Project Report

Title: The application of GIS, Remote Sensing techniques to field telemetry data to model and map sage-grouse (*Centrocercus* spp.) seasonal habitat-use on Seep Ridge and East Bench and other areas in Utah: Implications for Guiding Sustainable Energy Development in the West.

Proposal Type: Applied Research and Student Internship

Personnel: Terry A. Messmer and Todd A. Black, Quinney Professorship for Wildlife Damage Management, Jack. H. Berryman Institute, and Utah Community-Based Conservation Program. Utah State University, Logan, Utah 84322-5230

Chris McGinty and Ben Crabb, RS/GIS Laboratory, College of Natural Resources 5275 Old Main Hill Utah State University Logan, Utah 84322-5275

Beginning Date: July 1, 2010

Project Duration: March 31, 2011

Project Summary: Additional funding was needed to support a student and technician to model and map field telemetry data and to validate seasonal habitat use patterns for greater sage-grouse populations inhabiting Seep Ridge and East Bench in northeastern Utah. This information will assist the Utah Division of Wildlife Resources, the Bureau of Land Management and corporations, such as Anadarko Petroleum, in prioritizing energy development actions and activities which will contribute to economic sustainability in the region and species conservation. The methodology developed by this project will be applied across Utah to assist local sage-grouse working groups (LWGs), state and federal agencies, industry, and rural communities in making management decisions to mitigate sage-grouse threats. This research will result in a portable decision tool based on science-based models that will help LWGs range wide to better identify sage-grouse seasonal habitats population needs, predict suitable unoccupied habitat, and the effects of proposed energy development and conservation actions on local populations.

Executive Summary

Habitat loss and fragmentation have been implicated as the major cause for observed range wide declines in greater sage-grouse (*C. urophasianus*) populations. The U.S. Fish and Wildlife Service (USFWS) recently issued a “Warranted but Precluded” decision in response to petitions filed to list the species as threatened or endangered under the Endangered Species Act. One of the new threats identified by the USFWS in their decision was the unknown impacts of energy development on the species.

Declining greater sage-grouse populations exhibit low recruitment which is believed to be an artifact of poor brooding-rearing habitat quality. To date, much time and effort have been spent to map sage-grouse distributions in Utah and across its range. These distribution maps have been primarily developed using Delphi approaches by various agencies leading to different published renditions. The maps suffer from issues of scale and often were not developed using the best available science. Thus their accuracy in mapping sage-grouse habitat use pattern remains suspect. These cumulative efforts have created additional confusion among land management agencies, wildlife managers, private landowners, and industry regarding actual sage-grouse distribution and conservation priorities. Having access to an accurate science-based map will be crucial should the USFWS change the species designation from “candidate,” to threatened or endangered.

To mitigate this situation, we developed and test a habitat-use model incorporating greater sage-grouse movement data collected from field telemetry studies completed of the Seep Ridge and East Bench populations in northeastern Utah. We applied the same methodology to field data for the other sage-grouse populations across Utah. For the past decade, researchers across the state of Utah have been collecting detailed information on sage-grouse ecology including nesting, brooding, and locations using radio telemetry. These site-specific data enabled us to conduct a landscape level habitat analysis never before possible.

We used existing Geographical Information Systems (GIS) data layers project to further refine sage-grouse seasonal habitat for the state of Utah. Furthermore, we used existing brood-rearing data to classify and define the ecological amplitudes for occupied sites and match these sites to existing sage-grouse habitat maps. This model was then used to determine and prioritize brood rearing sites across the state for conservation purposes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Background

Current sage-grouse distribution maps for both the state of Utah and range wide were primarily developed by using data on the historical and current sagebrush habitat distribution maps. Research has shown that not all sagebrush constitutes suitable sage-grouse habitat and that different sagebrush species are more heavily used by sage-grouse throughout their seasonal movements. These historic sagebrush distribution maps have shown a substantial loss of sagebrush and are quite often used by researchers, land management agencies, and the environmental communities to show as the primary reason sage-grouse population have declined, because they correlated with the sagebrush disappearance.

To address this discrepancy, managers have attempted to modify sagebrush distribution maps to better identify sage-grouse potentials using habitat-use data collected from known locations of radio-collared grouse. Subsequently, alternative maps have been developed that attempt to better identify sage-grouse historical and contemporary distributions. However, because these mapping efforts employed different methodologies and scales, the information generated may be problematic for managers working at multiple scales. While these maps may provide a good 'big picture' of range wide habitat loss issues, the scales used may not be practical or portable for managers and local working groups who are developing projects to address habitat needs and threats of sage-grouse at a local level. Additionally, the mere fact that several different renditions of these maps exist further adds to the confusion of local and state managers.

In Utah, the initial sage-grouse habitat maps were developed by Utah Division of Wildlife Resource (UDWR) in the mid 1990's using primarily using anecdotal information and a Delphi approach. These maps are very broad in scope and often encompass smaller micro- habitats that are not typical of sage-grouse habitat. As such the UDWR and the Local Working Groups (LWGs) have identified a critical management need to have more detailed seasonal habitat maps (GIS layers) to better assist them in making management decisions to address specific threats and opportunities to mitigate or improve habitat and sage-grouse populations in these focus areas. Further, once these base layers are created managers can further model and address threats to sage-grouse as identified in these plans such as the decline of sagebrush, habitat fragmentation, connectivity of habitats, and the connectivity of existing populations.

The Community-Based Conservation Program at Utah State University (USU) has been asked by the Utah Division of Wildlife Resource (UDWR) and the Local Working Groups (LWGs) to assist in developing a portable process to prioritize sage-grouse conservation actions. For the past twelve years researchers from USU, Brigham Young University, and the UDWR have been collecting critical ecological and habitat data from radio-collared sage-grouse from eleven unique populations across the state of Utah. Concomitantly, they have complied thousands of Global Positioning System (GPS) locations of radio-collared grouse with associated habitat, vegetation, and other geographic data. However, to date little has been done to analyze these data to both map habitat and address threats to sage-grouse at different scales.

To initiate this process, we will use greater sage-grouse habitat-use location data collected for the Seep Ridge and East Bench populations to develop and validate a primary model (pilot study). We will than apply this methodology to sage-grouse location data collected from ten different sites in Utah and compile it into one database. Once this process is completed, we convert these data into usable GIS data layers and develop predicative presence/absence models and other population and habitat models for continued research to address specific sage-grouse threats.

Objectives

The objectives of this project are:

1. To provide Anadarko Petroleum, the UDWR, and BLM with a portable greater sage-grouse seasonal habitat-use model to guide energy development and species conservation on Seep Ridge and East Bench.

2. To use the methodology developed under objective 1 to provide the Utah's ten LWGs, UDWR and other land managers with GIS data layers designed to assist them in their efforts to make the best science-based management decisions.

Methods

Data collection

We compiled greater sage-grouse flush locations along with corresponding habitat and vegetation data from 14 different radio telemetry sites across the state of Utah. Upon reviewing the field data sheets from multiple research projects, much effort was needed to organize the data into one central database. Research efforts spanned 14 different research sites covering a time frame of 14 years and 30 different research projects (Table 1). Computer technicians at USU and BYU spent over 400 hours compiling, organizing, standardizing, and extracting the data into one central data base. These field data represented over 6,000 grouse telemetry locations. From these locations, we divided the data into the following tabs in the database; capture (sites where the grouse were captured), nesting (nest sites and habitat data), brood rearing (brood sites and habitat data), and non-brooding (male and non-brooding hen sites and also random bird locations and habitat data) (Table 2). Once the data was in a usable format, the database was sent to the USU RS/GIS lab for GIS analysis for initial model development and analysis. The area selected for a pilot/proof of concept was the East Bench area of Uintah County, Utah, which also corresponded with the Anadarko Petroleum EIS.

Table 1. Study sites across the state of Utah indicating the number of years research is being conducted on these sites.

Area	Utah County	Date	End date	years	projects	holder
Strawberry Valley	Wasatch/Utah/ Duchesne	2000	on going	11	5	BYU
Parker Mountain	Wayne/Piute	1998	on going	13	7	USU
West Box Elder County	Box Elder	2003	on going	8	2	USU
Alton Hoyt's Ranch	Kane/Garfield	2005	on going	7	2	USU
West Desert	Tooele	2005	on going	5	1	USU
East Bench	Uintah	2007	2008	2	1	USU
Emma Park West Taviputs	Carbon	2005	on going	10	2	UDWR
Anthro Mountain	Duchesne/Car-bon	2002	on going	10	2	USFS USU
Rich County	Rich	2002	on going	10	3	DLL USU UDWR
Wildcat & Horn Mountains	Emery/Utah/ Sevier	2007	2009	3	1	USU
Hamblin Valley	Iron/Beaver	2010		1	1	USU
Bald Hills	Beaver/Iron/ Millard	2009		2	1	USU
Diamond Mountain	Uintah	2009	on going	3	1	BYU

Table 2. Examples from the Access database, shown below are nest site data (Nesting), brood site data (Brooding), non-brooding hens and male grouse data (Non-Brood), and capture data (Captures).

Nesting									
Unique_id	**Date_ST**	**Capture_Year**	**Status**	**Nest_Success**	**Vegetation**	**Age**	**UTM_E**	**UTM_N**	**Sex_Female**
AM-02-148.768	5/14/2002	2002	NESTING	Yes	ARTR2	A	551114	4417746	Yes
AM-02-148.857	5/8/2002	2002	NESTING	Yes	ARNO	A	545766	4417461	Yes

Brooding								
Unique_Id	**Date_ST**	**Capture_Year**	**Status**	**Brood_Success**	**Vegitation**	**Age**	**UTM_E**	**UTM_N**
GC-07-151.529	5/29/2007	2007	BROODING	Yes	BRMA4	C	269192	4634956
GC-07-151.530	5/17/2007	2007	BROODING	Yes	ARTR2	A	258805	4629388
GC-07-151.530	5/23/2007	2007	BROODING	Yes	JUOS	A	259434	4629148

Non-Brood							
Unique_Id	**Date**	**Capture_Year**	**UTM_E**	**UTM_N**	**Habitat_Type**	**Sex**	**Status**
EB-07-149.442	6/11/2007	2007	625267	4411131	ARTR2	F	NON BROODING
EB-07-149.442	4/30/2007	2007	625397	4412854	ARTR2	U	NON BROODING
EB-07-149.442	5/7/2007	2007	625604	4413729	ARTR2	U	NON BROODING
EB-07-149.513	6/21/2007	2007	617114	4397179	ARNO	M	NON BROODING

Captures								
Unique_Id	**Date**	**Time**	**UTM_E**	**UTM_N**	**Sex**	**Age**	**Weight**	**Lek_Name**
WD-06-150.684	4/8/2006	1:45	356551	4435860	M	A	2700	SIMPSON SPRINGS
WD-06-151.031	4/9/2006	5:46	379860	4423444	F	A	1400	LITTLE VALLEY

The data and sage-grouse locations were collected in 2007 and 2008 in the southern portion of Uintah County, Utah, by Leah Smith as part of her Master's thesis at Utah State University (Smith, 2009). A total of 348 bird locations, representing 32 radio collared individuals and over 50 X bird locations (individual birds that were encountered at random while on the study site) were recorded with corresponding vegetation data at each flush location. Vegetation data were collected using standard protocol developed under the sage-grouse guidelines. Of the 348 total points collected 138 (40%) were located within the far southern portion of the Anadarko area of interest (AOI). Remaining observations were located south and west of the AOI. Observations were collected from March through November, with the majority collected from April through July (Figure 1).

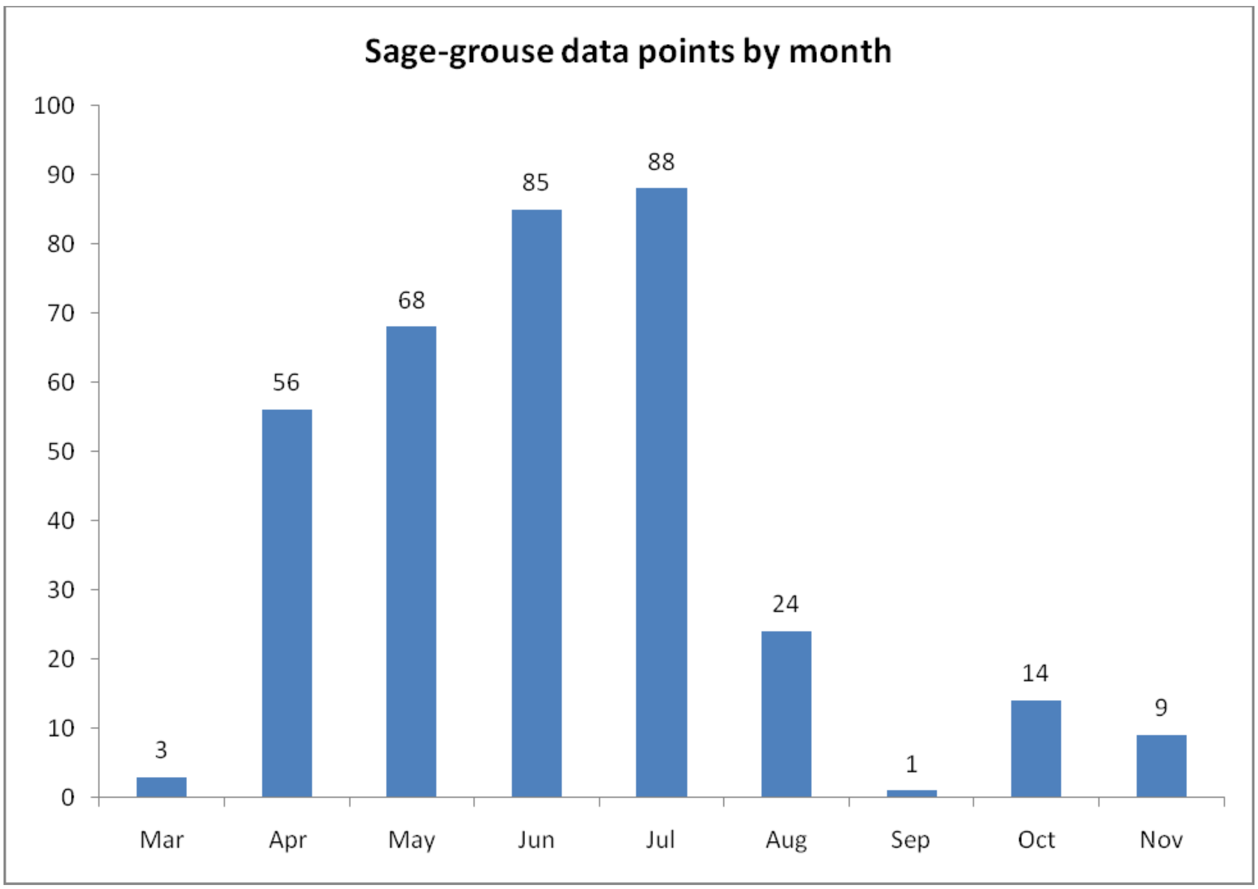

Figure 3: Number of sage-grouse observations by month. Data was collected by Leah Smith and provided by Todd Black. Total number of observations is 348.

The observation points are clustered in an arc that stretches about 15-km northward from approximately 39°42.2’N, 109°37.4’W, to 39°50.3’N, 109°36.9’W, then 15-km eastward to approximately 39°51.7’N, 109°27’W. Springtime observations are clustered in somewhat higher elevation, xeric sites in the northern and eastern portions of the arc while summertime observations are clustered at lower elevation, more mesic sites in the southern and western portions of the arc (Fig. 2).

The RS/GIS lab defined the study area boundary by a 10-km buffer around all of the sage grouse points in addition to the Anadarko AOI, which forms the northern and easternmost boundaries of the study area. In total, the study area encompasses 1769 km^2 in southern Uintah County, and measures approximately 60-km east-west and 53-km north-south (Fig. 2).

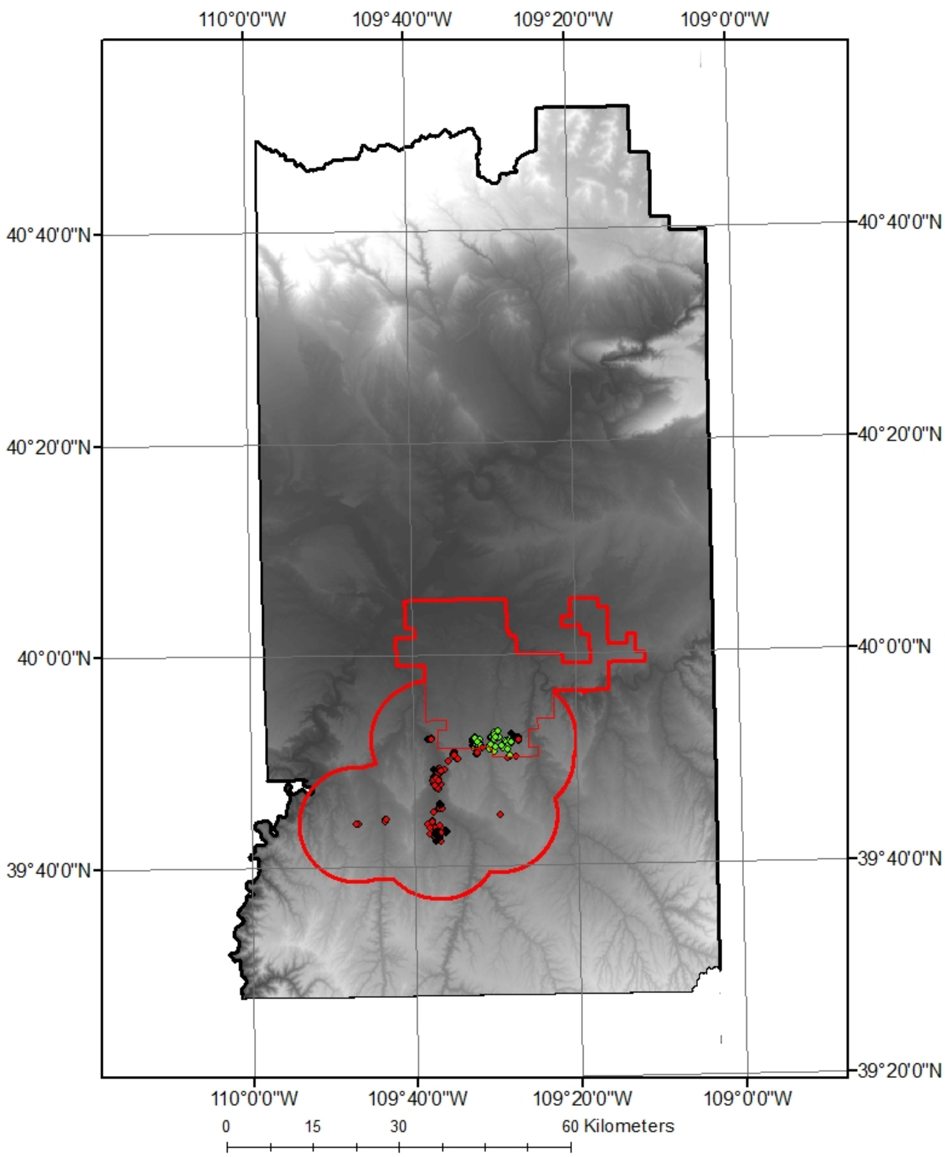

Figure 4: Map of the study area. The study area, located in Uintah County, Utah is indicated by the thick red line. Anadarko's EIS area of interest is indicated by the thin red line. March-May observations are in green. June though July observations are in red. Other observations are in black.

Modeling approach

We reviewed sage-grouse habitat modeling approaches from recent literature to identify approaches that may work with the grouse locations and habitat parameters available in the database. Yost et al. (2008) used Maximum Entropy (Phillips, Anderson, & Schapire, 2006) to model sage-grouse nesting habitat in southern Oregon. We chose not to use Maximum Entropy due to concerns about the accuracy and appropriateness of the approach for our application. Ecological Niche Factor Analysis (ENFA) used by Atamian, et al. (2010) to model sage-grouse brood rearing habitat in Nevada. However, an ENFA approach required modeling techniques beyond those we were able to glean in the amount of time requested.

Logistic regression is perhaps the most widely used habitat modeling tool. When working with presence-only datasets, logistic regression can be employed by generating pseudo-absences as random points (e.g, Zarnetske, 2006; Engler, Guisan, & Rechsteiner, 2004). In this project, we found logistic regression infeasible due to the clustered spatial configuration of observation points in a landscape highly impacted by oil and gas development. The dataset therefore provided a biased sample from which to model the relationship between anthropogenic or environmental variables and sage-grouse habitat use. Attempts to model seasonal habitat suitability via a logistic regression approach resulted in algorithms that would not converge and/or coefficients which were ecologically unrealistic (e.g. as oil and gas densities increased, logistic regression models had habitat suitability increasing), as well as unrealistic spatial depictions of habitat suitability (e.g. clustered around oil and gas wells). Additionally, 10-fold cross validation of logistic regression models indicated very poor predictive powers of the models.

Random Forests (RF) is an ensemble modeling approach related to classification and regression trees in which a large number of trees are grown, each of which uses a randomly selected 'bootstrap' sample of the data. At each node of each tree, a split is made using the best variable from a random subset m of all of the predictor variables M where $m << M$, e.g. the square root of M. Each tree in the forest is fully grown. Each sample receives a vote (in this context, for presence or absence) from each tree grown independently of that sample. For each sample, the forest chooses the classification which receives the most votes.

While RF models do not produce regression coefficients or other outputs amenable for traditional statistical inference, RF models have been shown to be highly accurate for classification in ecology (Cutler, et al., 2007), and have several advantages for the task at hand.

Classification by RF does not attempt to fit a linear relationship between variables, or transformations of variables, and the response. RF does not overfit models and provides a powerful means of identifying important variables. Additionally, the RandomForest package in R (Liaw & Wiener, 2002) allows the user to produce partial dependency plots to visualize the marginal impact of individual variables on the response. For these reasons, and because exploratory analysis indicated that RF models had better predictive powers and created more realistic spatial depictions of sage-grouse habitat suitability, we used an RF modeling approach for this project. The methods and results of the RF approach are described in this report.

The error rate of RF models is estimated based on the predicted classification of samples not in each bootstrap sample. These samples are called out-of-bag (OOB) data; on average, each data point would be out-of-bag approximately 36 percent of the time. The OOB error rate is calculated from the aggregation of the OOB predictions from each tree. RF estimates variable importance by assessing how much prediction error increases when the values of a given variable are permuted while the values for all other variables are left unchanged (Breiman, 2001).

Seasonal habitat use

For the purpose of this pilot study, we modeled habitat suitability in three time periods. This is based on seasonal movements of the grouse in the area. A spring habitat model was trained using observations collected during March, April, and May. A summer habitat was trained using observations collected during June and July. Finally, a year-round habitat model was trained using all available observations. We did not have observation points collected during the winter, so the year-round model could also be conceptualized as a 'non-winter' habitat model.

Of 348 total observations of sage-grouse presence, 127 were collected from March through May, and 173 were collected from June through July. No absence data is available, so the observations were paired with an equal number of unique pseudo-absences generated as random points across the study area using Esri ArcGIS 10 (Esri, Inc., 2010). The number of pseudo-absences was set equal to the number of presences because unbalanced ratios of presence to absence points can affect the accuracy of classification models (Manel, et al., 2001). Based on the daily movements of sage-grouse observed by Atamian, et al. (2010) and Aldridge and Boyce (2007), no pseudo-absences were generated within 510-m of a presence point.

Predictor variables

Fifty-seven geospatial predictor variables were produced at 30-m resolution using ERDAS Imagine (ERDAS, Inc., 2009) and ESRI ArcGIS 10 software (Esri, Inc., 2010). Thirty-six of these variables reflected neighborhood conditions around each pixel (e.g. mean, standard deviation). Thirty-four variables were calculated across a circular neighborhood of radius 510-m, and two variables were calculated across a circular neighborhood of radius 3-km. The 510-m neighborhood size was taken from Atamian, et al. (2010) and Aldridge and Boyce (2007), based on the observed daily movements of sage grouse in their study areas. The 3-km neighborhood size was taken from Holloran (2005), who studied sage-grouse response to natural gas field development in western Wyoming. Holloran found that as road densities within 3-km increased, lek attendance by males decreased, and that male lek attendance was sensitive to the number of producing natural gas wells within 3-km of the lek (Holloran, 2005).

The 57 predictor variables, their data sources, and the rationale for their development, are shown in Table 3. At each of the presence and pseudo-absence points the value of each predictor variable was extracted using the Sample tool in ArcGIS 10. All statistical analyses were performed in the statistical software *R* 2.12.2 (R Development Core Team, 2011), using the RandomForest (Liaw & Wiener, 2002), yaImpute (Crookston & Finley, 2007), and verification packages (NCAR - Research Application Program, 2010). The development of some of the predictor variables is described in more detail below.

Table 3: Predictor variables developed

Variable Name	**Description**	**Data source**	**Reasoning**
avg_cti	Mean Compound Topographic Index (CTI) (Gessler, Moore, McKenzie, & Ryan, 1995) over circular	Utah SGID	CTI is a moisture index; Ground moisture partly determines food availability

	neighborhood of radius 510-m.		
avg_slope	Mean slope (degrees) over circular neighborhood of radius 510-m.	Utah SGID	
avgbrt_apr	Average brightness over circular neighborhood of radius 510-m, April 2008	LANDSAT imagery from April 26 2008	Measure of overall reflectance; differentiates dry from wet soils
avgbrt_jul	Average brightness over circular neighborhood of radius 510-m, July 2008	LANDSAT imagery from July 15 2008	Measure of overall reflectance; differentiates dry from wet soils
avggrn_apr	Average greenness over circular neighborhood of radius 510-m, April 2008	LANDSAT imagery from April 26 2008	Measure of the presence and density of green vegetation
avggrn_jul	Average greenness over circular neighborhood of radius 510-m, July 2008	LANDSAT imagery from July 15 2008	Measure of the presence and density of green vegetation
avgherbcvr	Average herb cover (%) over circular neighborhood of radius 510-m	LANDFIRE 1.0.5	
avgherbht	Average herb height (m) over circular neighborhood of radius 510-m	LANDFIRE 1.0.5	
avgndmi_apr	Average NDMI over circular neighborhood of radius 510-m, April 2008	LANDSAT imagery from April 26 2008	Indication of moisture
avgndmi_jul	Average NDMI over circular neighborhood of radius 510-m, July 2008	LANDSAT imagery from July 15 2008	Indication of moisture
avgndvi_apr	Average NDVI over circular neighborhood of radius 510-m, April 2008	LANDSAT imagery from April 26 2008	Indication of live green vegetation
avgndvi_jul	Average NDVI over circular neighborhood of radius 510-m, July 2008	LANDSAT imagery from July 15 2008	Indication of live green vegetation

avgshrbcvr	Mean % Shrub Cover over circular neighborhood of radius 510-m	LANDFIRE 1.0.5	
avgshrbht	Mean Shrub Height over circular neighborhood of radius 510-m	LANDFIRE 1.0.5	
avgtreecvr	Mean % Tree Cover over circular neighborhood of radius 510-m	LANDFIRE 1.0.5	
avgwet_apr	Average wetness over circular neighborhood of radius 510-m, April 2008	LANDSAT imagery from April 26 2008	Measure of soil moisture content and vegetation density
avgwet_jul	Average wetness over circular neighborhood of radius 510-m, July 2008	LANDSAT imagery from July 15 2008	Measure of soil moisture content and vegetation density
brt_apr	Brightness, April 2008	LANDSAT imagery from April 26 2008	Measure of overall reflectance; differentiates dry from wet soils
brt_jul	Brightness, July 2008	LANDSAT imagery from July 15 2008	Measure of overall reflectance; differentiates dry from wet soils
cti	Compound Topographic Index (CTI) (Gessler, Moore, McKenzie, & Ryan, 1995)	Utah AGRC DEM derivative	CTI is a moisture index; Ground moisture partly determines food availability
dist_og	Distance to nearest Oil/Gas Well (m)	Utah SGID	Holloran (2005) found oil and gas development can affect sage-grouse habitat use
dist_peren	Distance to nearest perennial water source (m)	National Hydrography Dataset (NHD)	Proximity to water source may be important predictor of habitat use
dist_rd	Distance to road (m)	TIGER/Line files (US Census)	Holloran (2005) found roads can affect sage-grouse habitat use
evt11avg	Proportion of surrounding circular neighborhood of radius	LANDFIRE 1.0.5	EVT 11 = Open Water

	510-m composed of EVT 11		
evt2001avg	Proportion of surrounding circular neighborhood of radius 510-m composed EVT 2001	LANDFIRE 1.0.5	EVT 2001 = Inter-Mountain Basins Sparsely Vegetated Systems
evt2016avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2016	LANDFIRE 1.0.5	EVT 2016 = Colorado Plateau Pinyon-Juniper Woodland
evt2064avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2064	LANDFIRE 1.0.5	EVT 2064 = Colorado Plateau Mixed Low Sagebrush Shrubland
evt2066avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2066	LANDFIRE 1.0.5	EVT 2066 = Inter-Mountain Basins Mat Saltbush Shrubland
evt2080avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2080	LANDFIRE 1.0.5	EVT 2080 = Inter-Mountain Basins Big Sagebrush Shrubland
evt2081avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2081	LANDFIRE 1.0.5	EVT 2081 = Inter-Mountain Basins Mixed Salt Desert Scrub
evt2135avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2135	LANDFIRE 1.0.5	EVT 2135 = Inter-Mountain Basins Semi-Desert Grassland
evt2153avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2153	LANDFIRE 1.0.5	EVT 2153 = Inter-Mountain Basins Greasewood Flat

evt2159avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2159	LANDFIRE 1.0.5	EVT 2159 = Rocky Mountain Montane Riparian Systems
evt2180avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2180	LANDFIRE 1.0.5	EVT 2180 = Introduced Riparian Vegetation
evt2181avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2181	LANDFIRE 1.0.5	EVT 2181 = Introduced Upland Vegetation-Annual Grassland
evt2211avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 2211	LANDFIRE 1.0.5	EVT 2211 = Grayia spinosa Shrubland Alliance
evt31avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 31	LANDFIRE 1.0.5	EVT 31 = Barren
evt81avg	Proportion of surrounding circular neighborhood of radius 510-m composed of EVT 81	LANDFIRE 1.0.5	EVT 81 = Agriculture-Pasture and Hay
flow_acc	Flow accumulation (m^2) to each 30-m raster pixel	Utah SGID DEM derivative	May predict moisture & food availability
grn_apr	Greenness, April 2008	LANDSAT imagery from April 26 2008	Measure of the presence and density of green vegetation
grn_jul	Greenness, July 2008	LANDSAT imagery from July 15 2008	Measure of the presence and density of green vegetation
herb_ht	herb height (m)	LANDFIRE 1.0.5	

herbcvr	% herb cover	LANDFIRE 1.0.5	
ndmi_apr	Normalized Difference Moisture Index (NDMI), April 2008	LANDSAT imagery from April 26 2008	Indication of moisture
ndmi_jul	Normalized Difference Moisture Index (NDMI), July 2008	LANDSAT imagery from July 15 2008	Indication of moisture
ndvi_apr	Normalized Difference Vegetation Index (NDVI), April 2008	LANDSAT imagery from April 26 2008	Indication of live green vegetation
ndvi_jul	Normalized Difference Vegetation Index (NDVI), July 2008	LANDSAT imagery from July 15 2008	Indication of live green vegetation
nonforest	Nonforest EVTs	LANDFIRE 1.0.5	SG don't live in forested areas. Nonforest excludes EVTs with *SYSTMGRPPH* (system group) values *Conifer, Hardwood, Conifer-Hardwood*
og_dens3km	Number of active oil/gas wells within 3-km	Utah SGID	Holloran (2005) found oil and gas development within 3km can affect sage-grouse habitat use
rd_den3km	km of roads within 3-km	TIGER/Line files (US Census)	Methods described in text
shrb_ht	shrub height (m)	LANDFIRE 1.0.5	
shrbcvr	% shrub cover	LANDFIRE 1.0.5	
stdherbcvr	Standard deviation of percent herb cover over circular neighborhood of radius 510-m	LANDFIRE 1.0.5	Sage-grouse prefer variegated landscape

stdshrbcvr	Standard Deviation of % Shrub Cover over circular neighborhood of radius 510-m	LANDFIRE 1.0.5	Sage-grouse prefer variegated landscape
treecvr	% tree cover	LANDFIRE 1.0.5	
wet_apr	Wetness, April 2008	LANDSAT imagery from April 26 2008	Measure of soil moisture content and vegetation density
wet_jul	Wetness, July 2008	LANDSAT imagery from July 15 2008	Measure of soil moisture content and vegetation density

Spatial point data representing the surface location of oil and gas wells was acquired from the Utah State Geographic Information Database (SGID) (Utah SGID, 2011). Using this data, two metrics representing the presence of active oil and gas wells were developed: distance to the nearest active well (*dist_og*), and well density over a circular neighborhood with radius 3-km (*og_dens3km*). This analysis neighborhood was chosen based on Holloran (2005), who found that natural gas development up to 3-km away could impact sage grouse habitat use.

Because sage-grouse points were collected at various times in 2007 and 2008, attribute fields in the AGRC oil and gas well database were used to identify only those wells that were likely to have been active at the beginning of the year of the associated sage-grouse observation. For example, if a sage-grouse observation was collected in April of 2007, the proximity of that observation to oil and gas wells was calculated using only those wells that we were confident were active at the beginning of 2007 (i.e. 1 January, 2007). Sage-grouse observations were collected in 2007 and 2008, so we created two subsets of the AGRC oil and gas well database, representing wells active at on 1 January 2007, and 1 January 2008, and computed proximity and density metrics from these datasets.

Wells were identified as active on 1 January of 2007 or 2008 by excluding all wells with a value in the *ABNDONDATE* (abandon date) field reflecting well abandonment before 1 January 2007 or 1 January 2008, respectively. Wells with no *ABNDONDATE* but *WELLSTATUS* listed as *LA* ("Location abandoned; permit rescinded") or *PA* ("Plugged and abandoned") were also excluded. Measures of proximity to and density of oil and gas wells for all pseudo-absence points were taken from wells estimated as active as of 1 January 1 2007, because the majority (230 out of 348) of sage-grouse observations were collected in 2007. To apply the RF models to the creation habitat suitability maps reflecting current conditions, measures of proximity to and density of oil and gas wells were based on wells active as of 4 February 2011. These wells were defined exclusive of all LA wells, PA wells, and all wells with a valid *ABNDONDATE*.

The relationship of roads to sage-grouse observations was measured in two ways: (1) distance to the nearest road (*dist_rd*), and (2) road density (*rd_den3km*), calculated as the kilometers of road in a 3-km moving window. Roads data was taken from the U.S. Census Bureau's 2010 Topographically Integrated Geographic Encoding and Referencing system (TIGER) roads data (U.S. Census Bureau, Geography Division, 2010). We included primary, secondary, and local neighborhood roads, as well as roads for service vehicles, and excluded vehicular trails for 4WD drive vehicles only.

Predictor variables measuring vegetative landcover were derived from the LANDFIRE 1.0.5 Existing Vegetation Type spatial data layer (USGS, 2011). For each of the 15 most prevalent existing vegetation types (EVT) in the study area, the proportion of all landcover in a 510-m moving window composed of that vegetation type was calculated. The 15 vegetation types together accounted for 99.6% of all landcover in the study area (Table 4).

Table 4. The 15 most prevalent LANDFIRE existing vegetation types in the study area. Together these vegetation types account for 99.6% of all landcover. For each vegetation type, a continuous predictor variable was developed indicating the proportion of all landcover within 510-m composed of that vegetation type.

Value	Hectares	% study area	EVT_name
2080	60445	31.83	Inter-Mountain Basins Big Sagebrush Shrubland
2081	55104	29.02	Inter-Mountain Basins Mixed Salt Desert Scrub
2064	26847	14.14	Colorado Plateau Mixed Low Sagebrush Shrubland
31	11311	5.96	Barren
2016	10634	5.60	Colorado Plateau Pinyon-Juniper Woodland
2181	10309	5.43	Introduced Upland Vegetation-Annual Grassland
2153	3652	1.92	Inter-Mountain Basins Greasewood Flat
2001	3202	1.69	Inter-Mountain Basins Sparsely Vegetated Systems
2066	2971	1.56	Inter-Mountain Basins Mat Saltbush Shrubland
81	1659	0.87	Agriculture-Pasture and Hay
2180	1225	0.65	Introduced Riparian Vegetation
2211	800	0.42	Grayia spinosa Shrubland Alliance
2159	464	0.24	Rocky Mountain Montane Riparian Systems
2135	325	0.17	Inter-Mountain Basins Semi-Desert Grassland
11	175	0.09	Open Water
	189123	99.6	[Total area represented by these 15 vegetation types]
	189876	100.0	[Total study area]

The LANDFIRE 1.0.5 Existing Vegetation Height and Existing Vegetation Cover spatial layers (USGS, 2011) were used to produce predictor layers describing the height and percent cover of shrub and herb vegetation, and the percent cover of trees, for each 30-m pixel. These LANDFIRE layers describe vegetation height or cover in ordinal categories, e.g. 10%-20% tree cover. The middle value of each category, e.g. 15% tree cover, was used to transform these ordinal categories into continuous values. Variables reflecting the average and standard deviation of the height and percent cover of shrub and herb vegetation and the average percent tree cover were produced by calculating these values over a 510-m moving window.

Two cloud-free Landsat 5 satellite images of the study area captured during the same time period as Leah Smith's sage-grouse data collection were identified and acquired from the U.S. Geological Survey's Global Visualization Viewer database (USGS, 2011). The images were captured on April 26, 2008 and July 15, 2008. Using ERDAS Imagine 2010 software (ERDAS, Inc., 2009), ecogeographical gradients describing brightness, greenness, and wetness were produced from these images using a tasseled-cap transformation (Cris, Laurin, & Cicone, 1986). The Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Moisture Index (NDMI) were produced by converting the images to reflectance values and applying an atmospheric cost correction using the Chavez COST method (Chavez, 1996). For each of the satellite-imagery derived metrics (Brightness, Greenness, Wetness, NDVI, NDMI), the average value of the metric over a moving window with radius 510-m was also produced.

Variable reduction process

For each of the three habitat models, the number of predictor variables was reduced in three steps. First, a RF model was produced using all 57 predictor variables. The variables were ranked according to importance as indicated by the MeanDecreaseAccuracy measure provided in the randomForest package in R (Liaw & Wiener, 2002), and all but the 20 most important variables were eliminated. Next, a second RF model was produced using the top 20 variables and, again using the MeanDecreaseAccuracy measure, the lesser important of any highly correlated variables ($|r|>0.7$) were eliminated. Correlation was calculated over 10,000 random points across the study area. Third, another RF model was produced using the remaining variables, and all but the top eight variables were eliminated. These eight variables were used in a final RF model to produce the habitat suitability map.

We chose to use eight variables as a cut off because it represents a relatively small set of variables yet also retains a number of variables reflecting both direct and obvious anthropogenic influences (e.g. oil and gas wells, roads) as well as vegetation and satellite-imagery derived measurements of the natural environment. By choosing to use these eight variables we found we could produce RF models which accurately classify presences and pseudo-absences using only a few variables. However, these models did not seem ecologically realistic. For example, they would identify areas of a given proximity to and density of oil and gas wells as highly suitable. This is because the set of clustered observation points also fell within a similar distance to or density of oil and gas wells. The ability to correct for this bias is limited because the observation points are spatially clustered, and the sampling design used to collect the observation points is unknown. By retaining eight variables, we have attempted to develop a simple model in which the marginal effect of each predictor on the response can be accessed via partial dependence plots, and which also retains a sufficient representation from a variety of variables reflecting ecologically meaningful phenomena. While the use of an 8-variable cutoff is somewhat arbitrary, the resulting classification models are highly accurate in classifying the presences and pseudo-absences, and the resulting maps appear to more realistically reflect ecologically meaningful approximations of habitat suitability than do maps created from RF models built from fewer variables.

Accuracy assessment methods

For each RF model generated during the variable reduction process, five accuracy assessment metrics were generated. These included: sensitivity (the rate of success in identifying presences as presences); specificity (the rate of success in identifying absences as absences); percent of all points correctly classified; Cohen's kappa statistic; and the area under the receiver operating characteristic curve (AUC). Cohen's kappa statistic (Cohen, 1960) is a measure of agreement between the model and the data which takes into account the amount of agreement that would occur by chance. AUC measures the probability that the model will rank the habitat suitability of a randomly chosen presence higher than a randomly chosen pseudo-absence (Ling, Huang, & Zhang, 2003).

Results

Predictor variables retained

The eight predictor variables retained in each of the models are shown in order of importance in Table 5. Each of the habitat models retained predictor variables reflecting both anthropogenic effects as well as features of the natural environment. Each model retained two variables measuring direct anthropogenic influence on the landscape (i.e. *dist_og*, *og_dens3km*, and/or *rd_den3km*); one or two variables derived from July 2008 Landsat imagery (i.e. *avgndvi_jul* and/or *avgbrt_jul*); three or four variables reflecting vegetation conditions mapped by the LANDFIRE project (i.e. *evt2064avg*, *evt81avg*, *evt2016avg*, *avgtreecvr*, and/or *evt2081avg*); and the variable which measured the proximity to perennial streams or waterbodies (*dist_peren*). Interestingly, none of the variables derived from April 2008 Landsat imagery were retained in any of the habitat models.

Four common variables were retained in each of the models: *dist_peren* (distance to nearest perennial stream/waterbody), *og_dens3km* (number of oil/gas wells within circular neighborhood with radius = 3km), *evt2064avg* (proportion of landcover that is *Colorado Plateau Mixed Low Sagebrush Shrubland* within circular neighborhood with radius = 510m), and *evt81avg* (proportion of landcover that is *Agriculture-Pasture and Hay* within circular neighborhood with radius = 510m). Additionally, a variable measuring the amount of tree cover (*avgtreecvr*) or the amount of Colorado Plateau Pinyon-Juniper Woodland (*evt2016avg*) over a 510-m radius circular neighborhood was retained in each of the habitat models. These variables are highly correlated (r = 0.995); therefore, the amount of tree cover, specifically pinyon-juniper tree cover, can be considered a fifth variable retained in each of the habitat suitability models.

Table 5. Predictor variables retained in the 3 habitat models. Variables are listed in descending order of importance, as measured by the MeanDecreaseAccuracy measure available in the randomForest package in R. Four variables (*dist_peren, evt81avg, evt2064avg, og_dens3km*) were included in all 3 habitat models. Variables not included in all 3 habitat models are indicated by an asterisk (*).

Variable importance rank	Springtime model	Summertime model	Year-round model
1	*dist_og**	*dist_peren*	*dist_peren*
2	*evt2064avg*	*og_dens3km*	*og_dens3km*
3	*og_dens3km*	*evt2064avg*	*evt2064avg*
4	*evt2081avg**	*evt81avg*	*avgtreecvr**
5	*evt81avg*	*evt2081avg**	*evt81avg*
6	*dist_peren*	*avgtreecvr**	*avgbrt_jul**
7	*avgbrt_jul**	*dist_og**	*avgndvi_jul**
8	*evt2016avg**	*avgndvi_jul**	*rd_den3km**

Graphical and tabular descriptions of the results of the 3 seasonal habitat models are provided in Appendix 1. For each habitat model, one table shows the variable reduction process and model accuracy assessment metrics, and four figures show variable importance, partial dependency plots for the eight predictor variables, and habitat suitability maps for the study area. For the spring model, these tables and figures are shown on pages 27-32; for the summer model they are shown on pages 33-38; and for the year-round model they are shown on pages 39-44.

Accuracy assessment

Accuracy assessment metrics for each of the RF models indicates that the models fit the data very well (Tables A1.1 (spring model), A1.2 (summer model), A1.3 (year-round model)). Cross-validation is not necessary with RF models because the OOB estimate of the error rate is calculated using the randomly selected bootstrap samples of the data. The OOB estimate of the error rate is therefore already a cross-validated estimate of the error rate. The final models for the spring, summer, and year-round models correctly classified 93.7%, 94.22%, and 96.26% of all points, respectively. Sensitivity measures were higher than the total percent correctly classified, which is encouraging because these models were built using pseudo-absences, not true absences. Sensitivity measures for the spring, summer, and year-round models are 94.49%, 96.53%, and 97.99%, respectively, indicating that the models are excellent classifiers of sage-grouse presence points. The AUC criterion for the spring, summer, and year-round models are 0.9913, 0.9854, and 0.9948, respectively, indicating that the models are very good at ranking environmental conditions associated with presences as more suitable habitat than environmental conditions associated with pseudo-absences.

The marginal effect of each of the eight variables used in the habitat suitability models are shown in partial dependence plots in Figures A1.2 and A1.3 (spring model); Figures A1.6 and A1.7 (summer model); and Figures A1.10 and A1.11 (year-round model). These plots also help illuminate the effect of the clustered observation points on attempts to fit a model to the data. Partial dependence plots depict the value of the predictor variable on the x-axis and the logit of the predicted probability of presence (that is, the log of the fraction of votes) on the y-axis.

Visual inspection of the spring and summer habitat suitability maps show some suspect spatial patterns arising from the location of oil and gas wells in the study area. Inspection of the partial dependence plots for the variable *dist_og* for these models shows a spike in the relationship between this predictor and the response (Figures A1.2, A1.7), which is likely the cause of the suspect spatial patterns. These effects are less pronounced in the year-round habitat model, which not only did not retain *dist_og* as a predictor variable, but also used all available data points in training the model. For these reasons, the year-round model should be considered the most realistic of the three habitat suitability models.

Map products

Thirty-meter resolution maps of the study area which rank habitat suitability over a range from 0 to 1 were created using the final RF models and the R package yaImpute (Crookston & Finley, 2007). The rasters were reclassified into four categories to reflect differing levels of habitat suitability: 1 (likely not suitable habitat; values 0 – 0.25), 2 (somewhat suitable habitat; values 0.25 – 0.5), 3 (suitable habitat; values 0.5 – 0.75), and 4 (highly suitable habitat; values 0.75 – 1). These four habitat suitability classes were used to convert the rasters to polygons, which reflect this pilot project's estimates of seasonal habitat suitability for sage-grouse in the study area (Figures A1.4 (spring model), A1.8 (summer model), A1.12 (year-round model)).

Conclusions Management Applications

In any modeling effort, the results of the models are only as good as the assumptions and data used in their development. Accordingly, several limitations exist with respect to interpreting the results of these spatial models. These include the spatial clustering of data points and whether or not the data represent a true random sample, as well as the assumptions made with respect to the development and inclusion of predictor variables (e.g. what constitutes an active oil and gas well). However, given the available data and the objective of the modeling effort (the production of seasonal sage-grouse habitat use maps for Anadarko's AOI), these models are high-quality, data-driven, objective approximations of habitat suitability.

References

Aldridge, C., & Boyce, M. S. (2007). Linking occurrence and fitness to persistence: habitat-based approach for endangered greater sage-grouse. *Ecological Applications, 17*(2), 508-526.

Aldridge, C., Nielson, S., Beyer, H., Boyce, M., Connelly, J., Knick, S., et al. (2008). Range-Wide Patterns of Greater Sage-Grouse Persistence. *Diversity and Distributions, 14*, 983-994.

Atamian, M., Sedinger, J., Heaton, J., & Blomberg, E. (2010). Landscape-Level assessment of brood rearing habitat for Greater Sage-Grouse in Nevada. *Journal of Wildlife Management, 74*(7), 1533-1543.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32.

Chavez, P. (1996). Image-based atmospheric corrections - revisited and revised. *Photogrammetric Engineering and Remote Sensing, 69*(2), 1025-1036.

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

Cris, E., Laurin, R., & Cicone, R. (1986). Vegetation and soils information contained in transformed Thematic Mapper data. *Proceedings of IGARSS '86 Symposium*, 1465-1470.

Crookston, N., & Finley, A. (2007). yaImpute: An R Package for kNN Inputation. *Journal of Statistical Software, 23*(10).

Cutler, D., Edwards, T., Beard, K., Cutler, A., Hess, K., Gibson, J., et al. (2007). Random forests for classification in ecology. *Ecology, 88*(11), 2783-2792.

Engler, R., Guisan, A., & Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology, 41*, 263-274.

ERDAS, Inc. (2009). *ERDAS IMAGINE 2010.* Norcross, GA: ERDAS, Inc.

Esri, Inc. (2010). *ArcGIS 10.0.* Redlands, CA: ESRI, Inc.

Gessler, P., Moore, I., McKenzie, N., & Ryan, P. (1995). Soil-landscape modeling and spatial prediction of soil attributes. *International Journal of GIS, 9*, 421-432.

Holloran, M. (2005). *Sage-grouse population response to natural gas field development in western Wyoming.* Laramie, WY: University of Wyoming.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News, 2*(3), 18-22.

Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: A better measure than accuracy in comparing learning algorithms. *Advances in Artificial Intelligence Proceedings, 2671*, 329-341.

Manel, S., Williams, H., & Ormerod, S. (2001). Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology, 38*, 921-931.

NCAR - Research Application Program. (2010). verification: Forecast verification utilities. *R package version 1.31*.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling, 190*, 231-259.

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Smith, L. S. (2009). *Greater sage-grouse and energy development in northeastern Utah: Implications for management.* Logan, UT: Utah State University.

U.S. Census Bureau, Geography Division. (2010, November 30). *U.S. Census Bureau.* Retrieved February 4, 2011, from TIGER/Line: http://www.census.gov/geo/www/tiger/

USGS. (2011). *LANDFIRE 1.0.5, Existing Vegetation Type, Existing Vegetation Height, and Existing Vegetation Cover layers.* Retrieved February 4, 2011, from LANDFIRE: http://landfire.cr.usgs.gov/viewer/

USGS. (2011). *USGS Global Visualization Viewer.* Retrieved Feb 4, 2011, from Earth Resources Observation and Science Center (EROS): http://glovis.usgs.gov/

Utah SGID. (2011, February 4). *DNROilGasWells.* Retrieved February 4, 2011, from Utah State Geographic Information Database : http://gis.utah.gov/sgid-vector-download/utah-sgid-vector-gis-data-layer-download-index?fc=DNROilGasWells

Yost, A., Petersen, S., Gregg, M., & Miler, R. (2008). Predictive modeling and mapping sage grouse (Centrocercus urophasianus) nesting habitat using Maximum Entropy and a long-term dataset from Southern Oregon. *Ecological Informatics, 3*, 375-386.

Zarnetske, P. (2006). *Landscape-Scale habitat classification modeling with incomplete data: The generation of ecologically-based pseudo-absence points.* Logan, UT: Utah State University.

APPENDIX 1: GRAPHICAL AND TABULAR RESULTS OF SEASONAL HABITAT MODELS

Spring model results

Table A1.1. Spring habitat model variable selection and accuracy assessments. Number of observations: 127 presences, 127 pseudo-absences = 254 points total.

Variable reduction step 1: Identify top 20 vars	Variable reduction step 2: drop correlated vars		Variable reduction step 3: Drop all but top 8 vars		Final model
Variable name	Mean Decrease Accuracy	Drop variable?	Mean Decrease Accuracy	Drop variable?	Mean Decrease Accuracy
dist_og	1.79		1.86		1.98
og_dens3km	1.58		1.64		1.8
evt2064avg	1.4		1.55		1.83
evt2081avg	1.36		1.41		1.71
avgbrt_jul	1.34		1.39		1.55
dist_peren	1.22		1.27		1.68
evt81avg	1.22		1.3		1.71
avg_cti	1.22		1.2	#	
evt2016avg	1.17		1.36		1.51
avgndmi_jul	1.15		1.15	#	
grn_jul	1.14		1.22	#	
avgndmi_apr	1.12	# (cor w/ evt2081avg)			
avgwet_jul	1.12	# (cor w/ avgbrt_jul & avgndmi_jul)			
avgtreecvr	1.09	# (cor w/ evt2016avg)			
avgshrbcvr	1.07		1.17	#	
evt2211avg	1.07		1.06	#	
rd_den3km	1.04		1.22	#	
evt2181avg	1.03		1.05	#	
avgndvi_jul	1.01		1.15	#	
evt2066avg	0.92		0.9	#	
	20 vars		**17 vars**		**8 vars**
Sensitivity	**94.49**		**94.49**		**94.49**
Specificity	**91.34**		**92.13**		**92.91**
% Correctly Classified	**92.91**		**93.31**		**93.70**
Kappa	**0.8583**		**0.8661**		**0.8740**
AUC	**0.9908**		**0.9912**		**0.9913**

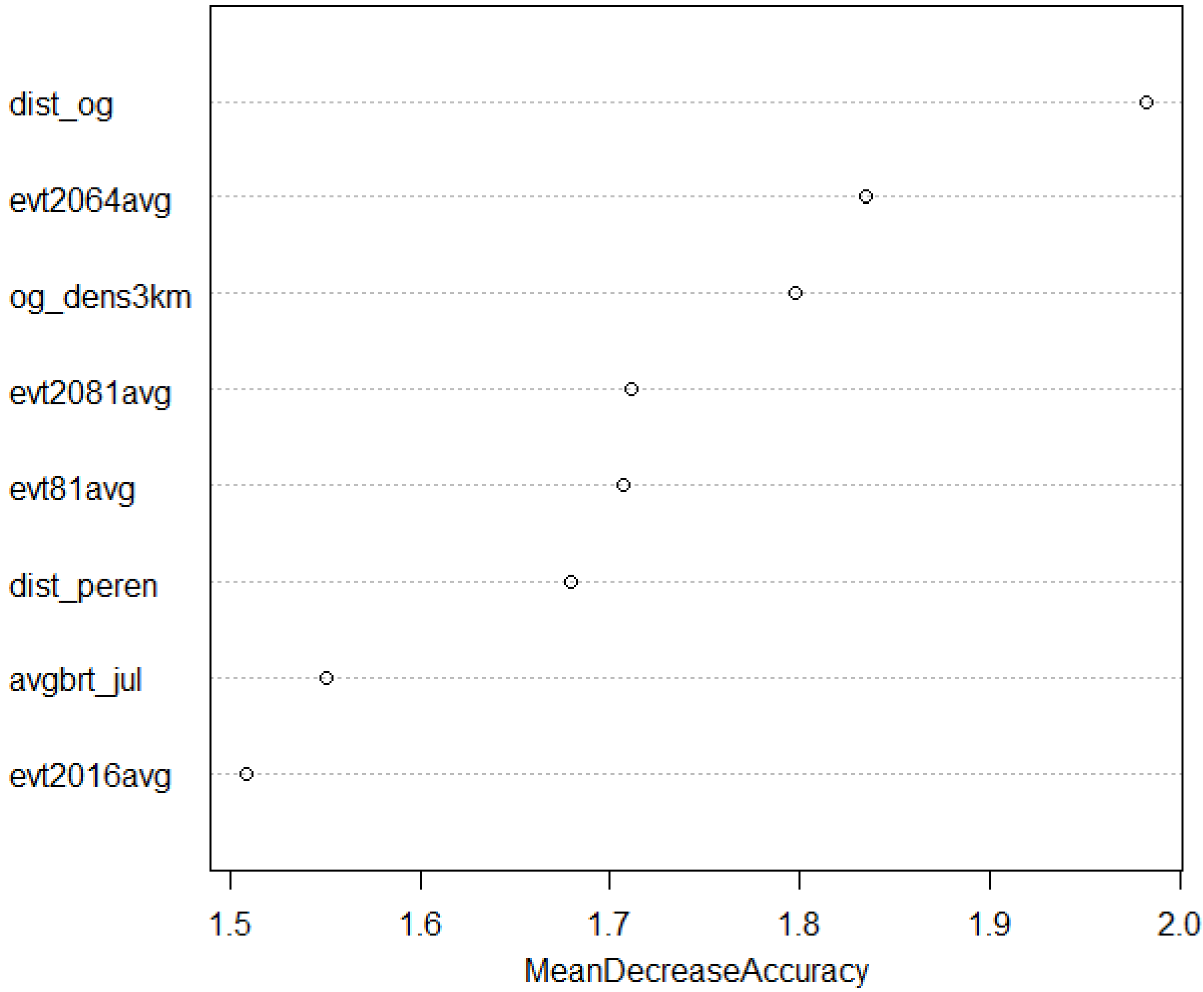

Figure A1.1: Spring model variable importance measures. Mean decrease in accuracy is calculated as the average percentage decrease in OOB classification accuracy for the random forest when the variable of interest is permuted in the OOB sample and accuracy is recalculated.

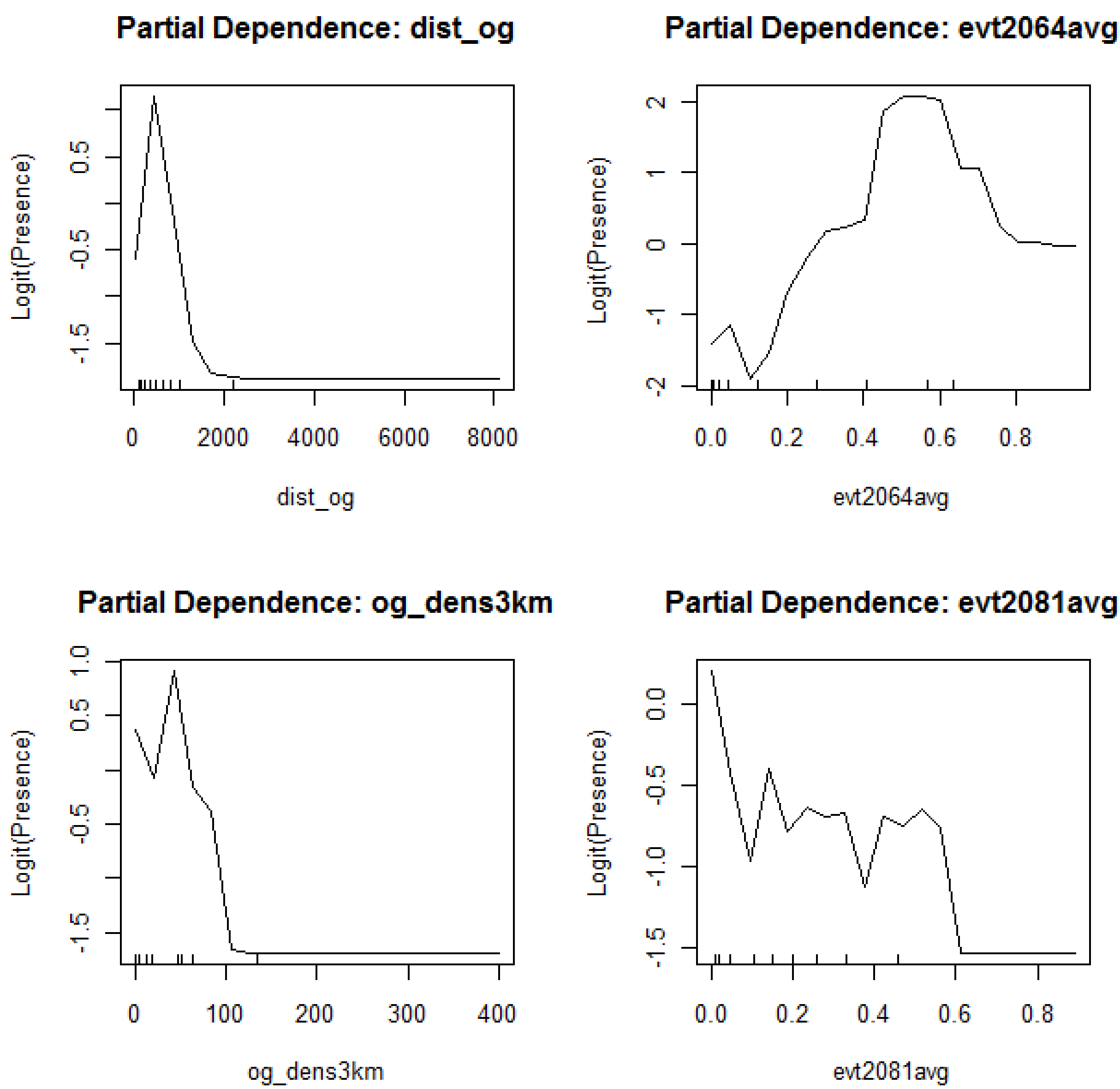

Figure A1.2: Spring model partial dependency plots depicting the marginal effect of *dist_og*, *evt2064avg*, *og_dens3km*,and *evt2081avg* on the response (sage-grouse presence or pseudo-absence). The partial plots indicate the marginal effect these variables have on the class probability of the response (y-axis units are the log of the fraction of ‘presence’ votes of all trees in the RF model). Clockwise from top left, X-axes measure: meters from nearest active oil/gas well (*dist_og*); proportion of vegetation within 510m classified as Colorado Plateau Mixed Low Sagebrush Shrubland (*evt2064avg*); number of active oil and gas wells within 3km (*og_dens3km*); proportion of vegetation within 510m classified as Inter-Mountain Basins Mixed Salt Desert Scrub (*evt2081avg*).

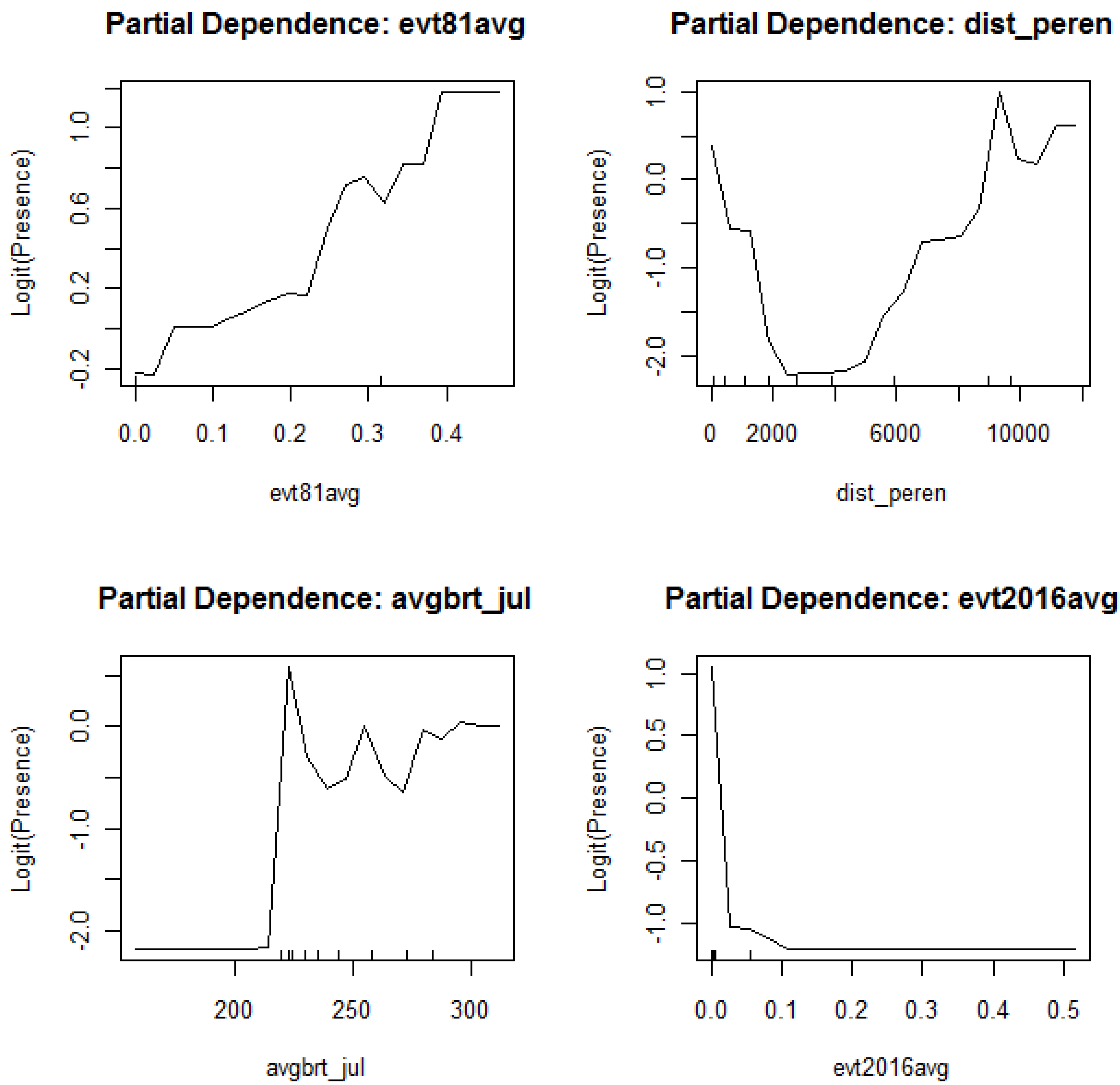

Figure A1.3: Spring model partial dependency plots depicting the marginal effect of *evt81_avg*, *dist_peren*, *avgbrt_jul*, and *evt2016avg* on the response (sage-grouse presence or pseudo-absence). The partial plots indicate the marginal effect these variables have on the class probability of the response (y-axis units are the log of the fraction of 'presence' votes of all trees in the RF model). Clockwise from top left, X-axes measure: proportion of vegetation within 510m classified as Agriculture-Pasture and Hay (*evt81avg*); distance (meters) from the nearest perennial stream/waterbody (*dist_peren*); average July 15, 2008 Brightness value over a 510m radius neighborhood (*avgbrt_jul*); proportion of vegetation within 510m classified as Colorado Plateau Pinyon-Juniper Woodland (*evt2016avg*).

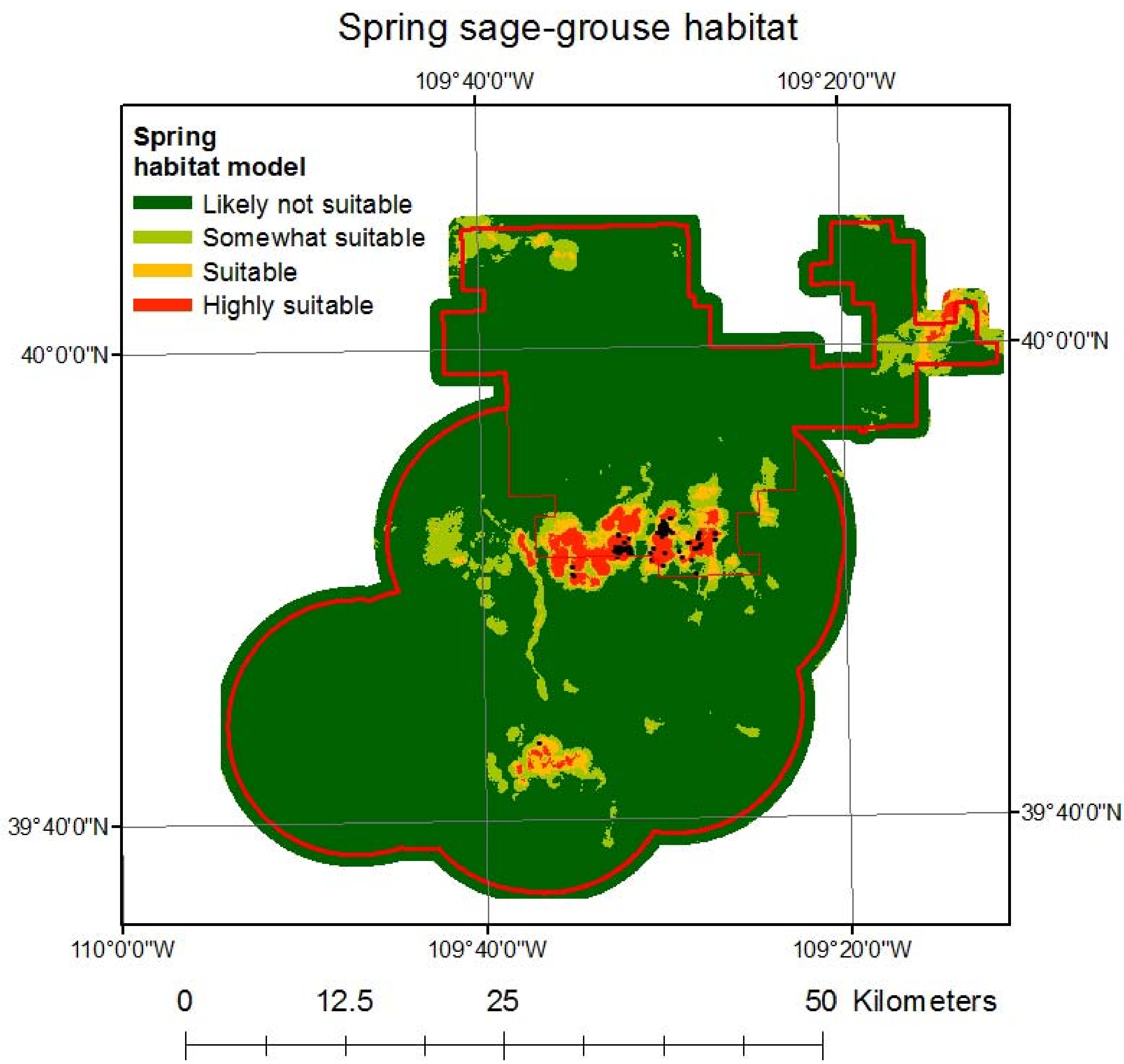

Figure A1.4: Spring sage-grouse habitat in the study area (thick red line). The southern boundary of Anadarko's AOI is indicated by the thin red line. Presence points used to train the model are shown as small black dots. One can see some suspect circular patterning, which is dependent on currently active oil and gas (not shown).

Summer model results

Table A1.2. Summer habitat model variable selection and accuracy assessments. Number of observations: 173 presences, 173 pseudo-absences = 346 points total.

Variable reduction step 1: Identify top 20 vars	Variable reduction step 2: drop correlated vars		Variable reduction step 3: Drop all but top 8 vars		Final model
Variable name	Mean Decrease Accuracy	Drop variable?	Mean Decrease Accuracy	Drop variable?	Mean Decrease Accuracy
og_dens3km	1.41		1.48		1.67
dist_peren	1.39		1.44		1.7
evt2064avg	1.28		1.39		1.62
evt81avg	1.26		1.34		1.61
evt2081avg	1.25		1.37		1.52
avgtreecvr	1.18		1.45		1.5
rd_den3km	1.17		1.19	#	
dist_og	1.16		1.23		1.44
avg_slope	1.16		1.2	#	
evt2016avg	1.13	# (cor w/ avgtreecvr)			
avgbrt_jul	1.13		1.17	#	1.44
avg_cti	1.1	# (cor w/ avg_slope)			
avggrn_jul	1.08	# (cor w/ avgbrt_jul)			
dist_rd	1.07		1.08	#	
avgndvi_jul	1.07		1.28		
avgndvi_apr	1.02	# (cor w/ avgndvi_jul)			
avgshrbcvr	1.01		1.03	#	
evt2066avg	1.01		0.96	#	
evt2001avg	0.98		1.04	#	
evt2211avg	0.95		0.99	#	
	20 vars		**16 vars**		**8 vars**
Sensitivity	**96.53**		**96.53**		**96.53**
Specificity	**92.49**		**93.06**		**91.91**
% Correctly Classified	**94.51**		**94.80**		**94.22**
Kappa	**0.8902**		**0.8960**		**0.8844**
AUC	**0.9822**		**0.9807**		**0.9854**

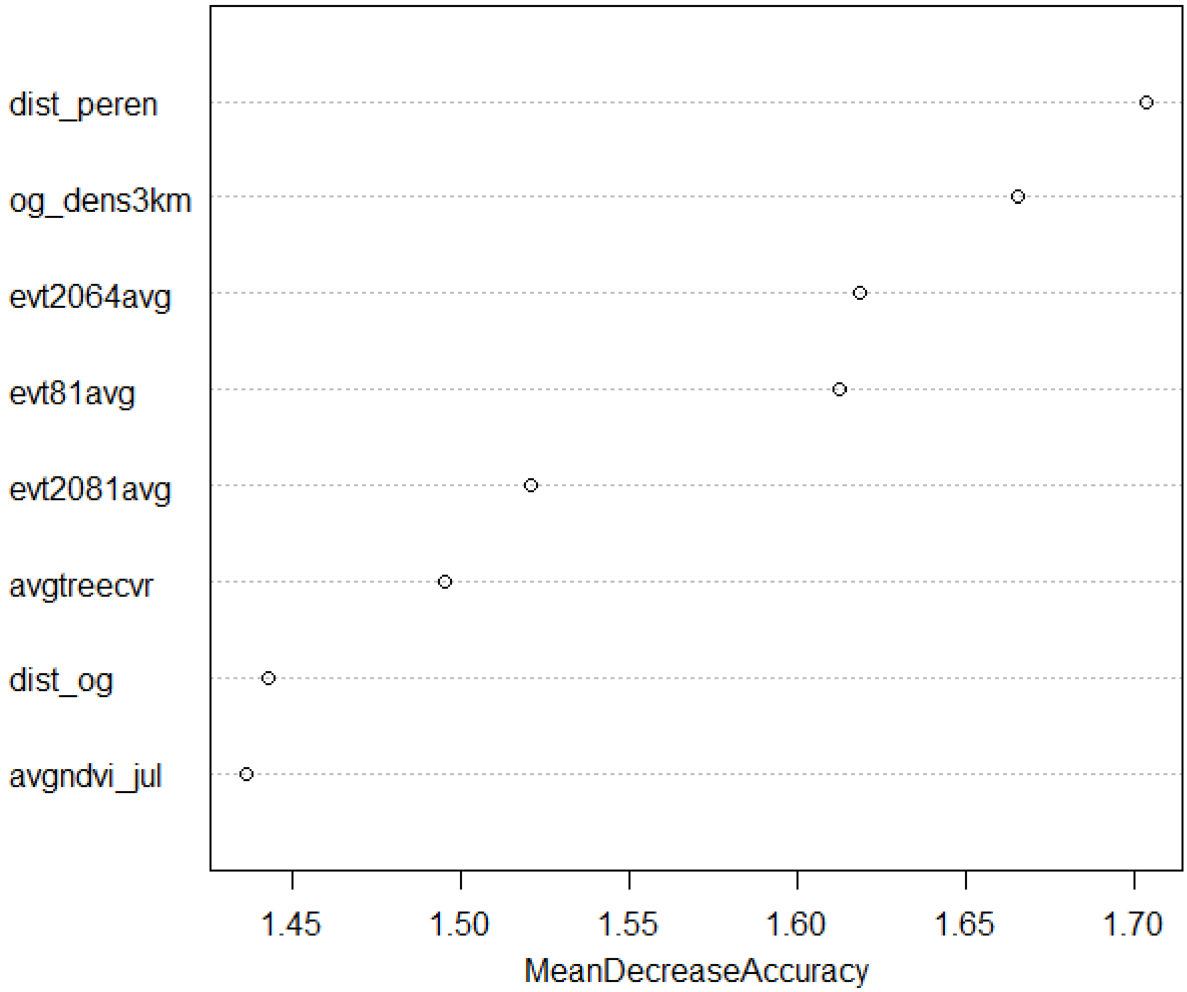

Figure A1.5: Summer model variable importance measures. Mean decrease in accuracy is calculated as the average percentage decrease in OOB classification accuracy for the random forest when the variable of interest is permuted in the OOB sample and accuracy is recalculated.

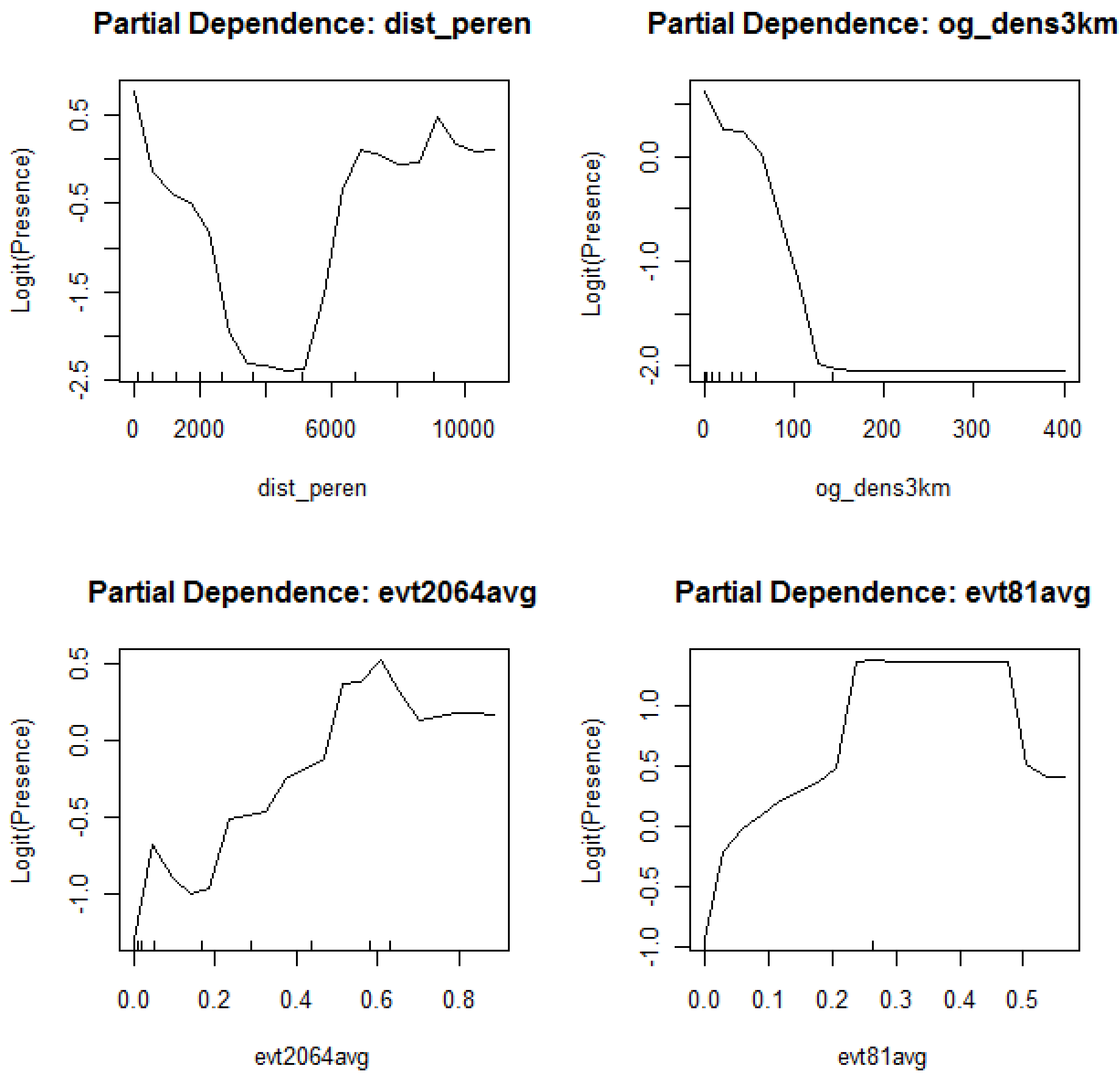

Figure A1.6: Summer model partial dependency plots depicting the marginal effect of *dist_peren, og_dens3km, evt2064avg,* and *evt81avg* on the response (sage-grouse presence or pseudo-absence). The partial plots indicate the marginal effect these variables have on the class probability of the response (y-axis units are the log of the fraction of 'presence' votes of all trees in the RF model). Clockwise from top left, X-axes measure: distance (meters) from the nearest perennial stream/waterbody (*dist_peren*); number of active oil and gas wells within 3km (*og_dens3km*); proportion of vegetation within 510m classified as Colorado Plateau Mixed Low Sagebrush Shrubland (*evt2064avg*); proportion of vegetation within 510m classified as Agriculture-Pasture and Hay (*evt81avg*).

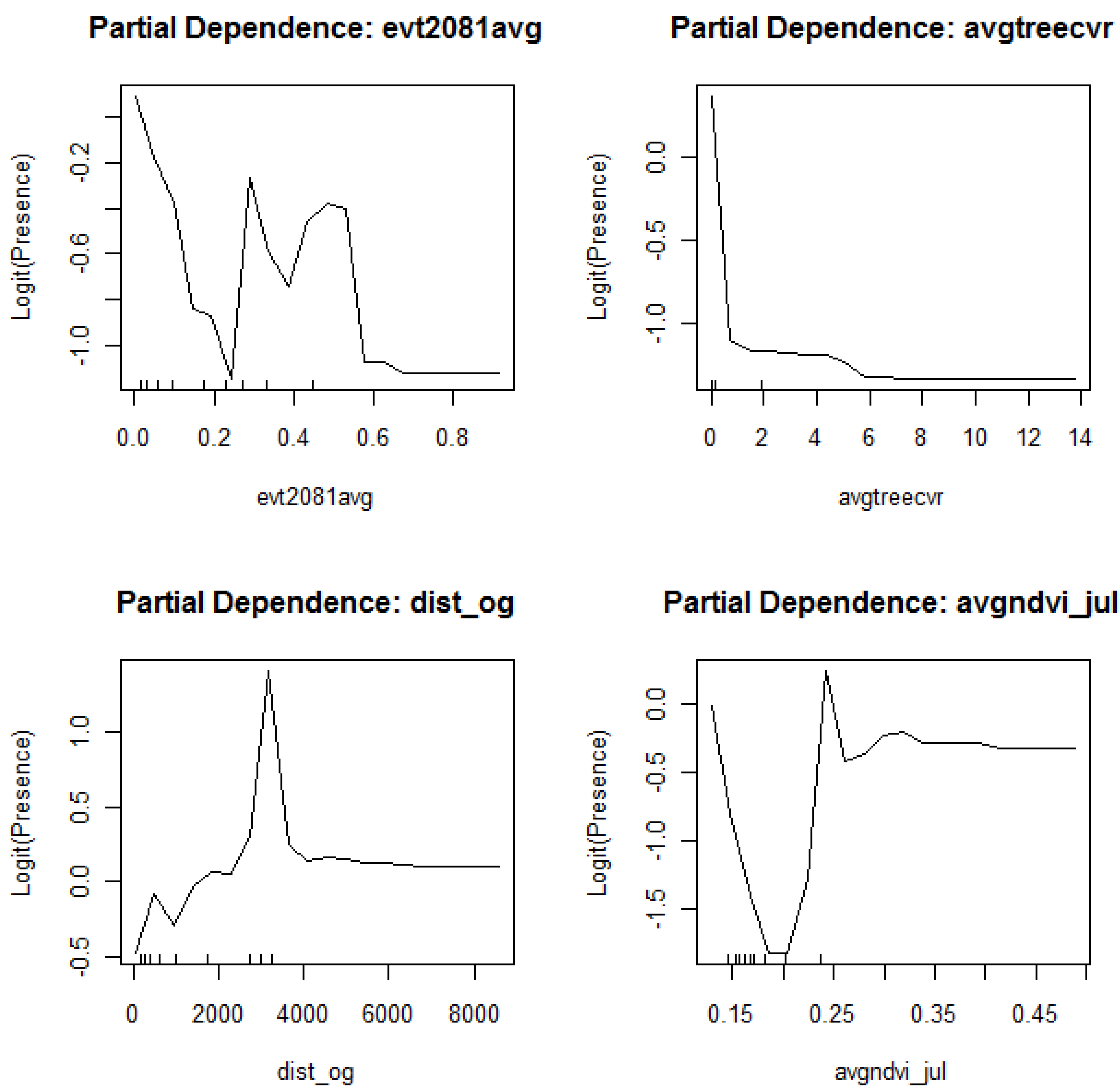

Figure A1.7: Summer model partial dependency plots depicting the marginal effect of *evt2081avg, avgtreecvr, dist_og,* and *avgndvi_jul* on the response (sage-grouse presence or pseudo-absence). The partial plots indicate the marginal effect these variables have on the class probability of the response (y-axis units are the log of the fraction of 'presence' votes of all trees in the RF model). Clockwise from top left, X-axes measure: proportion of vegetation within 510m classified as Inter-Mountain Basins Mixed Salt Desert Scrub (*evt2081avg*); average percent tree cover over a 510m radius neighborhood (*avgtreecvr*); meters from nearest active oil/gas well (*dist_og*); average July 15, 2008 NDVI value (range: (0,1)) over a 510m radius neighborhood (*avgndvi_jul*). The effect of the clustering of data points with respect to their distance from oil and gas wells is seen as the spike in the partial dependence plot for variable *dist_og*.

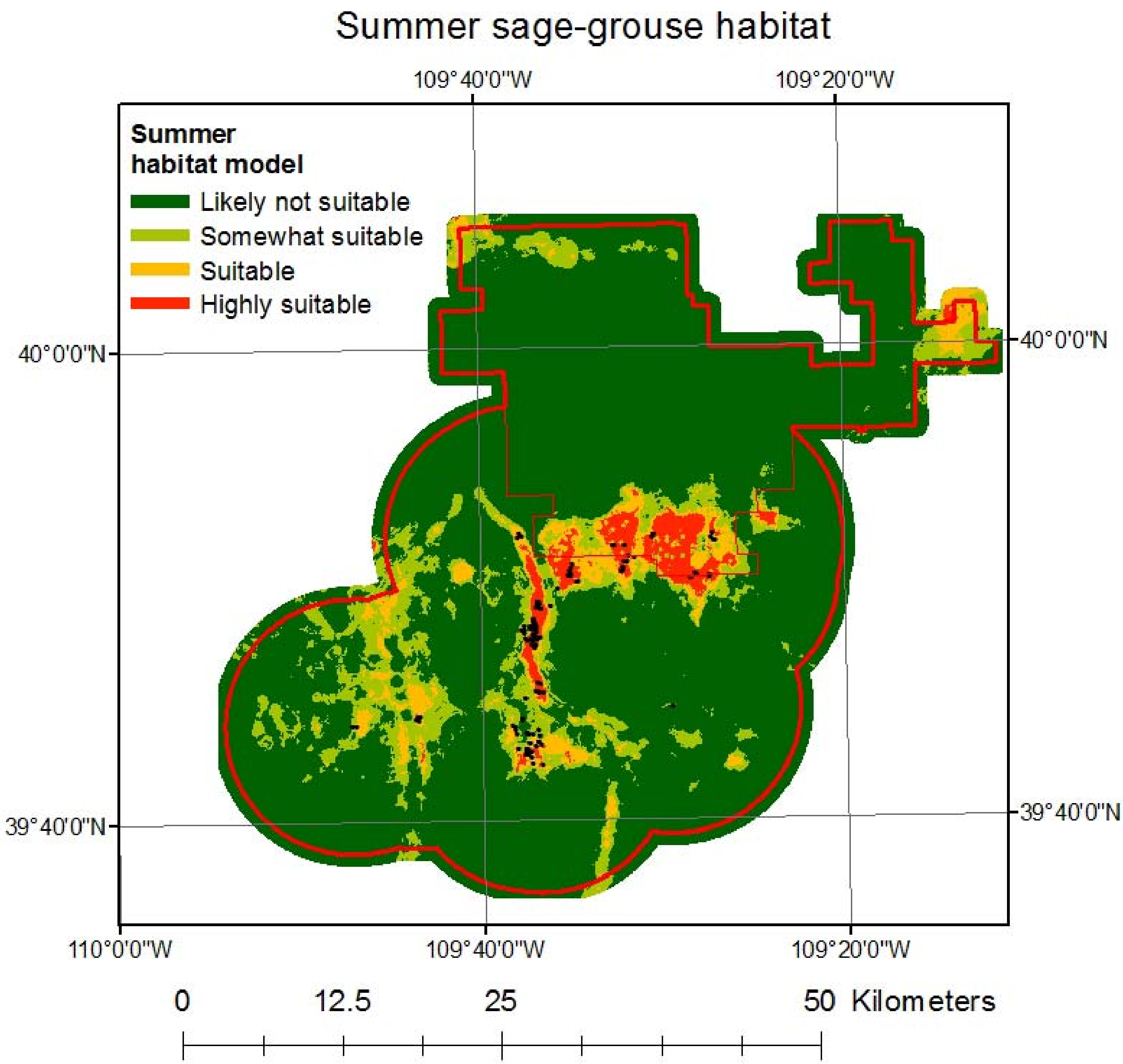

Figure A1.8: Summer sage-grouse habitat in the study area (thick red line). The southern boundary of Anadarko's AOI is indicated by the thin red line. Presence points used to train the model are shown as small black dots.

Year-round model results

Table A1.3. Year-round habitat model variable selection and accuracy assessments. Number of observations: 348 presences, 348 pseudo-absences = 696 points total.

Variable reduction step 1: Identify top 20 vars	Variable reduction step 2: drop correlated vars		Variable reduction step 3: Drop all but top 8 vars		Final model
Variable name	Mean Decrease Accuracy	Drop variable?	Mean Decrease Accuracy	Drop variable?	Mean Decrease Accuracy
og_dens3km	1.16		1.21		1.25
dist_peren	1.04		1.13		1.26
evt2064avg	1.03		1.11		1.21
rd_den3km	1.03		1.07		1.14
avgtreecvr	0.98		1.14		1.19
evt2016avg	0.96	# (cor w/ avgtreecvr)			
evt81avg	0.94		1.06		1.18
evt2066avg	0.93		0.93	#	
avgndvi_jul	0.92		1.08		1.15
avg_slope	0.9		0.94	#	
avggrn_jul	0.9	# (cor w/ avgndvi_jul)			
evt2081avg	0.89		1.02	#	
avgbrt_jul	0.87		1.05		1.15
avgndvi_apr	0.87	# (cor w/ avgndvi_jul)			
avggrn_apr	0.87	# (cor w/ avgbrt_jul)			
avgndmi_apr	0.85	# (cor w/ evt2081avg)			
avgbrt_apr	0.85	# (cor w/ avgbrt_jul)			
avg_cti	0.84	# (cor w/ avg_slope)			
avgwet_jul	0.83		0.91	#	
avgshrbcvr	0.82		0.93	#	
	<u>20 vars</u>		<u>13 vars</u>		<u>8 vars</u>
Sensitivity	97.41		97.70		97.99
Specificity	94.83		95.40		94.54
% Correctly Classified	96.12		96.55		96.26
Kappa	0.9224		0.9224		0.9224
AUC	0.9934		0.9940		0.9948

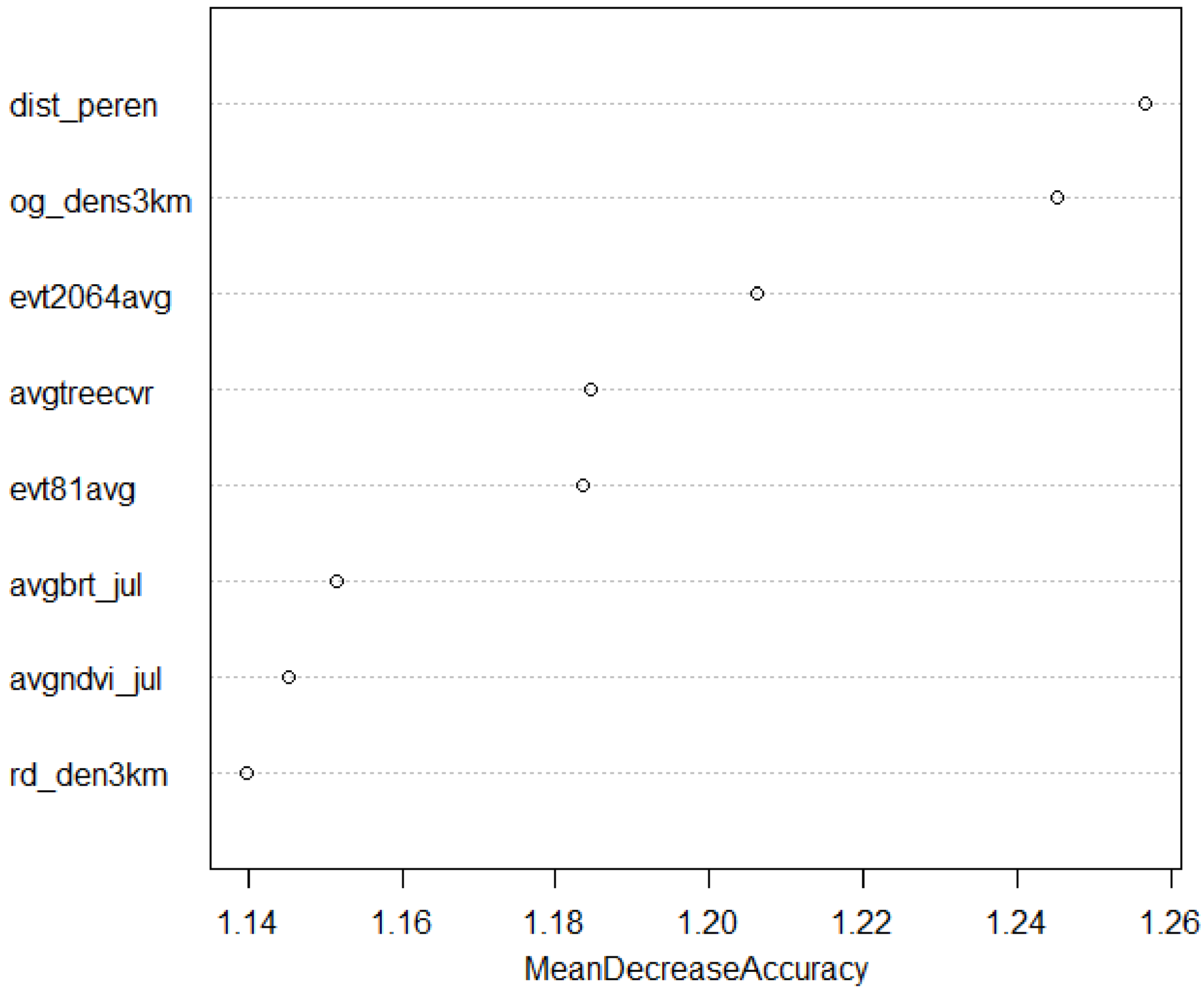

Figure A1.9: Variable importance plot for the year-round model. Mean decrease in accuracy is calculated as the average percentage decrease in OOB classification accuracy for the random forest when the variable of interest is permuted in the OOB sample and accuracy is recalculated.

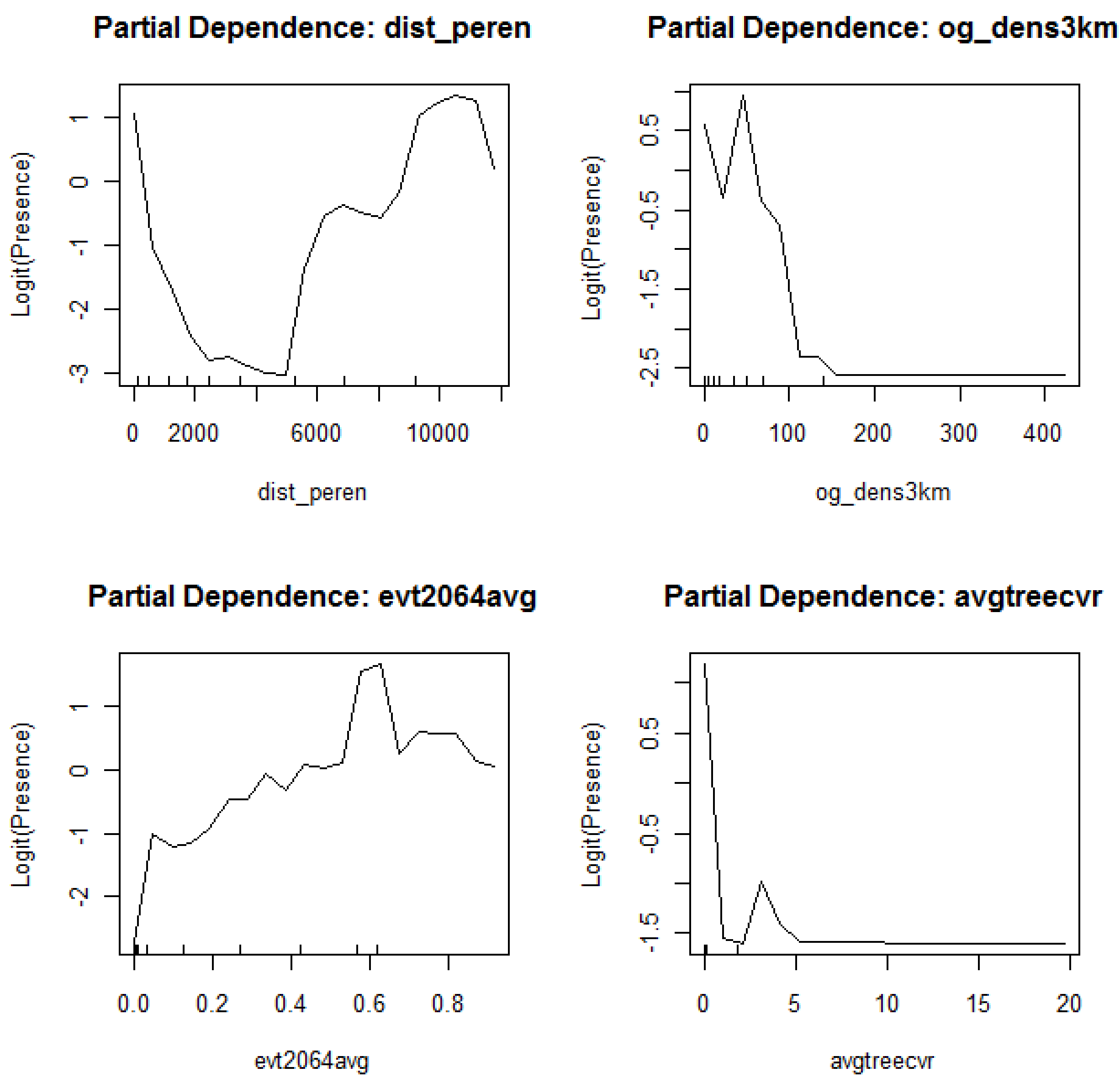

Figure A1.10: Year-round model partial dependency plots depicting the marginal effect of *dist_peren, og_dens3km, evt2064avg,* and *avgtreecvr,* on the response (sage-grouse presence or pseudo-absence). The partial plots indicate the marginal effect these variables have on the class probability of the response (y-axis units are the log of the fraction of 'presence' votes of all trees in the RF model). Clockwise from top left, X-axes measure: distance (meters) from the nearest perennial stream/waterbody (*dist_peren*); number of active oil and gas wells within 3km (*og_dens3km*); proportion of vegetation within 510m classified as Colorado Plateau Mixed Low Sagebrush Shrubland (*evt2064avg*); average percent tree cover over a 510m radius neighborhood (*avgtreecvr*).

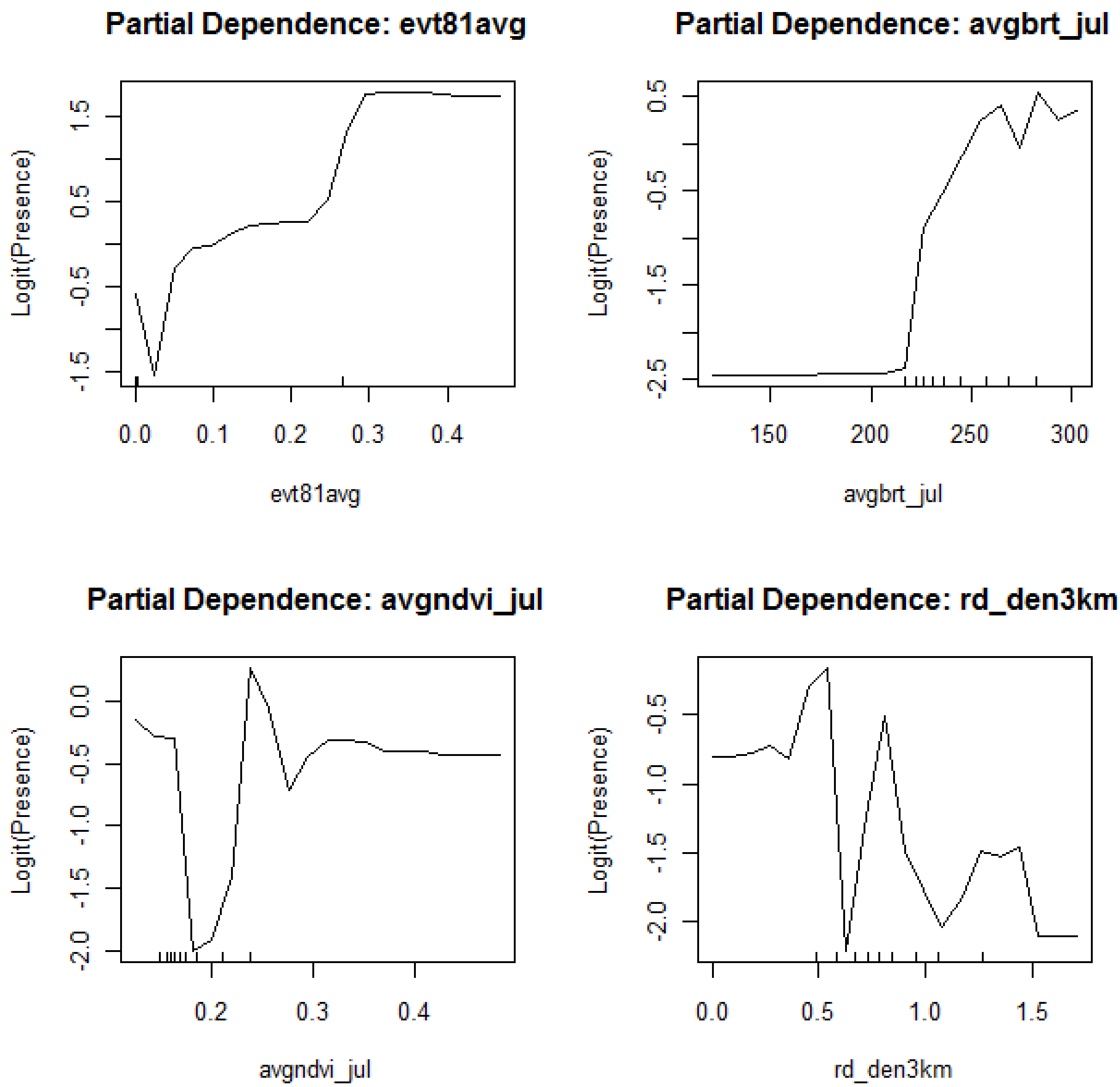

Figure A1.11: Year-round model partial dependency plots depicting the marginal effect of *avgtreecvr*, *rd_den3km*, *avgndvi_jul*, and *avgbrt_jul*, on the response (sage-grouse presence or pseudo-absence). The partial plots indicate the marginal effect these variables have on the class probability of the response (y-axis units are the log of the fraction of 'presence' votes of all trees in the RF model). Clockwise from top left, X-axes measure: proportion of vegetation within 510m classified as Agriculture-Pasture and Hay (*evt81avg*); average July 15, 2008 Brightness value over a 510m radius neighborhood (*avgbrt_jul*); average July 15, 2008 NDVI value (range: (0,1)) over a 510m radius neighborhood (*avgndvi_jul*); kilometers of roads within 3km of data points (*rd_den3km*).

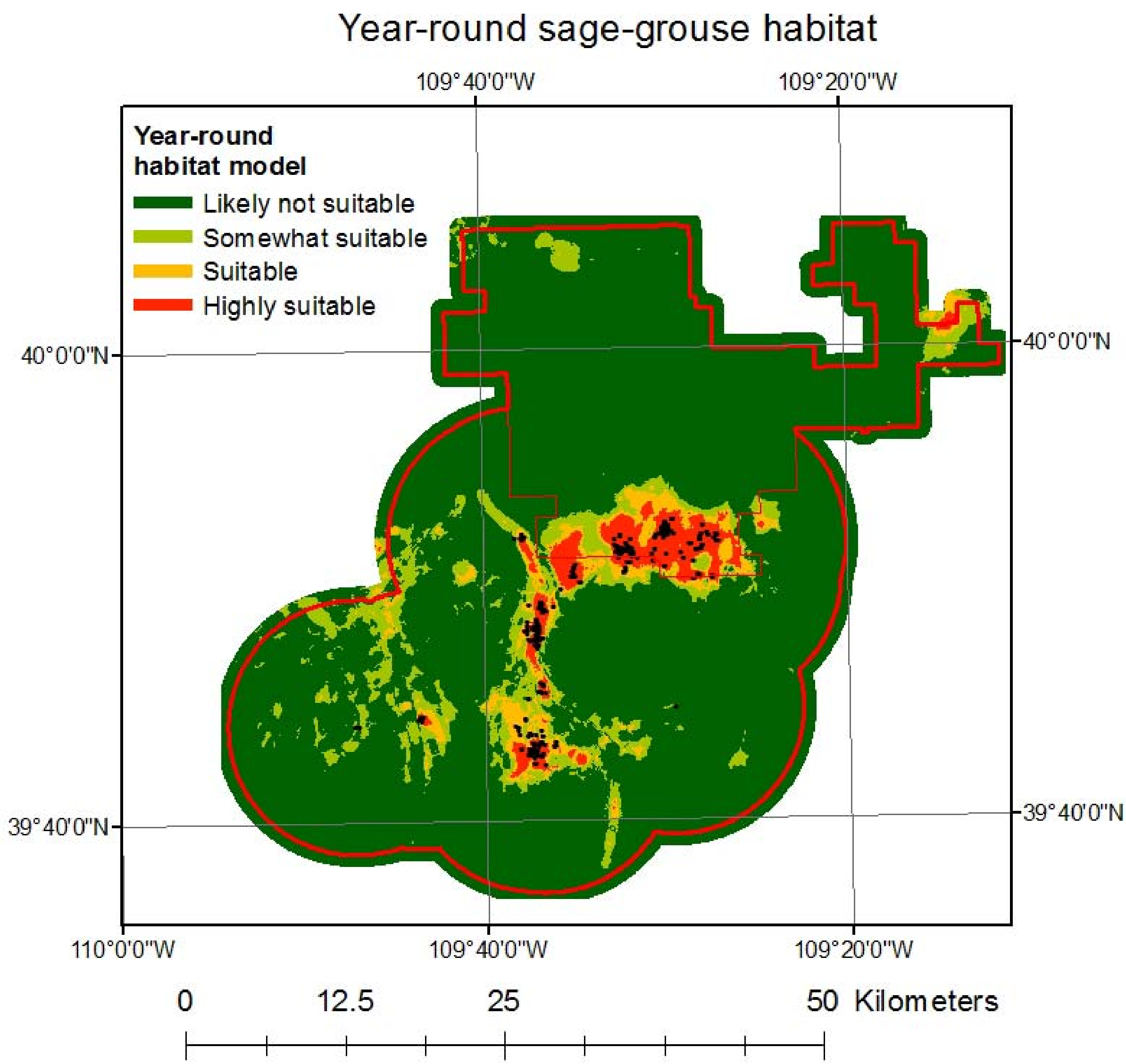

Figure A1.12: Year-round sage-grouse habitat in the study area (thick red line). The southern boundary of Anadarko's AOI is indicated by the thin red line. Presence points used to train the model are shown as small black dots.