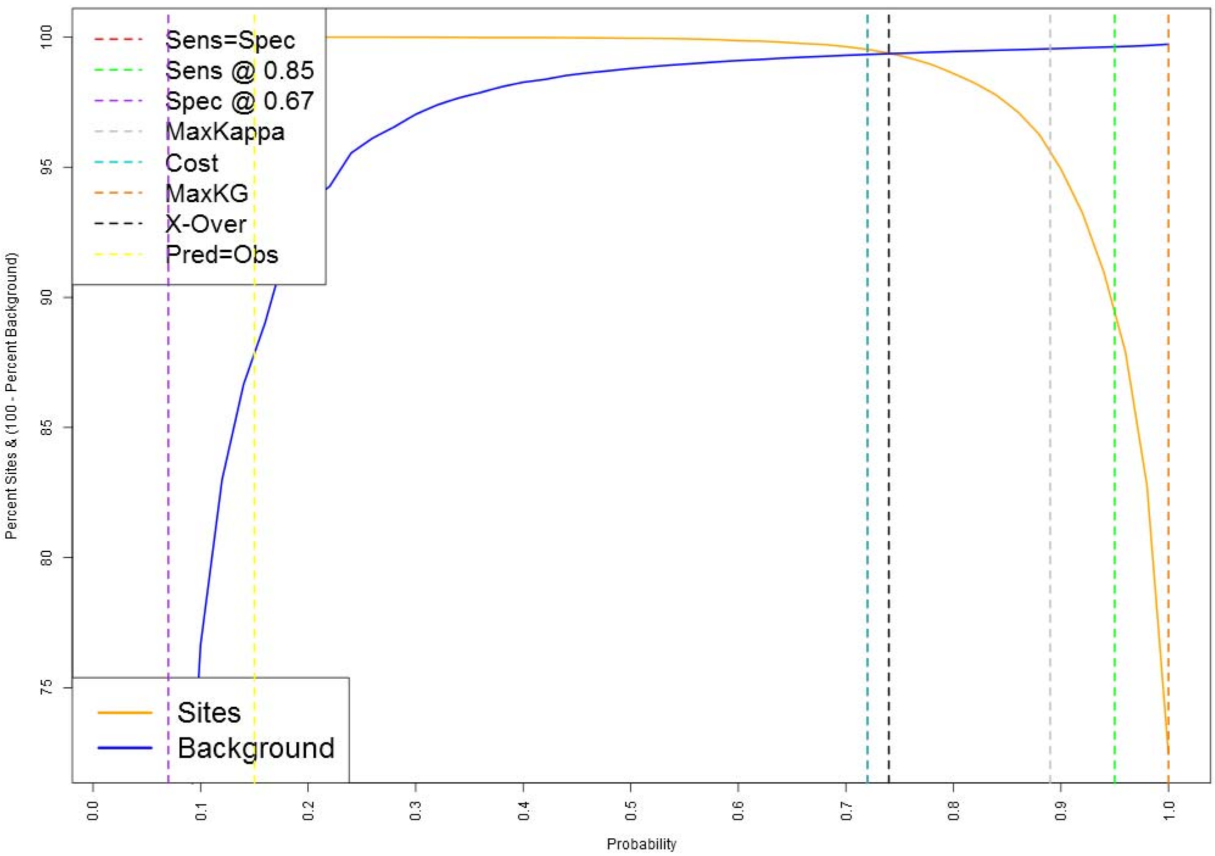
Region 9-10 All – Upland Section 4



Region 9-10 All – Upland Section 5

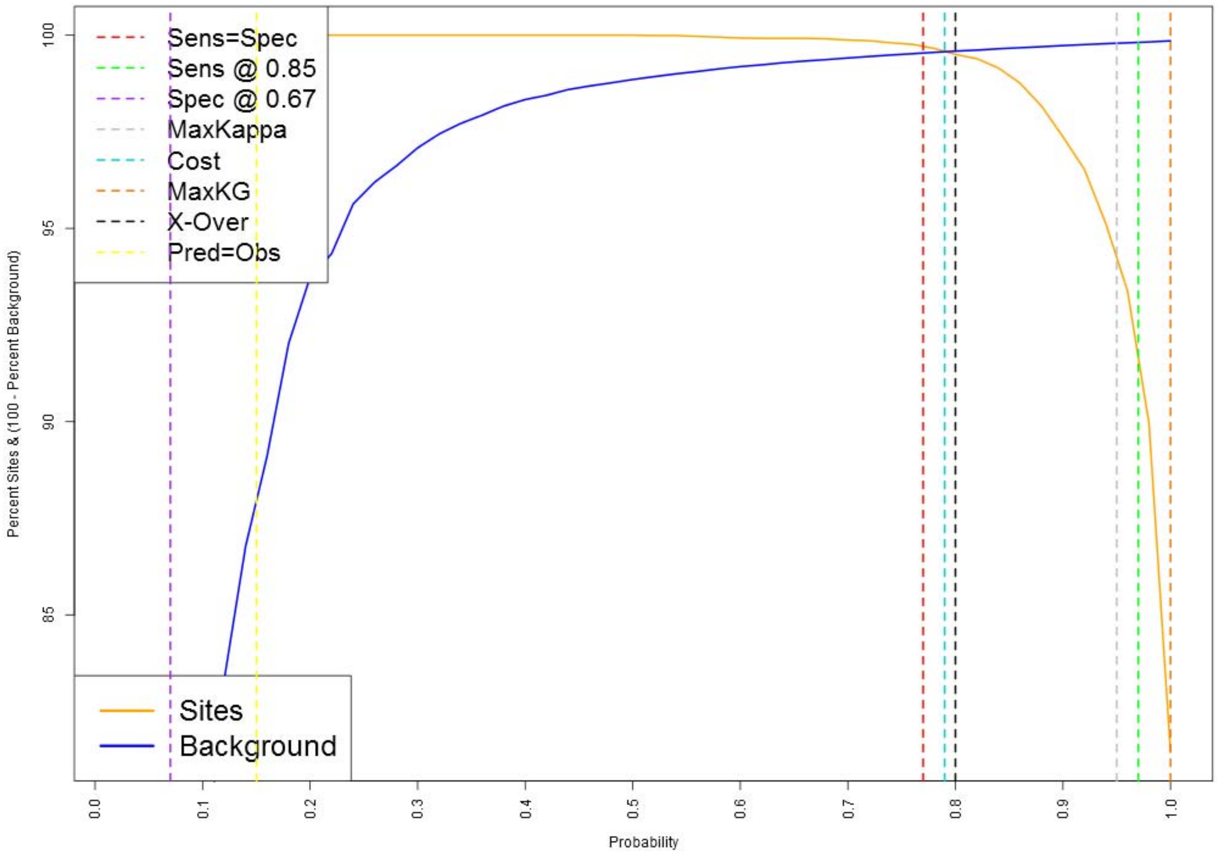


Region 9-10 All – Upland Section 6

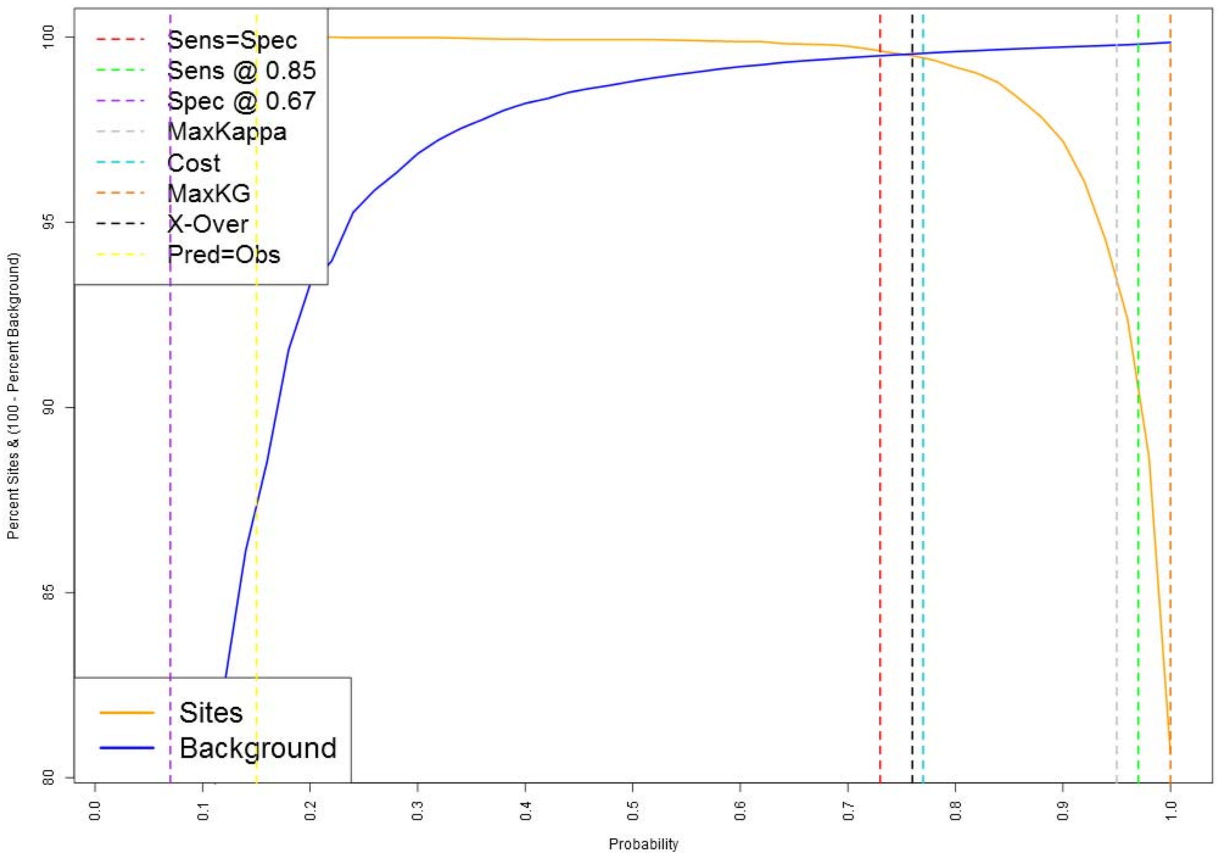


Note: The line for Sens=Spec is underneath the line for Cost and is not visible because the values are identical (0.72).

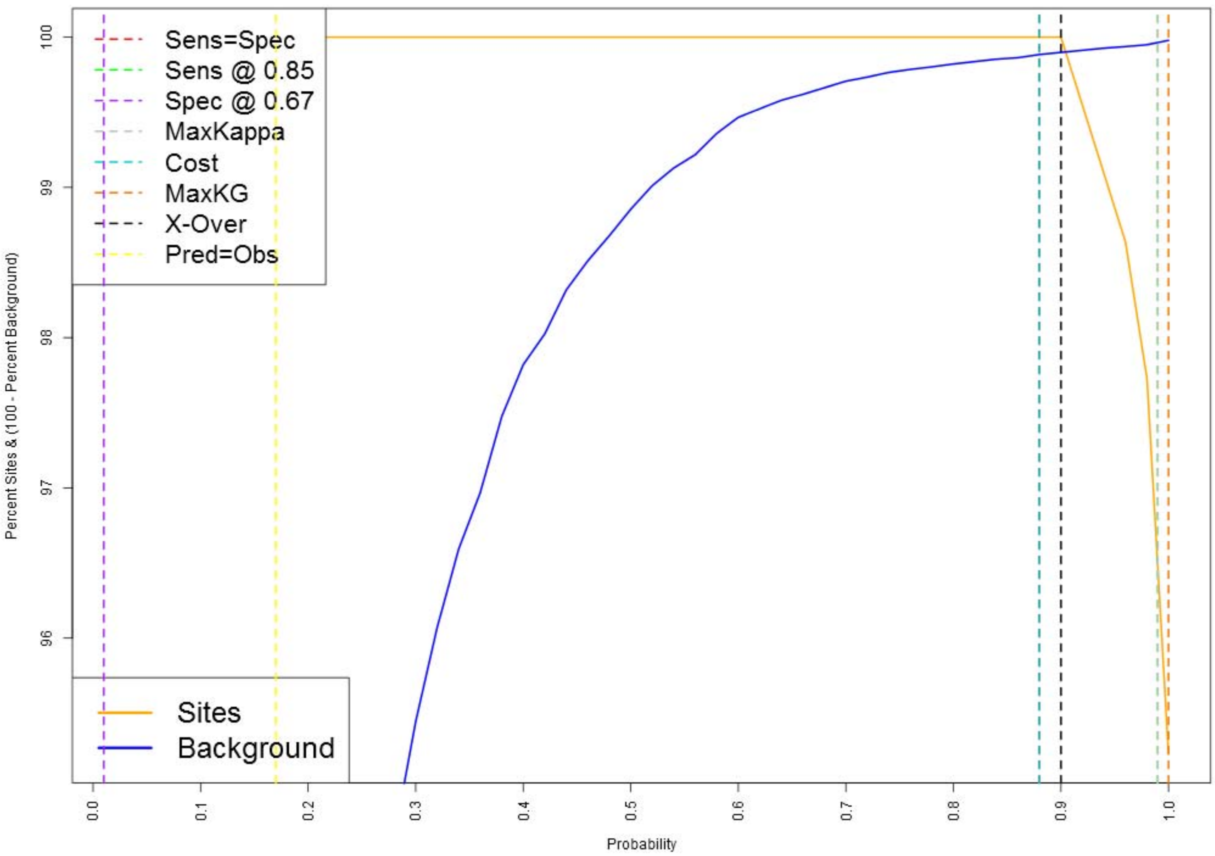
Region 9-10 All – Upland Section 7



Region 9-10 All – Upland Section 8

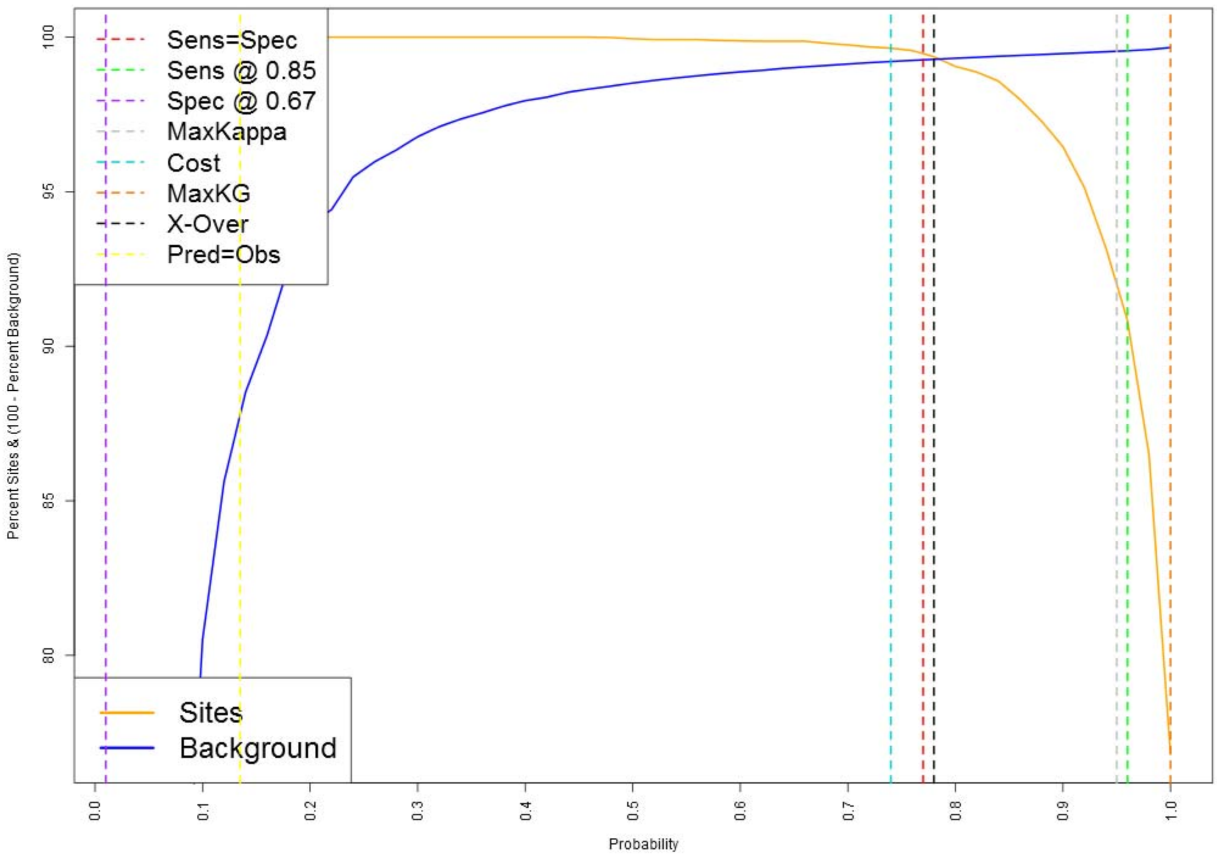


Region 9-10 All – Upland Section 9

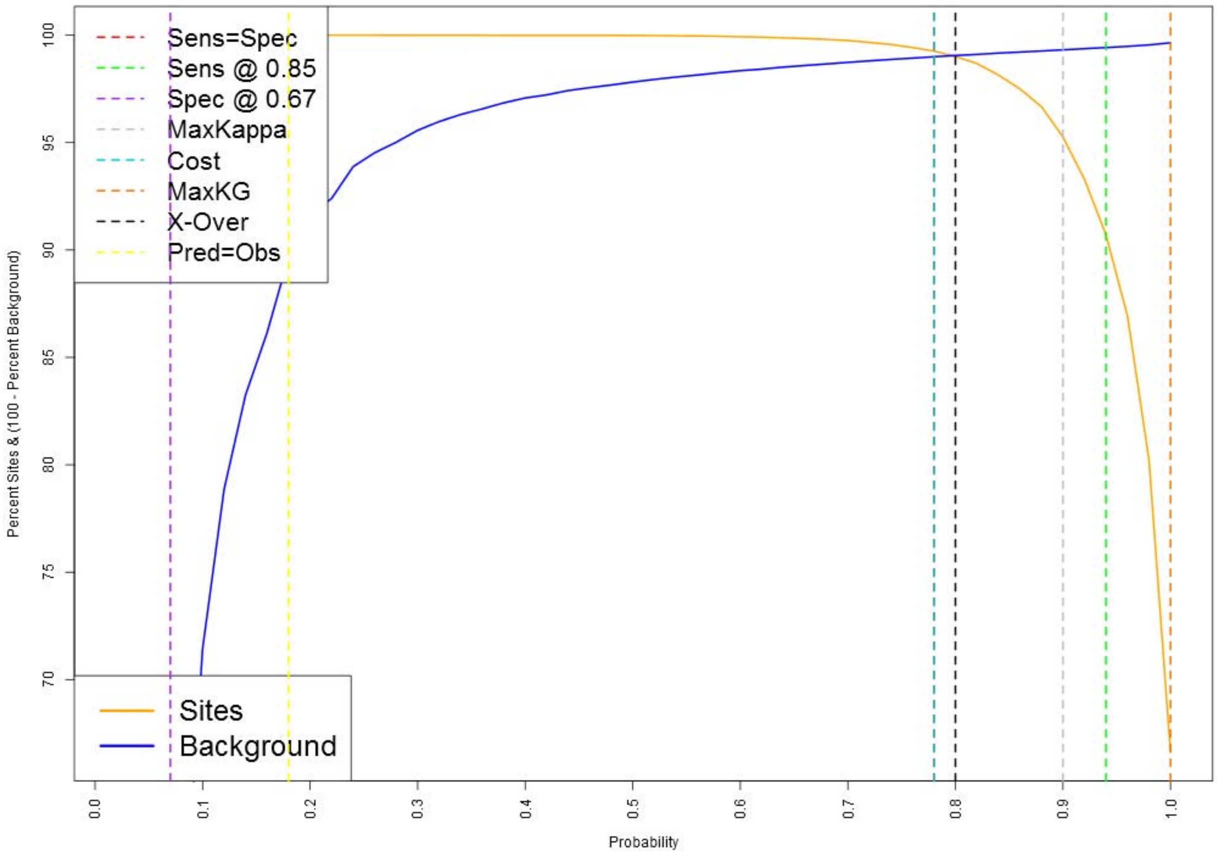


Note: The line for Sens @ 0.85 is underneath the line for MaxKappa and is not visible because the values are identical (0.99); similarly, the line for Sens=Spec is obscured by the line for Cost (0.88).

Region 9-10 All – Upland Section 10

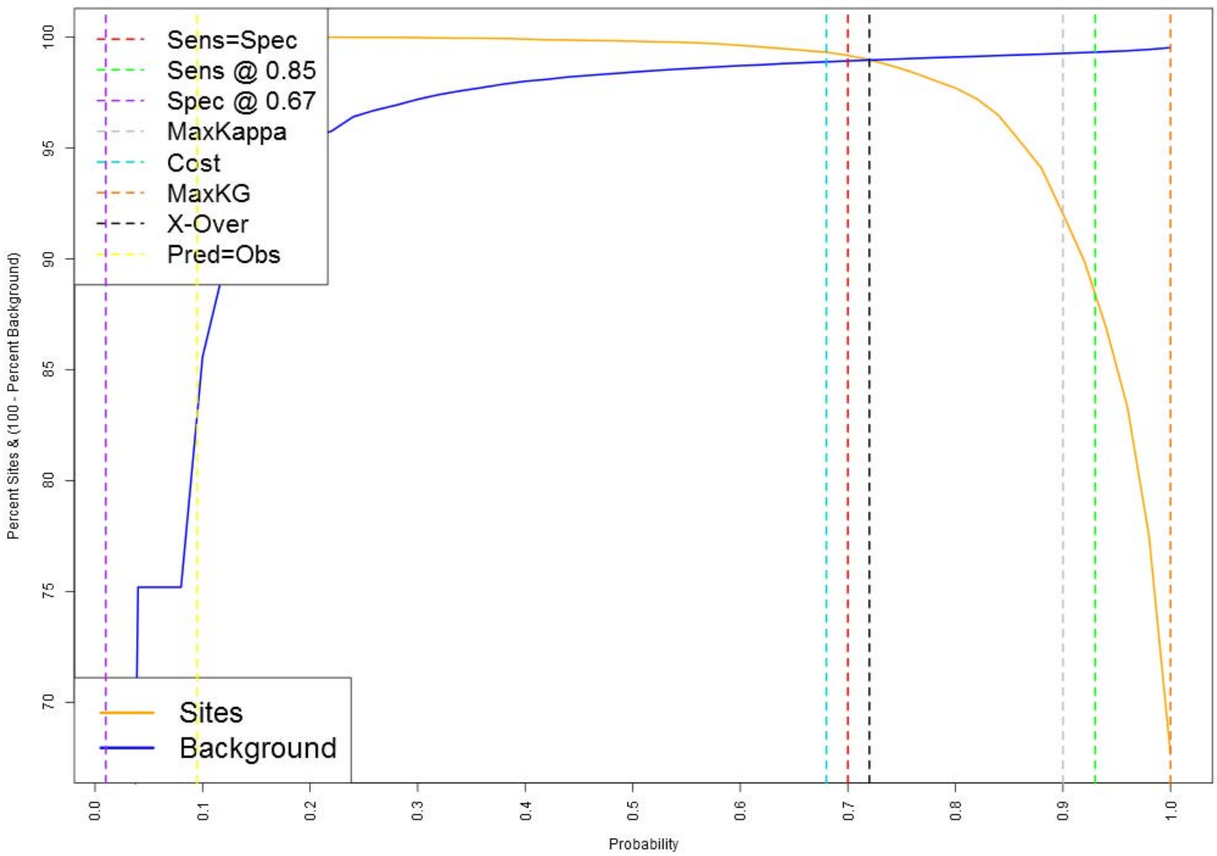


Region 9-10 All – Upland Section 11

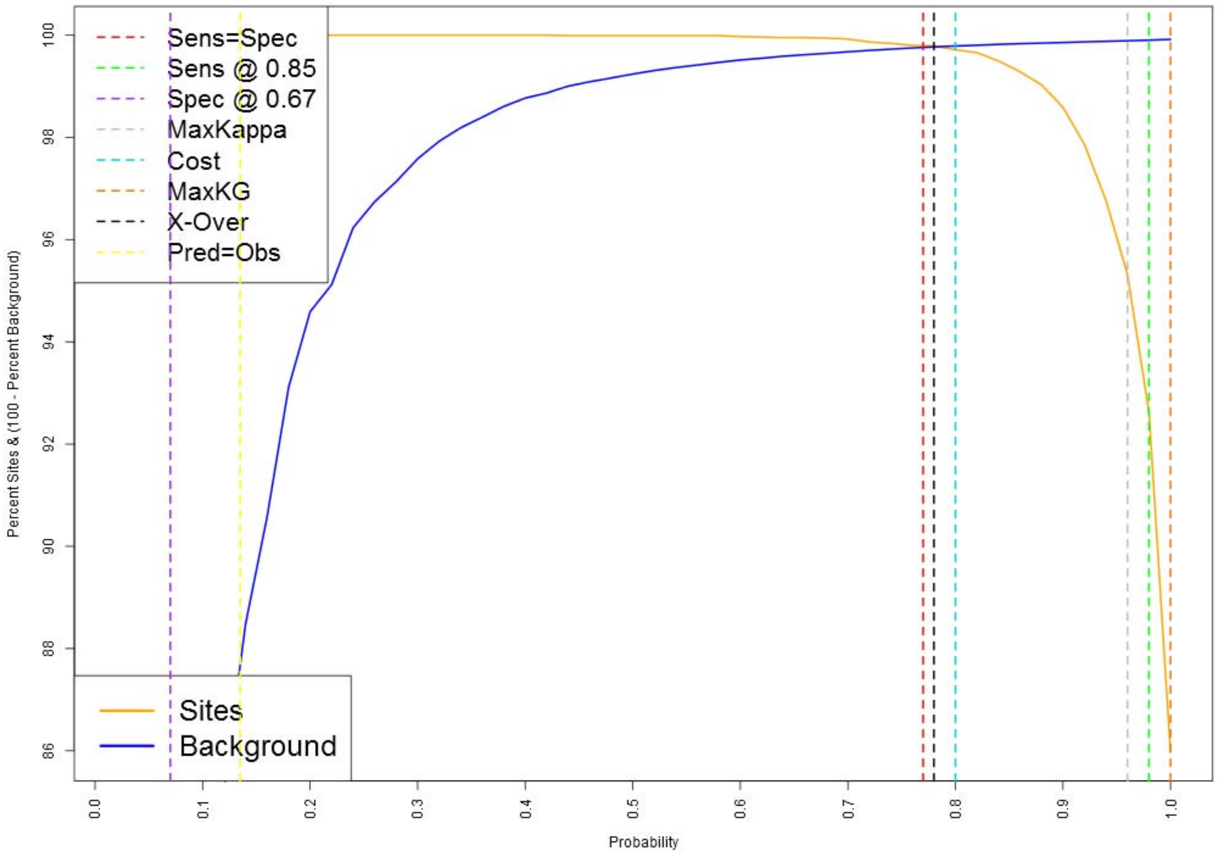


Note: The line for Sens=Spec is underneath the line for Cost and is not visible because the values are identical (0.78).

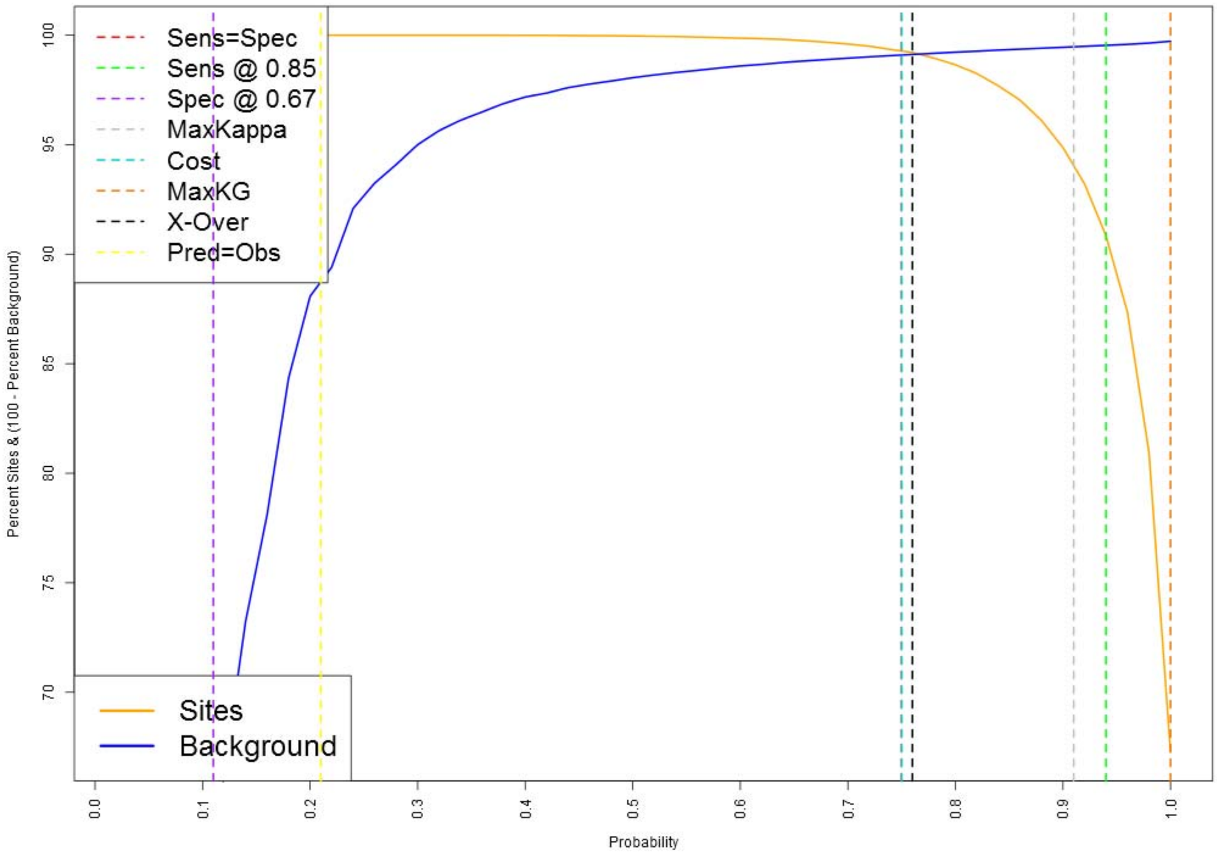
Region 9-10 All – Upland Section 12



Region 9-10 All – Upland Section 13

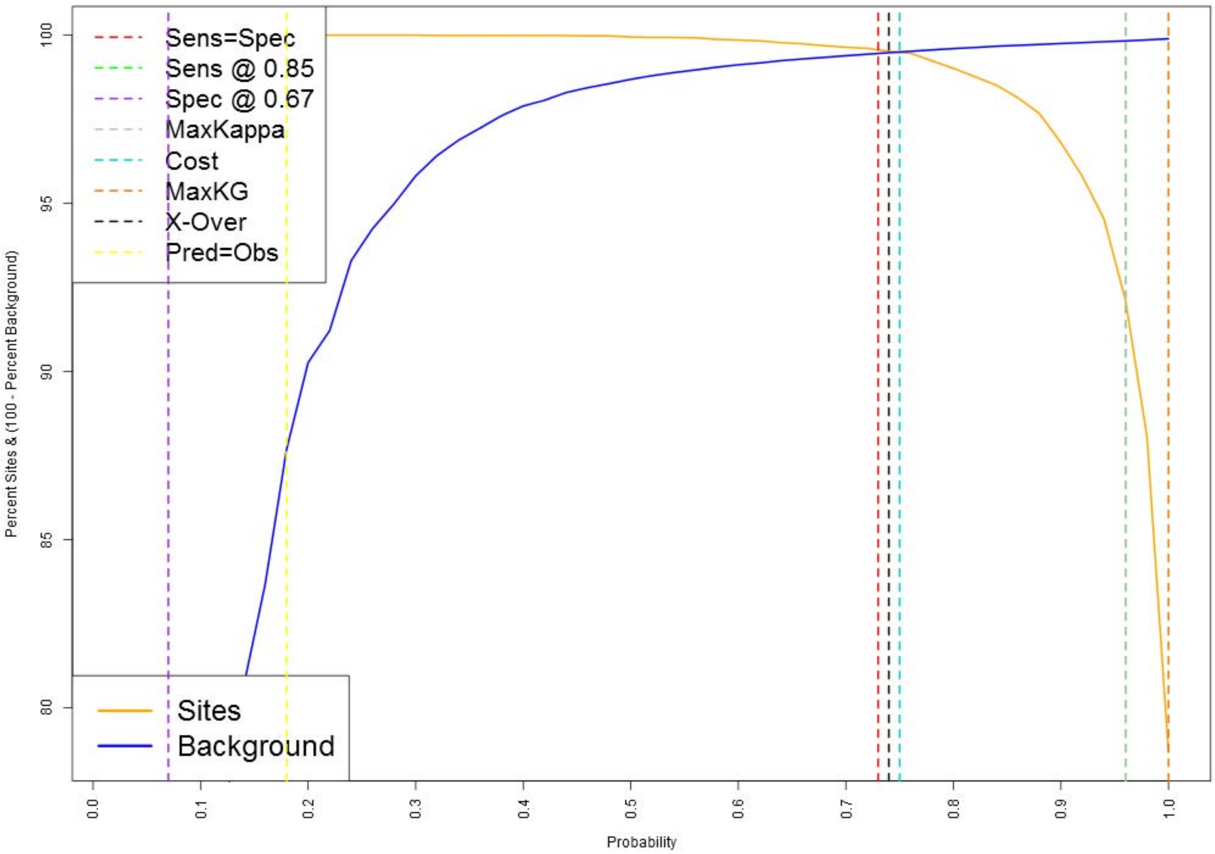


Region 9-10 All – Upland Section 14



Note: The line for Sens=Spec is underneath the line for Cost and is not visible because the values are identical (0.75).

Region 9-10 All – Upland Section 15



Note: The line for Sens @ 0.85 is underneath the line for MaxKappa and is not visible because the values are identical (0.96).

APPENDIX G

CONFUSION MATRICES

FOR EACH OF 66 MODELS

WITHIN REGIONS 7, 8, AND 9/10

Region 7 All - Riverine Section 1

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	39481	536819	576300
	Absent	1	1594274	1594275
		39482	2131093	2170575

Sensitivity / TPR =	1.000
Specificity / TNR =	0.748
Prevalence =	0.0182
Kvamme Gain (Kg) =	0.734
Accuracy =	0.753
Positive Prediction Value (PPV) =	0.069
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.018
Positive Prediction Gain (PPG) =	3.766
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.266

Region7 All - Riverine Section 2

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	6269	355723	361992
	Absent	0	938808	938808
		6269	1294531	1300800

Sensitivity / TPR =	1.000
Specificity / TNR =	0.725
Prevalence =	0.0048
Kvamme Gain (Kg) =	0.722
Accuracy =	0.727
Positive Prediction Value (PPV) =	0.017
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.005
Positive Prediction Gain (PPG) =	3.593
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.278

Region 7 All - Riverine Section 3

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	27405	248558	275963
	Absent	0	875974	875974
		27405	1124532	1151937

Sensitivity / TPR =	1.000
Specificity / TNR =	0.779
Prevalence =	0.0238
Kvamme Gain (Kg) =	0.760
Accuracy =	0.784
Positive Prediction Value (PPV) =	0.099
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.024
Positive Prediction Gain (PPG) =	4.174
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.240

Region 7 All - Riverine Section 4

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	9923	233008	242931
Model Prediction	Absent	0	713565	713565
		9923	946573	956496

Sensitivity / TPR =	1.000
Specificity / TNR =	0.754
Prevalence =	0.0104
Kvamme Gain (Kg) =	0.746
Accuracy =	0.756
Positive Prediction Value (PPV) =	0.041
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.010
Positive Prediction Gain (PPG) =	3.937
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.254

Region 7 All - Riverine Section 5

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	11281	370551	381832
	Absent	0	803099	803099
		11281	1173650	1184931

Sensitivity / TPR =	1.000
Specificity / TNR =	0.684
Prevalence =	0.0095
Kvamme Gain (Kg) =	0.678
Accuracy =	0.687
Positive Prediction Value (PPV) =	0.030
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.010
Positive Prediction Gain (PPG) =	3.103
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.322

Region 7 All - Riverine Section 6

		Known Sites		
		Present	Absent	
Model Prediction	Present	10046	440657	450703
	Absent	52	967809	967861
		10098	1408466	1418564

Sensitivity / TPR =	0.995
Specificity / TNR =	0.687
Prevalence =	0.0071
Kvamme Gain (Kg) =	0.681
Accuracy =	0.689
Positive Prediction Value (PPV) =	0.022
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.007
Positive Prediction Gain (PPG) =	3.131
Negative Prediction Gain (NPG) =	0.008
False Negative Rate (FNR) =	0.005
Detection Prevalence =	0.318

Region 7 All - Riverine Section 7

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	8660	345760	354420
	Absent	0	790603	790603
		8660	1136363	1145023

Sensitivity / TPR =	1.000
Specificity / TNR =	0.696
Prevalence =	0.0076
Kvamme Gain (Kg) =	0.690
Accuracy =	0.698
Positive Prediction Value (PPV) =	0.024
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.008
Positive Prediction Gain (PPG) =	3.231
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.310

Region 7 All - Riverine Section 8

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	3048	449042	452090
	Absent	51	959657	959708
		3099	1408699	1411798

Sensitivity / TPR =	0.984
Specificity / TNR =	0.681
Prevalence =	0.0022
Kvamme Gain (Kg) =	0.674
Accuracy =	0.682
Positive Prediction Value (PPV) =	0.007
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.002
Positive Prediction Gain (PPG) =	3.071
Negative Prediction Gain (NPG) =	0.024
False Negative Rate (FNR) =	0.016
Detection Prevalence =	0.320

Region 7 All - Riverine Section 9

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	5977	289297	295274
	Absent	0	642402	642402
		5977	931699	937676

Sensitivity / TPR =	1.000
Specificity / TNR =	0.689
Prevalence =	0.0064
Kvamme Gain (Kg) =	0.685
Accuracy =	0.691
Positive Prediction Value (PPV) =	0.020
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.006
Positive Prediction Gain (PPG) =	3.176
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.315

Region 7 All - Upland Section 1

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	9911	6819761	6829672
	Absent	359	15686579	15686938
		10270	22506340	22516610

Sensitivity / TPR =	0.965
Specificity / TNR =	0.697
Prevalence =	0.0005
Kvamme Gain (Kg) =	0.686
Accuracy =	0.697
Positive Prediction Value (PPV) =	0.001
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.000
Positive Prediction Gain (PPG) =	3.182
Negative Prediction Gain (NPG) =	0.050
False Negative Rate (FNR) =	0.035
Detection Prevalence =	0.303

Region 7 All - Upland Section 2

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	3249	1812816	1816065
	Absent	0	12256904	12256904
		3249	14069720	14072969

Sensitivity / TPR =	1.000
Specificity / TNR =	0.871
Prevalence =	0.0002
Kvamme Gain (Kg) =	0.871
Accuracy =	0.871
Positive Prediction Value (PPV) =	0.002
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.000
Positive Prediction Gain (PPG) =	7.749
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.129

Region 7 All - Upland Section 3

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	7510	3383251	3390761
	Absent	0	9249106	9249106
		7510	12632357	12639867

Sensitivity / TPR =	1.000
Specificity / TNR =	0.732
Prevalence =	0.0006
Kvamme Gain (Kg) =	0.732
Accuracy =	0.732
Positive Prediction Value (PPV) =	0.002
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.001
Positive Prediction Gain (PPG) =	3.728
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.268

Region 7 All - Upland Section 4

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	1820	2014825	2016645
	Absent	0	6393008	6393008
		1820	8407833	8409653

Sensitivity / TPR =	1.000
Specificity / TNR =	0.760
Prevalence =	0.0002
Kvamme Gain (Kg) =	0.760
Accuracy =	0.760
Positive Prediction Value (PPV) =	0.001
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.000
Positive Prediction Gain (PPG) =	4.170
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.240

Region 7 All - Upland Section 5

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	11018	2903709	2914727
	Absent	0	6761809	6761809
		11018	9665518	9676536

Sensitivity / TPR =	1.000
Specificity / TNR =	0.700
Prevalence =	0.0011
Kvamme Gain (Kg) =	0.699
Accuracy =	0.700
Positive Prediction Value (PPV) =	0.004
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.001
Positive Prediction Gain (PPG) =	3.320
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.301

Region 7 All - Upland Section 6

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	1122	3495225	3496347
	Absent	108	7596911	7597019
		1230	11092136	11093366

Sensitivity / TPR =	0.912
Specificity / TNR =	0.685
Prevalence =	0.0001
Kvamme Gain (Kg) =	0.654
Accuracy =	0.685
Positive Prediction Value (PPV) =	0.000
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.000
Positive Prediction Gain (PPG) =	2.894
Negative Prediction Gain (NPG) =	0.128
False Negative Rate (FNR) =	0.088
Detection Prevalence =	0.315

Region 7 All - Upland Section 7

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	2351	3330223	3332574
	Absent	0	7905534	7905534
		2351	11235757	11238108

Sensitivity / TPR =	1.000
Specificity / TNR =	0.704
Prevalence =	0.0002
Kvamme Gain (Kg) =	0.703
Accuracy =	0.704
Positive Prediction Value (PPV) =	0.001
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.000
Positive Prediction Gain (PPG) =	3.372
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.297

Region 7 All - Upland Section 8

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	1090	3463866	3464956
	Absent	8	7298529	7298537
		1098	10762395	10763493

Sensitivity / TPR =	0.993
Specificity / TNR =	0.678
Prevalence =	0.0001
Kvamme Gain (Kg) =	0.676
Accuracy =	0.678
Positive Prediction Value (PPV) =	0.000
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.000
Positive Prediction Gain (PPG) =	3.084
Negative Prediction Gain (NPG) =	0.011
False Negative Rate (FNR) =	0.007
Detection Prevalence =	0.322

Region 7 All - Upland Section 9

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	820	2604021	2604841
	Absent	0	5959020	5959020
		820	8563041	8563861

Sensitivity / TPR =	1.000
Specificity / TNR =	0.696
Prevalence =	0.0001
Kvamme Gain (Kg) =	0.696
Accuracy =	0.696
Positive Prediction Value (PPV) =	0.000
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.000
Positive Prediction Gain (PPG) =	3.288
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.304

Region 8 All - Riverine Section 1

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	28382	572156	600538
Model Prediction	Absent	0	1358949	1358949
		28382	1931105	1959487

Sensitivity / TPR =	1.000
Specificity / TNR =	0.704
Prevalence =	0.0145
Kvamme Gain (Kg) =	0.694
Accuracy =	0.708
Positive Prediction Value (PPV) =	0.047
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.014
Positive Prediction Gain (PPG) =	3.263
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.306

Region 8 All - Riverine Section 2

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	1792	66052	67844
	Absent	134	152283	152417
		1926	218335	220261

Sensitivity / TPR =	0.930
Specificity / TNR =	0.697
Prevalence =	0.0087
Kvamme Gain (Kg) =	0.669
Accuracy =	0.700
Positive Prediction Value (PPV) =	0.026
Negative Prediction Value (NPV) =	0.999
Unexpected Discovery Rate (UDR) =	0.001
Detection Rate =	0.008
Positive Prediction Gain (PPG) =	3.021
Negative Prediction Gain (NPG) =	0.101
False Negative Rate (FNR) =	0.070
Detection Prevalence =	0.308

Region 8 All - Riverine Section 3

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	2208	94918	97126
	Absent	24	230333	230357
		2232	325251	327483

Sensitivity / TPR =	0.989
Specificity / TNR =	0.708
Prevalence =	0.0068
Kvamme Gain (Kg) =	0.700
Accuracy =	0.710
Positive Prediction Value (PPV) =	0.023
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.007
Positive Prediction Gain (PPG) =	3.335
Negative Prediction Gain (NPG) =	0.015
False Negative Rate (FNR) =	0.011
Detection Prevalence =	0.297

Region 8 All - Riverine Section 4

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	20921	529924	550845
	Absent	0	1278019	1278019
		20921	1807943	1828864

Sensitivity / TPR =	1.000
Specificity / TNR =	0.707
Prevalence =	0.0114
Kvamme Gain (Kg) =	0.699
Accuracy =	0.710
Positive Prediction Value (PPV) =	0.038
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.011
Positive Prediction Gain (PPG) =	3.320
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.301

Region 8 All - Riverine Section 5

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	19555	401022	420577
	Absent	0	827010	827010
		19555	1228032	1247587

Sensitivity / TPR =	1.000
Specificity / TNR =	0.673
Prevalence =	0.0157
Kvamme Gain (Kg) =	0.663
Accuracy =	0.679
Positive Prediction Value (PPV) =	0.046
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.016
Positive Prediction Gain (PPG) =	2.966
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.337

Region 8 All - Riverine Section 6

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	22062	309951	332013
Model Prediction	Absent	1	717284	717285
		22063	1027235	1049298

Sensitivity / TPR =	1.000
Specificity / TNR =	0.698
Prevalence =	0.0210
Kvamme Gain (Kg) =	0.684
Accuracy =	0.705
Positive Prediction Value (PPV) =	0.066
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.021
Positive Prediction Gain (PPG) =	3.160
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.316

Region 8 All - Riverine Section 7

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	11337	242349	253686
	Absent	16	498670	498686
		11353	741019	752372

Sensitivity / TPR =	0.999
Specificity / TNR =	0.673
Prevalence =	0.0151
Kvamme Gain (Kg) =	0.662
Accuracy =	0.678
Positive Prediction Value (PPV) =	0.045
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.015
Positive Prediction Gain (PPG) =	2.962
Negative Prediction Gain (NPG) =	0.002
False Negative Rate (FNR) =	0.001
Detection Prevalence =	0.337

Region 8 All - Riverine Section 8

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	23843	255537	279380
	Absent	25	522996	523021
		23868	778533	802401

Sensitivity / TPR =	0.999
Specificity / TNR =	0.672
Prevalence =	0.0297
Kvamme Gain (Kg) =	0.651
Accuracy =	0.682
Positive Prediction Value (PPV) =	0.085
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.030
Positive Prediction Gain (PPG) =	2.869
Negative Prediction Gain (NPG) =	0.002
False Negative Rate (FNR) =	0.001
Detection Prevalence =	0.348

Region 8 All - Riverine Section 9

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	34271	396440	430711
	Absent	1	1063889	1063890
		34272	1460329	1494601

Sensitivity / TPR =	1.000
Specificity / TNR =	0.729
Prevalence =	0.0229
Kvamme Gain (Kg) =	0.712
Accuracy =	0.735
Positive Prediction Value (PPV) =	0.080
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.023
Positive Prediction Gain (PPG) =	3.470
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.288

Region 8 All - Upland Section 1

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	21629	3728292	3749921
	Absent	0	11668075	11668075
		21629	15396367	15417996

Sensitivity / TPR =	1.000
Specificity / TNR =	0.758
Prevalence =	0.0014
Kvamme Gain (Kg) =	0.757
Accuracy =	0.758
Positive Prediction Value (PPV) =	0.006
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.001
Positive Prediction Gain (PPG) =	4.112
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.243

Region 8 All - Upland Section 2

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	25368	1615778	1641146
	Absent	0	3413211	3413211
		25368	5028989	5054357

Sensitivity / TPR =	1.000
Specificity / TNR =	0.679
Prevalence =	0.0050
Kvamme Gain (Kg) =	0.675
Accuracy =	0.680
Positive Prediction Value (PPV) =	0.015
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.005
Positive Prediction Gain (PPG) =	3.080
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.325

Region 8 All - Upland Section 3

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	5007	1364913	1369920
	Absent	0	5125704	5125704
		5007	6490617	6495624

Sensitivity / TPR =	1.000
Specificity / TNR =	0.790
Prevalence =	0.0008
Kvamme Gain (Kg) =	0.789
Accuracy =	0.790
Positive Prediction Value (PPV) =	0.004
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.001
Positive Prediction Gain (PPG) =	4.742
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.211

Region 8 All - Upland Section 4

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	18500	2751892	2770392
	Absent	0	8202032	8202032
		18500	10953924	10972424

Sensitivity / TPR =	1.000
Specificity / TNR =	0.749
Prevalence =	0.0017
Kvamme Gain (Kg) =	0.748
Accuracy =	0.749
Positive Prediction Value (PPV) =	0.007
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.002
Positive Prediction Gain (PPG) =	3.961
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.252

Region 8 All - Upland Section 5

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	16780	2396096	2412876
Model Prediction	Absent	0	5247893	5247893
		16780	7643989	7660769

Sensitivity / TPR =	1.000
Specificity / TNR =	0.687
Prevalence =	0.0022
Kvamme Gain (Kg) =	0.685
Accuracy =	0.687
Positive Prediction Value (PPV) =	0.007
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.002
Positive Prediction Gain (PPG) =	3.175
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.315

Region 8 All - Upland Section 6

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	18561	1941148	1959709
	Absent	0	4524980	4524980
		18561	6466128	6484689

Sensitivity / TPR =	1.000
Specificity / TNR =	0.700
Prevalence =	0.0029
Kvamme Gain (Kg) =	0.698
Accuracy =	0.701
Positive Prediction Value (PPV) =	0.009
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.003
Positive Prediction Gain (PPG) =	3.309
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.302

Region 8 All - Upland Section 7

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	15123	1512994	1528117
	Absent	0	3224822	3224822
		15123	4737816	4752939

Sensitivity / TPR =	1.000
Specificity / TNR =	0.681
Prevalence =	0.0032
Kvamme Gain (Kg) =	0.678
Accuracy =	0.682
Positive Prediction Value (PPV) =	0.010
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.003
Positive Prediction Gain (PPG) =	3.110
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.322

Region 8 All - Upland Section 8

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	62986	1215790	1278776
	Absent	0	4513700	4513700
		62986	5729490	5792476

Sensitivity / TPR =	1.000
Specificity / TNR =	0.788
Prevalence =	0.0109
Kvamme Gain (Kg) =	0.779
Accuracy =	0.790
Positive Prediction Value (PPV) =	0.049
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.011
Positive Prediction Gain (PPG) =	4.530
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.221

Region 8 All - Upland Section 9

		Known Sites		
		Present	Absent	
Model Prediction	Present	55394	2381656	2437050
	Absent	0	9684075	9684075
		55394	12065731	12121125

Sensitivity / TPR =	1.000
Specificity / TNR =	0.803
Prevalence =	0.0046
Kvamme Gain (Kg) =	0.799
Accuracy =	0.804
Positive Prediction Value (PPV) =	0.023
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.005
Positive Prediction Gain (PPG) =	4.974
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.201

Region 9/10 All - Riverine Section 1

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	8736	283130	291866
Model Prediction	Absent	1454	578401	579855
		10190	861531	871721

Sensitivity / TPR =	0.857
Specificity / TNR =	0.671
Prevalence =	0.0117
Kvamme Gain (Kg) =	0.609
Accuracy =	0.674
Positive Prediction Value (PPV) =	0.030
Negative Prediction Value (NPV) =	0.997
Unexpected Discovery Rate (UDR) =	0.003
Detection Rate =	0.010
Positive Prediction Gain (PPG) =	2.561
Negative Prediction Gain (NPG) =	0.215
False Negative Rate (FNR) =	0.143
Detection Prevalence =	0.335

Region 9/10 All - Riverine Section 2

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	27386	558561	585947
	Absent	4238	1143852	1148090
		31624	1702413	1734037

Sensitivity / TPR =	0.866
Specificity / TNR =	0.672
Prevalence =	0.0182
Kvamme Gain (Kg) =	0.610
Accuracy =	0.675
Positive Prediction Value (PPV) =	0.047
Negative Prediction Value (NPV) =	0.996
Unexpected Discovery Rate (UDR) =	0.004
Detection Rate =	0.016
Positive Prediction Gain (PPG) =	2.563
Negative Prediction Gain (NPG) =	0.202
False Negative Rate (FNR) =	0.134
Detection Prevalence =	0.338

Region 9/10 All - Riverine Section 3

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	10967	198032	208999
	Absent	0	442731	442731
		10967	640763	651730

Sensitivity / TPR =	1.000
Specificity / TNR =	0.691
Prevalence =	0.0168
Kvamme Gain (Kg) =	0.679
Accuracy =	0.696
Positive Prediction Value (PPV) =	0.052
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.017
Positive Prediction Gain (PPG) =	3.118
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.321

Region 9/10 All - Riverine Section 4

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	957	64134	65091
	Absent	0	153052	153052
		957	217186	218143

Sensitivity / TPR =	1.000
Specificity / TNR =	0.705
Prevalence =	0.0044
Kvamme Gain (Kg) =	0.702
Accuracy =	0.706
Positive Prediction Value (PPV) =	0.015
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.004
Positive Prediction Gain (PPG) =	3.351
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.298

Region 9/10 All - Riverine Section 5

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	25405	266340	291745
	Absent	5927	551647	557574
		31332	817987	849319

Sensitivity / TPR =	0.811
Specificity / TNR =	0.674
Prevalence =	0.0369
Kvamme Gain (Kg) =	0.576
Accuracy =	0.679
Positive Prediction Value (PPV) =	0.087
Negative Prediction Value (NPV) =	0.989
Unexpected Discovery Rate (UDR) =	0.011
Detection Rate =	0.030
Positive Prediction Gain (PPG) =	2.360
Negative Prediction Gain (NPG) =	0.288
False Negative Rate (FNR) =	0.189
Detection Prevalence =	0.344

Region 9/10 All - Riverine Section 6

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	15892	313975	329867
	Absent	538	639692	640230
		16430	953667	970097

Sensitivity / TPR =	0.967
Specificity / TNR =	0.671
Prevalence =	0.0169
Kvamme Gain (Kg) =	0.648
Accuracy =	0.676
Positive Prediction Value (PPV) =	0.048
Negative Prediction Value (NPV) =	0.999
Unexpected Discovery Rate (UDR) =	0.001
Detection Rate =	0.016
Positive Prediction Gain (PPG) =	2.845
Negative Prediction Gain (NPG) =	0.050
False Negative Rate (FNR) =	0.033
Detection Prevalence =	0.340

Region 9/10 All - Riverine Section 7

		Known Sites		
		Present	Absent	
Model Prediction	Present	9113	293906	303019
	Absent	780	611118	611898
		9893	905024	914917

Sensitivity / TPR =	0.921
Specificity / TNR =	0.675
Prevalence =	0.0108
Kvamme Gain (Kg) =	0.640
Accuracy =	0.678
Positive Prediction Value (PPV) =	0.030
Negative Prediction Value (NPV) =	0.999
Unexpected Discovery Rate (UDR) =	0.001
Detection Rate =	0.010
Positive Prediction Gain (PPG) =	2.781
Negative Prediction Gain (NPG) =	0.118
False Negative Rate (FNR) =	0.079
Detection Prevalence =	0.331

Region 9/10 All - Riverine Section 8

		Known Sites		
		Present	Absent	
Model Prediction	Present	5508	292367	297875
	Absent	0	595263	595263
		5508	887630	893138

Sensitivity / TPR =	1.000
Specificity / TNR =	0.671
Prevalence =	0.0062
Kvamme Gain (Kg) =	0.666
Accuracy =	0.673
Positive Prediction Value (PPV) =	0.018
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.006
Positive Prediction Gain (PPG) =	2.998
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.334

Region 9/10 All - Riverine Section 9

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	2151	492482	494633
	Absent	0	1273253	1273253
		2151	1765735	1767886

Sensitivity / TPR =	1.000
Specificity / TNR =	0.721
Prevalence =	0.0012
Kvamme Gain (Kg) =	0.720
Accuracy =	0.721
Positive Prediction Value (PPV) =	0.004
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.001
Positive Prediction Gain (PPG) =	3.574
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.280

Region 9/10 All - Riverine Section 10

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	10153	147060	157213
Model Prediction	Absent	0	313542	313542
		10153	460602	470755

Sensitivity / TPR =	1.000
Specificity / TNR =	0.681
Prevalence =	0.0216
Kvamme Gain (Kg) =	0.666
Accuracy =	0.688
Positive Prediction Value (PPV) =	0.065
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.022
Positive Prediction Gain (PPG) =	2.994
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.334

Region 9/10 All - Riverine Section 11

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	15429	547026	562455
	Absent	0	1198681	1198681
		15429	1745707	1761136

Sensitivity / TPR =	1.000
Specificity / TNR =	0.687
Prevalence =	0.0088
Kvamme Gain (Kg) =	0.681
Accuracy =	0.689
Positive Prediction Value (PPV) =	0.027
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.009
Positive Prediction Gain (PPG) =	3.131
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.319

Region 9/10 All - Riverine Section 12

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	4886	113606	118492
	Absent	606	238490	239096
		5492	352096	357588

Sensitivity / TPR =	0.890
Specificity / TNR =	0.677
Prevalence =	0.0154
Kvamme Gain (Kg) =	0.628
Accuracy =	0.681
Positive Prediction Value (PPV) =	0.041
Negative Prediction Value (NPV) =	0.997
Unexpected Discovery Rate (UDR) =	0.003
Detection Rate =	0.014
Positive Prediction Gain (PPG) =	2.685
Negative Prediction Gain (NPG) =	0.165
False Negative Rate (FNR) =	0.110
Detection Prevalence =	0.331

Region 9/10 All - Riverine Section 13

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	6013	424461	430474
	Absent	0	989669	989669
		6013	1414130	1420143

Sensitivity / TPR =	1.000
Specificity / TNR =	0.700
Prevalence =	0.0042
Kvamme Gain (Kg) =	0.697
Accuracy =	0.701
Positive Prediction Value (PPV) =	0.014
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.004
Positive Prediction Gain (PPG) =	3.299
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.303

Region 9/10 All - Riverine Section 14

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	15436	541790	557226
Model Prediction	Absent	0	1161420	1161420
		15436	1703210	1718646

Sensitivity / TPR =	1.000
Specificity / TNR =	0.682
Prevalence =	0.0090
Kvamme Gain (Kg) =	0.676
Accuracy =	0.685
Positive Prediction Value (PPV) =	0.028
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.009
Positive Prediction Gain (PPG) =	3.084
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.324

Region 9/10 All - Riverine Section 15

		Known Sites		
		Present	Absent	
Model Prediction	Present	3473	240377	243850
	Absent	0	641391	641391
		3473	881768	885241

Sensitivity / TPR =	1.000
Specificity / TNR =	0.727
Prevalence =	0.0039
Kvamme Gain (Kg) =	0.725
Accuracy =	0.728
Positive Prediction Value (PPV) =	0.014
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.004
Positive Prediction Gain (PPG) =	3.630
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.275

Region 9/10 All - Upland Section 1

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	14632	782437	797069
	Absent	0	2270569	2270569
		14632	3053006	3067638

Sensitivity / TPR =	1.000
Specificity / TNR =	0.744
Prevalence =	0.0048
Kvamme Gain (Kg) =	0.740
Accuracy =	0.745
Positive Prediction Value (PPV) =	0.018
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.005
Positive Prediction Gain (PPG) =	3.849
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.260

Region 9/10 All - Upland Section 2

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	13336	2388926	2402262
	Absent	0	6228123	6228123
		13336	8617049	8630385

Sensitivity / TPR =	1.000
Specificity / TNR =	0.723
Prevalence =	0.0015
Kvamme Gain (Kg) =	0.722
Accuracy =	0.723
Positive Prediction Value (PPV) =	0.006
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.002
Positive Prediction Gain (PPG) =	3.593
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.278

Region 9/10 All - Upland Section 3

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	4301	792031	796332
	Absent	0	2591782	2591782
		4301	3383813	3388114

Sensitivity / TPR =	1.000
Specificity / TNR =	0.766
Prevalence =	0.0013
Kvamme Gain (Kg) =	0.765
Accuracy =	0.766
Positive Prediction Value (PPV) =	0.005
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.001
Positive Prediction Gain (PPG) =	4.255
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.235

Region 9/10 All - Upland Section 4

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	6664	976701	983365
	Absent	0	2000635	2000635
		6664	2977336	2984000

Sensitivity / TPR =	1.000
Specificity / TNR =	0.672
Prevalence =	0.0022
Kvamme Gain (Kg) =	0.670
Accuracy =	0.673
Positive Prediction Value (PPV) =	0.007
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.002
Positive Prediction Gain (PPG) =	3.034
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.330

Region 9/10 All - Upland Section 5

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	19985	1232617	1252602
	Absent	0	3742455	3742455
		19985	4975072	4995057

Sensitivity / TPR =	1.000
Specificity / TNR =	0.752
Prevalence =	0.0040
Kvamme Gain (Kg) =	0.749
Accuracy =	0.753
Positive Prediction Value (PPV) =	0.016
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.004
Positive Prediction Gain (PPG) =	3.988
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.251

Region 9/10 All - Upland Section 6

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	23566	1540341	1563907
	Absent	0	5055624	5055624
		23566	6595965	6619531

Sensitivity / TPR =	1.000
Specificity / TNR =	0.766
Prevalence =	0.0036
Kvamme Gain (Kg) =	0.764
Accuracy =	0.767
Positive Prediction Value (PPV) =	0.015
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.004
Positive Prediction Gain (PPG) =	4.233
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.236

Region 9/10 All - Upland Section 7

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	9658	1345552	1355210
	Absent	0	4635188	4635188
		9658	5980740	5990398

Sensitivity / TPR =	1.000
Specificity / TNR =	0.775
Prevalence =	0.0016
Kvamme Gain (Kg) =	0.774
Accuracy =	0.775
Positive Prediction Value (PPV) =	0.007
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.002
Positive Prediction Gain (PPG) =	4.420
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.226

Region 9/10 All - Upland Section 8

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	7399	1222618	1230017
	Absent	0	3911497	3911497
		7399	5134115	5141514

Sensitivity / TPR =	1.000
Specificity / TNR =	0.762
Prevalence =	0.0014
Kvamme Gain (Kg) =	0.761
Accuracy =	0.762
Positive Prediction Value (PPV) =	0.006
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.001
Positive Prediction Gain (PPG) =	4.180
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.239

Region 9/10 All - Upland Section 9

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	440	986488	986928
	Absent	0	2476417	2476417
		440	3462905	3463345

Sensitivity / TPR =	1.000
Specificity / TNR =	0.715
Prevalence =	0.0001
Kvamme Gain (Kg) =	0.715
Accuracy =	0.715
Positive Prediction Value (PPV) =	0.000
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.000
Positive Prediction Gain (PPG) =	3.509
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.285

Region 9/10 All - Upland Section 10

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	7591	577405	584996
	Absent	0	1263913	1263913
		7591	1841318	1848909

Sensitivity / TPR =	1.000
Specificity / TNR =	0.686
Prevalence =	0.0041
Kvamme Gain (Kg) =	0.684
Accuracy =	0.688
Positive Prediction Value (PPV) =	0.013
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.004
Positive Prediction Gain (PPG) =	3.161
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.316

Region 9/10 All - Upland Section 11

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	48491	2732166	2780657
	Absent	0	6829859	6829859
		48491	9562025	9610516

Sensitivity / TPR =	1.000
Specificity / TNR =	0.714
Prevalence =	0.0050
Kvamme Gain (Kg) =	0.711
Accuracy =	0.716
Positive Prediction Value (PPV) =	0.017
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.005
Positive Prediction Gain (PPG) =	3.456
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.289

Region 9/10 All - Upland Section 12

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	9983	374340	384323
	Absent	0	1134547	1134547
		9983	1508887	1518870

Sensitivity / TPR =	1.000
Specificity / TNR =	0.752
Prevalence =	0.0066
Kvamme Gain (Kg) =	0.747
Accuracy =	0.754
Positive Prediction Value (PPV) =	0.026
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.007
Positive Prediction Gain (PPG) =	3.952
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.253

Region 9/10 All - Upland Section 13

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	10485	2555153	2565638
	Absent	0	9992625	9992625
		10485	12547778	12558263

Sensitivity / TPR =	1.000
Specificity / TNR =	0.796
Prevalence =	0.0008
Kvamme Gain (Kg) =	0.796
Accuracy =	0.797
Positive Prediction Value (PPV) =	0.004
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.001
Positive Prediction Gain (PPG) =	4.895
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.204

Region 9/10 All - Upland Section 14

		Known Sites: Present	Known Sites: Absent	
Model Prediction	Present	61548	4351256	4412804
	Absent	0	11907055	11907055
		61548	16258311	16319859

Sensitivity / TPR =	1.000
Specificity / TNR =	0.732
Prevalence =	0.0038
Kvamme Gain (Kg) =	0.730
Accuracy =	0.733
Positive Prediction Value (PPV) =	0.014
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.004
Positive Prediction Gain (PPG) =	3.698
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.270

Region 9/10 All - Upland Section 15

		Known Sites		
		Present	Absent	
Model Prediction	Present	8910	2537300	2546210
	Absent	0	5741775	5741775
		8910	8279075	8287985

Sensitivity / TPR =	1.000
Specificity / TNR =	0.694
Prevalence =	0.0011
Kvamme Gain (Kg) =	0.693
Accuracy =	0.694
Positive Prediction Value (PPV) =	0.003
Negative Prediction Value (NPV) =	1.000
Unexpected Discovery Rate (UDR) =	0.000
Detection Rate =	0.001
Positive Prediction Gain (PPG) =	3.255
Negative Prediction Gain (NPG) =	0.000
False Negative Rate (FNR) =	0.000
Detection Prevalence =	0.307