ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/292970319

Importance of Regional Variation in Conservation Planning and Defining Thresholds for a Declining Species: A Range-wide Example of the Greater Sage-grouse

Technical Report · December 2015

READS
191

5 authors, including:

Jeffrey S Evans
The Nature Conservancy
64 PUBLICATIONS **2,452** CITATIONS
SEE PROFILE

Peter S. Coates
United States Geological Survey
51 PUBLICATIONS **331** CITATIONS
SEE PROFILE

Lara M Juliusson
U.S. Fish and Wildlife Service
11 PUBLICATIONS **106** CITATIONS
SEE PROFILE

Brad Fedy
University of Waterloo
28 PUBLICATIONS **328** CITATIONS
SEE PROFILE

Available from: Jeffrey S Evans
Retrieved on: 13 June 2016

Importance of Regional Variation in Conservation Planning and Defining Thresholds for a Declining Species: A Range-wide Example of the Greater Sage-grouse

Kevin E. Doherty[1*], Jeffrey S. Evans[2], Peter S. Coates[3], Lara Juliusson[1], Bradley C. Fedy[4]

[1] United States Fish & Wildlife Service, 134 Union Blvd, Lakewood, Colorado U.S.A.
*Corresponding Author, kevin_doherty@fws.gov, (303) 921-0524

[2] The Nature Conservancy, Fort Collins, CO 80524, U.S.A. & Department of Zoology and Physiology, University of Wyoming, Laramie, WY 82071, U.S.A.

[3] U.S. Geological Survey, Western Ecological Research Center, Dixon Field Station, Dixon CA, 95620.

[4] Environment and Resource Studies, University of Waterloo, Waterloo, Ontario, Canada

Table of Contents

INTRODUCTION 3

OBJECTIVES 5

STUDY AREA 5

METHODS 7

Breeding Habitat Model 7

Lek Survey Data 7

Habitat Selection and Scale 7

Pseudo-absence Data 7

Statistical Model 10

Evaluation of Model Fit and Predictions 10

Regional Variation in Habitat Selection and Disturbance Thresholds 11

Breeding Population Index Model 11

Kernel Index 12

Population Index 12

Aggregation using Population Index Volumes 12

RESULTS 13

Breeding Habitat Model 13

Breeding Population Index Model 15

DISCUSSION 15

Acknowledgments 16

LITERATURE CITED 22

APPENDIX I 25

APPENDIX II 32

APPENDIX III 33

Background photo: *Sagebrush in Wyoming* / K. Theule, USFWS

Cover background photo: *Greater sage-grouse hens* / Tom Koerner, USFWS
Cover subphotos: *Greater sage-grouse flying over Seedskadee National Wildlife Refuge* / Tom Koerner, USFWS, *Male and female greater sage-grouse* / Tom Koerner, USFWS, *Pronghorn herd* / Theo Stein, USFWS, *Greater sage-grouse in sagebrush* / USFWS

Abstract

We developed range-wide population and habitat models for Greater Sage-grouse *(Centrocercus urophasianus)* that account for regional variation in habitat selection and relative densities of birds for use in conservation planning and risk assessments. We developed a probabilistic model of occupied breeding habitat by statistically linking habitat characteristics within 4-miles of an occupied lek using a non-linear machine learning technique (Random Forests). Habitat characteristics used were quantified in GIS and represent standard abiotic and biotic variables related to sage-grouse biology. Statistical model fit was high (mean correctly classified = 82.0%, range = 75.4%¬ – 88.0%) as were cross-validation statistics (mean = 80.9%, range = 75.1% – 85.8%). We also developed a spatially explicit model to quantify the relative density of breeding birds across each Greater Sage-grouse management zone. The models demonstrate distinct clustering of relative abundance of sage-grouse populations across all management zones. On average approximately half of the breeding population is predicted to be within 10% of the occupied range. We also found 80% of sage-grouse populations were contained in 25 – 34% of the occupied range within each management zone. Our range-wide population and habitat models account for regional variation in habitat selection and the relative densities of birds, thus they can serve as a consistent and common currency to assess how sage-grouse habitat and populations overlap with conservation actions or threats over the entire sage-grouse range. We also quantified differences in functional habitat responses and disturbance thresholds across the Western Association of Fish and Wildlife (WAFWA) management zones using statistical relationships identified during habitat modeling. Even for a species as specialized as Greater sage-grouse, our results show that ecological context matters in both the strength of habitat selection (i.e. functional response curves) and response to disturbance.

Greater sage-grouse / Jeannie Stafford, USFS

INTRODUCTION

In an increasingly anthropogenic world where funding for conservation activities is limited, effective landscape-scale conservation planning tools have been progressively embraced by resource management agencies to maximize both conservation investments and reduce impacts of anthropogenic disturbances. This has corresponded with rapid expansion of landscape-scale, spatially-explicit models of species' habitat, such as resource selection functions (RSF) [1-3], which simultaneously give insight into the ecology of species and can be used to produce maps to help guide where conservation actions should be most effective. Often, RSF models do not encompass the entire range of a focal species and therefore biological relationships are extrapolated to novel areas, not included in the development of the RSF models, when decisions must be made. While extrapolating known relationships represents the best available information to decision makers, past work has shown limits in the transferability of RSF models to novel areas for a wide variety of taxa from low mobility plants [4] to semi-migratory birds [5].

A species response to particular habitat components can change as a function of the prevalence of the resource, which is referred to as the functional response of a species [6]. Understanding functional responses related to habitat selection through RSF modeling can elucidate, disturbance threshold values for habitat

quantity and quality, tolerance to perturbations, and cumulative effects [7]. This is important as conservation plans generally require targets for the amount of habitat required for specific species in order for managers to make cost effective decisions and balance competing interests [8]. Unfortunately, setting conservation targets based upon thresholds defined for other regions is precarious [7] because thresholds can vary tremendously across species and landscapes [9]. Landscape-scale modeling across broad extents is important in understanding how functional responses may vary for wide-ranging species, as landscapes are seldom homogeneous across large extents.

Data on the distributions of abundance are rare for most taxa, yet if available they can provide crucial baseline data for monitoring populations and conservation actions [10]. Abundance is often clustered and highly variable among sites across the range of a species, typically being high in relatively few sites and low in the majority [11]. Knowledge and mapping of population centers or 'hotspots' can be critically important for conservation planning as many species with broad distributions occur in densities of several orders of magnitude higher in hotspots compared to occupied habitats outside of hotspot boundaries [12]. Population centers of many species can be stable over several decades even while population sizes fluctuate [12]. Consequently, habitat protection can affect drastically different proportions of target populations depending on overlap with population centers. Ideally, conservation planning would make the best use of information related to population abundance and habitat requirements, while accounting for regional gradients and differences in functional responses. Spatial and temporal data in numerous ecological studies are often used for independent methods to derive proxies for population abundance and measures of habitat quality, which are helpful to inform environmental management and policy decisions [13]. However, when broad scale population survey data exist, probabilistic surfaces of density indices and habitat selection indices can be integrated to create analytical tools across broad spatial scales [14]. Integrative methodology could create composite spatially-explicit indices that reflect demographic and habitat information and make predictions to guide landscape-level conservation actions. Unfortunately, such data are rare in conservation planning because the broad scale population surveys are lacking for many species and habitat modeling is generally conducted at scales smaller than a species range.

Greater sage-grouse (*Centrocercus urophasianus*; hereafter sage-grouse) is a wide ranging species of conservation concern that occurs throughout the sagebrush ecosystem in the Intermountain West of the United States [15: see **Figure 1**]. Currently sage-grouse are considered "warranted, but precluded" for listing under the United States Endangered Species Act of 1973 (ESA; U.S. Fish and Wildlife Service [USFWS] 2010), Endangered under the Canadian Species at Risk Act, and "Near Threatened" on the International Union for Conservation of Nature (IUCN) Red List. Because of the wide reaching implications of an ESA listing on western lands within North America, monitoring sage-grouse populations is imperative to help inform land and wildlife management agencies responsible for regulatory actions and policies. Lek sites (traditional breeding grounds) provide opportunity to count sage-grouse annually and monitor demographic responses. Leks are typically located in nesting habitat where males are most likely to encounter females for breeding opportunities [16], and several studies support this hypothesis for both Greater Prairie chickens *(Tympanuchus cupido)* and sage-grouse [17-22]. Although sage-grouse leks have been counted each year since the 1950s, wildlife agencies have drastically increased their efforts in surveying known leks and searching for new lek sites since the 1990s (personal communication, Tom Remington Western Association of Fish and Wildlife Agencies). Broad-scale sage-grouse lek survey data managed by each state with sage-grouse provides a unique opportunity to identify sources of temporal and spatial variation in functional responses across the entire range of a species that inhabits most of western United States. Furthermore, findings from such analysis could be used to target thresholds for conservation planning activities for a species of increasingly high conservation concern.

OBJECTIVES

1. *Develop range-wide habitat and population models that identify regional variation in habitat selection and relative densities of sage-grouse for use in conservation planning and risk assessments.*

2. *Assess the importance of variability in habitat selection and thresholds of disturbance, and identify differences in functional responses across the range of sage-grouse.*

STUDY AREA

Our study area includes the entire range of North American sage-grouse populations with the exception of six active leks located in Canada (**Figure 1**). Canadian leks were not included in our modeling because of significant differences in available spatial data between the U.S. and Canada. Loss and degradation of native vegetation have affected much of the sagebrush *(Artemisia spp.)* ecosystem in western North America, and this ecosystem has become increasingly fragmented because of conifer encroachment, exotic annual grass invasion, and anthropogenic development [23]. The WAFWA Conservation Strategy for Greater Sage-grouse [24] delineated seven sage-grouse management zones to guide conservation and management (**Table 1**). The boundaries of these management zones were delineated based on differences in ecological and biological attributes (i.e. floristic provinces) rather than on arbitrary political boundaries [24] (**Figure 1**). Maps representing the major ecological gradients and subsequent dominant landcover types are shown in **Appendix I**. We stratified our analyses by sage-grouse management zones because spatial partitioning of data improves model fit where regional niche variation occurs [25] because of fundamental differences in the ecological gradients and different functional responses at regional scales.

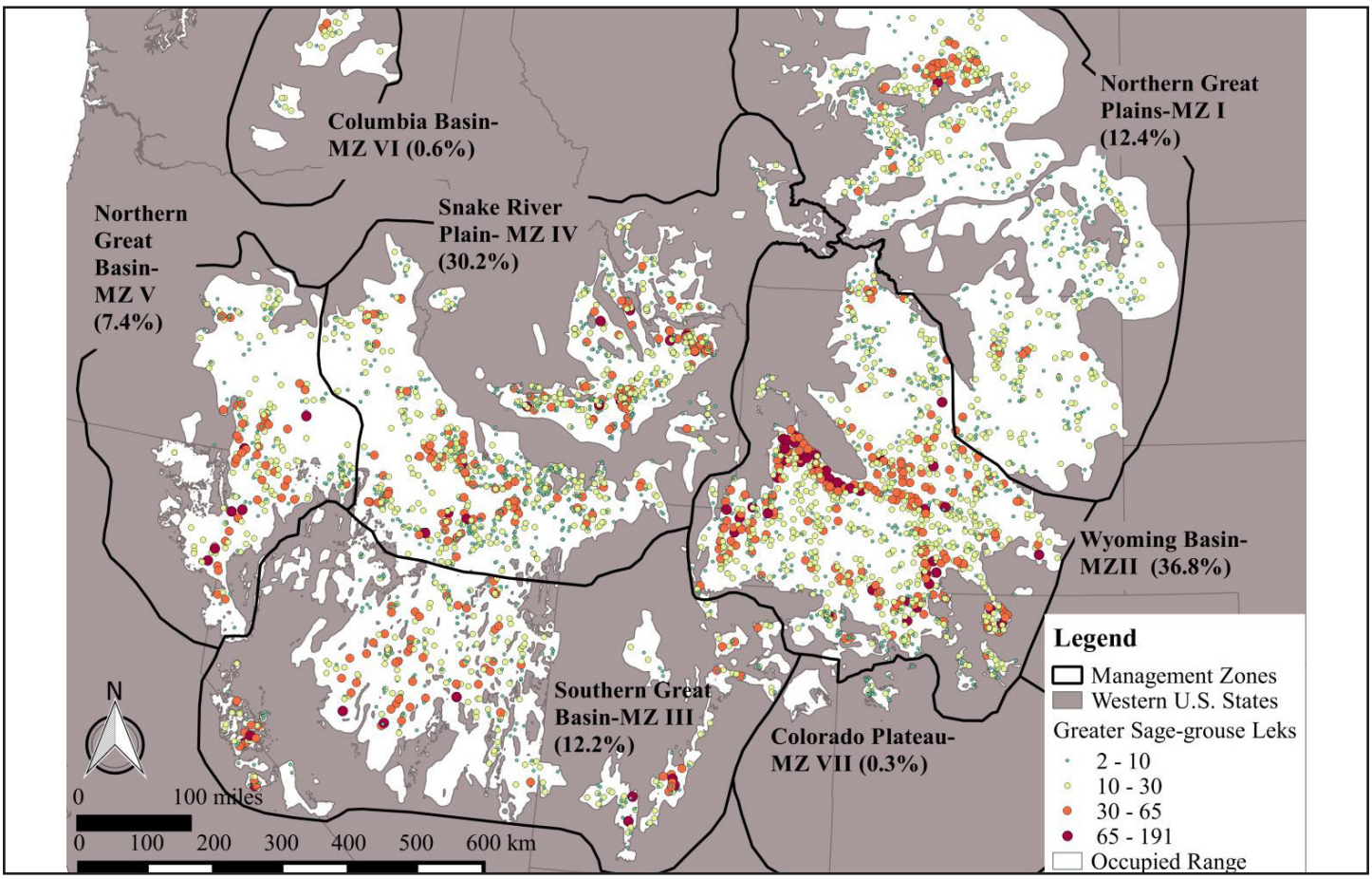


Figure 1. Location of Greater Sage-grouse Management Zones used to spatially subset analyses and the location of active Greater Sage-grouse leks counted during 2010-2014. Percentages are derived from the sum of the mean peak count of displaying sage-grouse at individual leks during 2010-2014 within each management zone divided by the range-wide total, to give context to the amount of known populations within each management zone.

Table 1. Ecological descriptions of Western Association of Fish and Wildlife Agencies Greater Sage-grouse Management Zones.

Management Zone	Ecological Descriptions of Management Zones [a b]
Northern Great Plains (MZ 1)	The Northern Great Plains includes the north-eastern portions of the sage-grouse range. This management zone experiences the most precipitation, thus it contains larger portions of the landscape dominated by grasslands, smaller patches of sagebrush, and contains more silver sagebrush *(Artesmisa cana var. cana)* than other management zones. MZ I also has the highest amount of land in private ownership and compared to other management zones, it has the highest amount of cropland.
Wyoming Basin (MZ II)	The Wyoming Basin is characterized by large expanses of Wyoming big sagebrush *(Artemisia tridentata var. wyomingensis)* with little fragmentation; however it experiences the greatest amount of oil and gas development. Most of the precipitation in this management zone comes in the form of winter snowfall. MZ II contains the highest densities of sage-grouse across their range.
Southern Great Basin (MZ III)	The Southern Great Basin includes the southern and western most populations of sage-grouse. MZ III is the most arid of all the management zones and includes a mix of Wyoming big sagebrush, mountain big sagebrush *(A. tridentata var. vaseyana)*, low sagebrush *(A. arbuscula)*, and black sagebrush *(A. nova)*. Topography is rugged with sagebrush on many of the valley floors transitioning to arid coniferous forests at higher elevations on the mountain slopes.
Snake River Plain (MZ IV)	The Snake River Plain encompasses the north-central populations of sage-grouse. Like MZ's III and IV, it is characterized by salt deserts in the lower elevations and conifer forests at higher elevations. Wyoming big sagebrush and basin big sagebrush *(A. tridentata var. tridentata)* are the dominant species, with mountain big sagebrush at higher elevations. MZ VI contains the second highest density of sage-grouse across the species range. The Snake River Plains management zone also experiences dense cropland areas, however they are clustered at lower elevations[b].
Northern Great Basin (MZ V)	The Northern Great Basin is similar to the Southern Great Basin, but it is less arid with precipitation occurring primarily in the winter and spring. Similar to MZ's III & IV, lower elevations are dominated by salt deserts and higher elevations are dominated by conifer forest.
Columbia Basin (MZ VI)	The Colombia Basin is isolated from the rest of the sage-grouse range and is contained entirely within Washington state. Wyoming big sagebrush and basin big sagebrush are predominate species. MZ VI contains the lowest elevation sagebrush across the range and experiences high amounts of cropland in comparison to all other management zones with the exception of the Northern Great Plains.
Colorado Plateau (MZ VII)	The Colorado Plateau is the south-eastern most management zone and contains a small fraction of the overall sage-grouse populations. It is similar to the Southern Great Basin MZ, but it receives more precipitation. Soil types within the Colorado Plateau greatly restrict the sagebrush distribution and it contains a very small portion of the overall occupied habitat.

[a] Descriptions of management zones were originally summarized [1] and adapted by WAFWA for analyses for both the 2004 Conservation Assessment of Greater Sage-grouse and Sagebrush Habitats [2] and 2006 Greater Sage-grouse Comprehensive Conservation Strategy [3].

[b] We created maps of the ecological gradients and major land cover types between Greater Sage-grouse Management Zones for further reference in **Appendix I**. We focused maps on the major ecological gradients and subsequent landcover **(Figures 3-9)**.

METHODS

Breeding Habitat Model

We developed a binomial probabilistic model of occupied breeding habitat by quantifying habitat characteristics, within 6.4 km (4-miles) of an occupied sage-grouse lek, for available points using a classification instance of the non-parametric model Random Forests [26-29]. Model predictions produce an estimated probability for each 120-m2 grid cell within each sage-grouse management zone. Components of sage-grouse habitat were compiled into a GIS database from various sources, but generally represent standard abiotic and biotic variables used in past work to represent sage-grouse habitat (**Table 2**).

Lek Survey Data

We compared active lek locations to pseudo-absence locations to generate models of breeding sage-grouse habitat across the range. The hotspot hypothesis of lek evolution suggests leks are typically located in nesting habitat where males will most likely encounter pre-nesting females who are attracted by important resources [17,30], such as forbs required for pre-breeding [31] and sagebrush cover for nesting [32]. Additionally, most sage-grouse nesting locations are located within 6.4 km of a lek [19,33,34]. Further, recent studies have shown that telemetry-based models of nesting sage-grouse predicted almost 2 times more nesting habitat around leks than at random locations [5,20]. We therefore believe sage-grouse lek locations are a good predictor of important breeding areas. We used lek data assembled and proofed by WAFWA to develop both our breeding habitat model and breeding population index model. For the purposes of both models, a lek was defined as active if > 2 males were counted during 2010-2014 and the last count was not a zero.

Habitat Selection and Scale

The desired geographic scale of understanding is paramount in studies aimed at obtaining inference on selection behavior [31]. Our study was specifically designed to assess first-order selection of sage-grouse seasonal home ranges [35,36]. The rationale for using the first-order scale was two-fold: (1) the primary objective was to develop population and habitat models that account for regional variation within each sage-grouse management zone, and (2) broad-scale lek data represent locations of populations and are not adequate to appropriately model second or third-order habitat selection [35,36]. Lower orders of habitat selection are generally derived from finer scale telemetry data at the individual level. Our results will be highly relevant to conservation practitioners and researchers working at regional spatial scales, because research conducted at range-wide extents are rare for most species, but can be critical to making informed decisions. For example, findings could inform the first-order of multi-scale habitat selection models using integrated resource selection functions [37], or hierarchical Bayesian models [38]. Using first-order assessments here that produce a relative probability for each 120-m2 grid cell across the range of the species allows for later integration with other research including at finer scales (e.g. 2nd – 4th orders). We believe investigating first-order habitat selection across the entire sage-grouse range is important, because understanding landscape context can elucidate why results of second and third-order habitat selection studies can seemingly give conflicting results and varying thresholds, even for well-studied topics [39].

Pseudo-absence Data

Recent survey efforts have been intensive enough that although not all leks have been identified, we are confident that the spatial process is well represented in the data. To generate pseudo-absence locations we modeled the spatial process of known leks, using an isotropic kernel estimate [40], and use the inverse of the density estimate to weight samples. A gradient function allowed for a tension parameter to control the proximity of pseudo-absence locations in relation to known lek locations. We utilized the pseudo-absence model available in the spatialEco library [41] and defined the sigma (distance smoothing for the kernel; bandwidth) as 18 km and the gradient as 1 thus, providing no weighting to the pseudo-absence diffusion process. This ensured that we were sampling the range of habitat variation within each sage-grouse management zone. To avoid class imbalance [42] (ie., zero-inflation) issues, we generated an equal ratio of pseudo-absence to lek locations and compared resulting sample variation against population data (rasters) to evaluate if we had an adequate sample to support model fit, spatial estimation, and inference. We chose an 18 km bandwidth

Table 2. Description of explanatory variables used to predict the occupied Greater Sage-grouse breeding habitat across 11 western U.S. States during 2010–2014. All variables with the exception of the climate date predictor group were quantified using at a 6.4 km buffer moving window (130.1 km²).

Name	Abbreviation	Source (years)	Native resolution	Resampled resolution	Description [a]	Justification [references]
	General Habitat Predictor Group					
Low Sagebrush	LowSB4mi FM	LANDFIRE EVT 1.2 (2010) [b]	30 m²	120 m²	% of grid cells classified as Low Sagebrush	Established positive relationship between sage-grouse abundance and sagebrush [7]
Tall Sagebrush	TalSB4mi FM	LANDFIRE EVT 1.2 (2010) [b]	30 m²	120 m²	% of grid cells classified as Low Sagebrush	Established positive relationship between sage-grouse abundance and sagebrush [7]
All Sagebrush	ALLSB4mi FM	LANDFIRE EVT 1.2 (2010) [b]	30 m²	120 m²	% of grid cells classified as Low Sagebrush	Established positive relationship between sage-grouse abundance and sagebrush [7]
Canopy Cover	CC4miFM	LANDFIRE Fuels 1.2 (2010)	30 m²	120 m²	% Canopy Cover in 10% increments from 15% – 95%	Established negative relationship between sage-grouse and conifers [8-10]
Grassland/ Herbaceous	GH4mFM	LANDFIRE Fuels 1.2 (2010)	30 m²	120 m²	% of grid cells classified as Grassland	Established negative relationship between sage-grouse abundance and grasslands [7]
Perennial Water	NHD_Perennial4miFM	National Hydrological Dataset NHD (2012)	Vector of Lines & Polygons	120 m²	NHD perennial flow lines within a 6440m moving window, multiplied by the average line length per cell (133.2m)	Established negative relationship of riparian areas to nest site selection [11]and established positive relationship between sage-grouse populations and riparian habitats [12]
Intermittent Water	NHD_Intermittent-4miFM	NHD (2012)	Vector of Lines & Polygons	120 m2	See Perennial Water	See Perennial Water
Springs and Seeps	SpringsandSeepsDensity4miFM	NHD (2012)	Vector of Lines & Polygons	120 m2	See Perennial Water	See Perennial Water
Topographic Wetness Index	TWI4mi FM(SD)	NHD (2012) & NED elevation Data (2013)	30 m²	120 m²	Index of wetness	See Perennial Water
	Climatic Data Predictor Group					
Gross Primary Production	GPP(years) FM(SD)	MODIS NASA EODP (2009-13)	1 km²	120 m²	Index of early Brood Rearing Habitat (mean of GPP from 5-15 through 6-15)	Forbs are important predictors of early brood survival and habitat selection [11]
Degree Days > 5°C	dd5120m [c]	USFS (1961-1990) [13]	1 km²	120m²	Number of days that reach a temperature ≥ 5°C	Large scale ecological driver of land types. Hypothesized regional scale relationship between sagebrush landscapes with higher production. Documented carry over effects [12]

Mean Annual Precipitation	map120m [c]	USFS (1961-1990) [13]	1 km²	120m²	Mean annual precipitation (mm)	Large scale ecological driver of land types. Hypothesized regional scale relationship between sagebrush landscapes with higher production. Documented carry over effects [12,14]
Annual Drought Index	adi120m [c]	USFS (1961-1990) [13]	1 km²	120 m²	Ratio = dd5 / map	Large scale ecological driver of land types. Hypothesized regional scale relationship between sagebrush landscapes with higher production. Documented carry over effects [12]
	Landform Variables Predictor Group					
Roughness	Rough4miSD	National Elevation Data NED (2013)	30 m²	120m²	Standard deviation in elevation within a 6440m buffer of a grid cell	Established negative relationship between sage-grouse and rough terrain [8,10]
Elevation	Elev4mi FM	NED (2013)	30 m²	120m²	Average elevation within a 6440m buffer of the grid cell	Hypothesized relationship between grouse populations and areas with higher productivity because of elevation
Steep	Steep4mi FM	NED (2013)	30 m²	120m²	% of landscape classified as steep using Theobald LCAP tool	Established negative relationship between sage-grouse and rough terrain [8,10]
	Disturbance Variables Predictor Group					
Human Disturbance Index	dist_noag4m-FM	NLCD Disturbed Classes[d] (2011)	30 m²	120m²	Landcover types associated with human presence	Established negative relationship between sage-grouse and human activity [15,16]
Oil & Gas Wells	ogptdsm4(2) miFM	IHS oil and gas database (19xx-2014)	Point	120m²	Density of oil and gas well locations [a]	Established negative relationship between sage-grouse and oil and gas development [17,18]
Burned Landscapes	F(-Years)4miFM	WFDSS-GeoMac Fire Perimeters (2000-08, 2009-2013, 1984-2013)	Vector of Polygons	120m²	Proportion of grid cells that are burned within a 6440 m area	Established negative relationship between fire and sagebrush habitat [19,20]
Agriculture Lands	tillage4mi FM	NASS (2008-2014)	30 m²	120m²	Proportion of grid cells that have been tilled since 2008 within a 6440 m area	Established negative relationship between sage-grouse and cropland [10,21]

[a] All variables were resampled to a 120m² pixel. All moving windows were calculated at a 6440-m (4-mile) buffer. Oil and gas layers were also calculated at a 2-mile moving window because of variations in the distance the impact was detected [15]. We did not use the 120m2 pixels for modeling because leks are a surrogate of habitat at a larger scale.

[b] Landfire vegetation groupings defined in Johnson et al. SAB [22].

[c] Because climate grids native resolution change at a 1-km scale and are highly spatially correlated, we did not resample the grids using a 6440m moving window.

[d] NLCD Urban development classes: Developed-High Intensity, Developed-Low Intensity, Developed-Medium Intensity, Developed-Open Space and NLCD impervious surfaces. The index also included roads (TIGER), oil & gas wells (compiled by each state), Wind Turbines (FCC obstruction database), Transmission lines (Ventyx), Pipelines (Ventyx)

because recent research has shown this scale represents the scale at which breeding populations move across the landscape to fulfill other seasonal habitat needs [43] and because we specifically designed our study to capture large first-order habitat selection. To accurately define 1st order sage-grouse habitat availability extents, we matched the spatial scale of availability to the desired scale of inference because matching such scales is critical to obtaining reliable estimates on selection behavior [44].

Statistical Model

Nonparametric methods are becoming much more common in ecological modeling, supporting inference of nonlinear and spatial dynamics [26-29]. Random Forests uses multiple realizations of the data, with no distributional assumptions, that effectively converges on a stable estimate in very high dimensional statistical spaces [27,45]. Model interpretation and inference was supported following methods presented in [26,27,45]. The expected complexity in interaction effects, potential latent variables, high spatial variability representing both global and local effects and nonlinear relationships, all support a non-linear model such as Random Forests as an appropriate choice.

We modeled selection of breeding season habitat within the species range [35,36] using Random Forests, which is a bootstrapped Classification and Regression Tree (CART) approach [46]. Random Forests is based on the principle of weak learning, where a set of weak subsample models converge on a stable global model. This method has been shown to provide stable estimates while being robust to many of the issues associated with spatial data (e.g., autocorrelation, non-stationarity). It also fits complex, nonlinear relationships, accounts for high dimensional interaction effects and accounts for hierarchically structured data inherent in non-stationary processes [26,27]. We expected both global trends in sage-grouse habitat selection as well as localized variation in habitat selection within each of the 7 sage-grouse management zones. First and second order variation are addressed in the hierarchical nature of the iterative node partitioning, making this a good model to implement when global trend and local variation [47] are expected to occur in the same model [27]. Analysis was conducted in program R [48] using the rgdal [49], sp [50] and raster [51] libraries to read spatial data, assign values from spatial covariates to the point observations of our dependent variable, and make spatial predictions. We used the implementation of Random Forests [52] in the R library randomForest [53] and followed the model selection method introduced in Murphy et al. (2010) using the rfUtilities library [54]. Parsimony in Random Forests is important not only for producing a more interpretable model but also reduces any fitting of the model to statistical noise, thus providing a better model fit [27,45].

Evaluation of Model Fit and Predictions

Model fit is an important assessment in evaluating the utility of a model, however one can argue that measures of prediction and model stability are even more important for conservation planning and risk analyses. This is because the biggest danger to generalization and thus spatial predictions to novel areas, is over-fitting of the training data [55]. Hence validation should consist of evaluating the degree of over-fitting, by comparing model performance on training and holdout sets. If performance is significantly better on the training set, over-fitting is implied [55]. To assess model fit, we used OOB error (out of bag) and confusion matrixes [53]. The OOB error represents the internal evaluation of global and class error against the withheld data from the Bootstrap, and represents an error distribution across all Bootstrap replicates in the ensemble where the median error is used to represent the OOB error. We evaluated model stability and performance using cross-validation methods [27], where 10% of data were withheld from training the model and used as a validation dataset. Over-fitting was assessed by comparing error rates between OOB and cross-validation. We would have concerns of over-fitting the data if error rates were much higher in the OOB sample vs. our k-fold cross validation error rates.

We also tested the sensitivity of the fitted model to errors in classification between used vs. available locations in the rfUtilities library [54] by randomly changing known lek locations to pseudo-absence points and evaluating cross-classification errors. We systematically changed known lek locations to zeros in 5% increments to understand the influence of pseudo-absence errors on overall error rates and model stability. This was done because an unknown

portion of our pseudo-absence locations were expected to fall within suitable sage-grouse breeding habitat. The primary motivation behind implementing a sensitivity test was to address model sensitivity to any lack of independence. A pseudo-replication problem would also affect the independence (correlation) of the Bootstraps and potentially overfit the model. Since ensemble models are based on the premise of weak learning and variation in the Bootstrap, if the data is homogenous, the Bootstraps would not be independent and the ensemble would exhibit considerable correlation and effectively overfit the model. In evaluating model fit and convergence we did not observe any indication of ensemble correlation. The sensitivity test allowed better understanding of overall error rates within our model and more importantly, it allowed assessment of model stability and prediction congruency across a range of lek locations that are misclassified as pseudo-absence.

Regional Variation in Habitat Selection and Disturbance Thresholds

We used probability partial plots to elucidate habitat relationships of the modeled covariates after partialling out (holding constant) the other variables in the model. To improve interpretability, we plotted each given covariate for all management zones on the same plot. The probability partial plots were derived using the rfUtilities library [54]. Management Zone VII: Management zone VII, while modeled, has a very small sample size (~ 0.3% of counted birds between 2010 – 2014) and only contains 652 km² of the 192,381 km² modeled breeding habitat (see **Table 5)**. Therefore we did not focus on these results in the general manuscript or include MZ VII in Figures highlighting functional habitat responses.

Breeding Population Index Model

Knowledge of high-abundance population centers for priority species represents a starting point to frame regional conservation initiatives and can direct management actions to landscapes where they will have the largest benefit to regional populations [56,57]. We developed a model to quantify the relative density of breeding birds within each management zone. This was motivated by past work across the range that showed sage-grouse populations are highly clustered [24,33,58]. Fortunately, sage-grouse are one of the few species in which extensive data sets exist on distribution and relative abundance across their entire breeding distribution, making an analysis of this scale possible [15,58].

To map high-abundance To map high-abundance population centers, we followed methods and logic very similar to models developed by the U.S. Geological Survey (USGS) for Nevada and the Bi-state populations of sage-grouse [38]. Distribution models that combine information about habitat quality and abundance of sage-grouse from multiple data sources are valuable given recent intensification of sage-grouse management and policymaking [38]. We modified their methods [38] to better represent a sage-grouse population index, because their original technique was developed to highlight management priority areas. Our final population index model incorporated two standardized kernel based point density models, representing local and regional scales and our Breeding Habitat Model described above. The results of our models are grids that represent an index to the relative amount of breeding birds for each 120 m² within each management zone. Our final population index model incorporates spatial patterns of sage-grouse habitat selection with contemporary information of abundance allowing the use of the available data [13,38]. Population indexes, such as ours, allow conservation actions to be targeted to the right landscapes, and help identify threats to a species that are occurring in areas which could impact large proportions of sage-grouse populations [13,14]. proportions of the sage-grouse populations [13,14].

Greater sage-grouse

Kernel Index

Kernel density functions have been commonly used in ecology to delineate home ranges of individual animals and to map concentrated areas of use by populations [59,60]. Within our study, we used the kernel density function to groups cells of concentrated use by attributing count data to a grid placed over top of a sage-grouse management zone [59,60]. Using kernels to define population concentrations is consistent with past work defining core areas for sage-grouse [33]. We created two kernel models based on two separate bandwidth values (i.e., 6.4 and 18 km), which reflect published information on sage-grouse movement and seasonal space use patterns. The 6.4 km bandwidth was chosen to correspond with utilization distribution of areas conducive for reproduction in relation to lek sites (e.g., breeding, nesting, brood-rearing), as demonstrated in populations at multiple sites [19,33,34]. Although leking areas generally serve as hubs for nesting and are usually centered across seasonal areas [34], some sage-grouse move relatively long distances to access wintering areas [34,43]. Thus, we incorporated the larger spatial scale of 18 km to reflect these life history patterns [43]. Combining the scales appropriately placed greater emphasis on adjacent areas, thus preventing oversmoothing, but still allowed for representation of sage-grouse occurrence at further distances. We used SAGA GIS version 2.1.0 [61] to create two Gaussian kernel density functions. The same set of active lek locations from our habitat model defined the point density for our kernel models and each point was weighted by the mean peak count of displaying sage-grouse from 2010-2014. Following logic of Coates et al. [14], we standardized each kernel using a row standardization. We then added each grid together and divided by 2, using the raster library [51] in R. The output is a 120 m2 raster that represents a multi-scale density process of sage-grouse lek counts across two biologically meaningful scales (**Equation 1**).

Equation 1: Kernel Index = (standardized 6.4 km kernel + standardized 18 km kernel) / 2

Population Index

Our Kernel Index summarizes the best available information on the relative density of birds across the entire sage-grouse range. We selected bandwidths to correspond with linear movement distance of sage-grouse within the breeding season [33], as well as movements between breeding and other seasonal habitats [43]. We believe the combination of both kernels into a single Kernel Index represent ecologically meaningful areas for sage-grouse. However, kernel functions are inherently an estimator of the spatial point-density process, thus they are not explicitly linked to habitat features.

We wanted to create a population index to further refine our Kernel Index. First, we wanted a method that would reduce the importance of lands with low probabilities of being habitat based upon known sage-grouse habitat relationships. Secondly, we wanted to increase the value of lands with high probabilities of being occupied habitat, but further away from known leks, thus having lower value in the Kernel Index. We did this by multiplying the Kernel Index by the probability of our Breeding Habitat Model (**Equation 2**).

Equation 2: Population Index = (Kernel Index * Breeding Habitat Model)

Highest population index values arise where high breeding habitat probabilities co-occur with landscapes having higher lek counts. The use of this equation also effectively reduces the value of landscapes near larger sage-grouse leks predicted to be non-habitat. Lastly, multiplying the Kernel Index by the breeding habitat model increases the value of lands further from known sage-grouse leks that have high probabilities of containing breeding sage-grouse. We felt this was important because our data set utilized all known sage-grouse population survey data across their range, however our survey data do not represent all leks.

Aggregation using Population Index Volumes

We ordered all population index values from each grid cell within a management zone from the highest to lowest density. We selected the highest density cells in order until they summed to 10% of the total population index within a management zone. We repeated the selection process in 10% increments selecting the highest remaining grid cell densities first until we had 10 bins (i.e. highest density bin represented the top 10% of the population, 100% bin representing all breeding areas identified in modeling). Results are cumulative, such that all bins contain all preceding bins of 10% increments. We then calculated the percent of the occupied distribution that each incremental 10% population bin covered.

RESULTS

Breeding Habitat Model

On average our breeding habitat model correctly classified 82.0% (Range 75.4% – 88.0%) of hold-out data from out-of-bag bootstrap samples (**Table 3**). Our models also correctly classified independent k-fold hold-out data (mean across management zones = 80.9%; range 75.0 – 85.8%) (**Table 3**). General agreement between out-of-bag error rates and k-fold cross validation indicate stability in our model to predict independent data and lack of over-fitting (**Table 3**). We documented higher error rates within pseudo-absence classes compared to our active lek class (**Table 4**), however simulations indicated estimates were stable across a wide range of pseudo-absence errors (0 – 30% simulated errors in 5% increments). For example, the mean standard error (SE) across the 7 management zones with 20% simulated pseudo-absence errors = 0.032. Low SEs indicates model stability and the ability of the random forests to predict though statistical noise arising from false absences. We documented an ~ 3% error increase for every 5% increase in false-absences.

Models demonstrate breeding habitat is highly condensed within the current occupied range of sage-grouse (**Figure 2**). All currently active leks occurred on probabilities > 0.65; we therefore used this threshold to quantify the amount of breeding habitat. When we use this threshold value, 26% of the current occupied range is predicted to be breeding habitat (**Table 5, Figure 2**). Across the range of sage-grouse, general habitat variables and climatic gradient variables had greater

Table 3. Percent of K-fold cross validation hold out data set locations (10%) which were correctly classified by a model built with 90% of the data set. These results are compared to internal model fit statistics generated via bootstrap resampling (1 – Out of Bag Error Bootstrap Error Rates).

Management Zone	1 - Out of Bag Error	K-Fold Cross Validation % Correctly Classified
MZ I – Northern Great Plains	76.3	75.9
MZ II – Wyoming Basin	75.4	75.0
MZ III – Southern Great Basin	85.9	85.3
MZ IV – Snake River Plain	83.9	83.6
MZ V – Northern Great Basin	76.3	75.1
MZ VI – Columbia Basin	88.0	85.8
MZ VII – Colorado Plateau	88.0	85.4
Avg.	82.0	80.9

Table 4. Classification confusion error rates for leks and pseudo-absence locations. Error rates were generated from bootstrap resampling. Across management zones there was a general pattern of higher errors in the pseudo-absence class, with the exception of the two smallest management zones, the Columbia Basin and the Colorado Plateau.

	Pseudo-absence	Leks
MZ I – Northern Great Plains	29.8%	16.9%
MZ II – Wyoming Basin	32.7%	16.5%
MZ III – Southern Great Basin	18.0%	9.9%
MZ IV – Snake River Plain	21.0%	11.3%
MZ V – Northern Great Basin	29.4%	19.7%
MZ VI – Columbia Basin	12.0%	12.0%
MZ VII – Colorado Plateau	12.0%	10.0%

Table 5. Area (km^2) of Occupied Range [23] and modeled breeding habitat across the Greater Sage-grouse range in North America[a]. Breeding habitat probabilities were calculated using a 0.65 threshold, because all current active leks had a probability > 0.65.

Management Zone	Occupied Range	Modeled Breeding Habitat	Percent of Occupied Range
MZ I – Northern Great Plains [a]	186,480	41,731	22%
MZ II – Wyoming Basin	149,820	48,189	32%
MZ III – Southern Great Basin	124,057	36,629	30%
MZ IV – Snake River Plain	156,360	46,700	30%
MZ V – Northern Great Basin	78,293	14,018	18%
MZ VI – Columbia Basin	11,161	4,462	40%
MZ VII – Colorado Plateau	4,777	652	14%
Range wide 1	710,948	192,381	26%

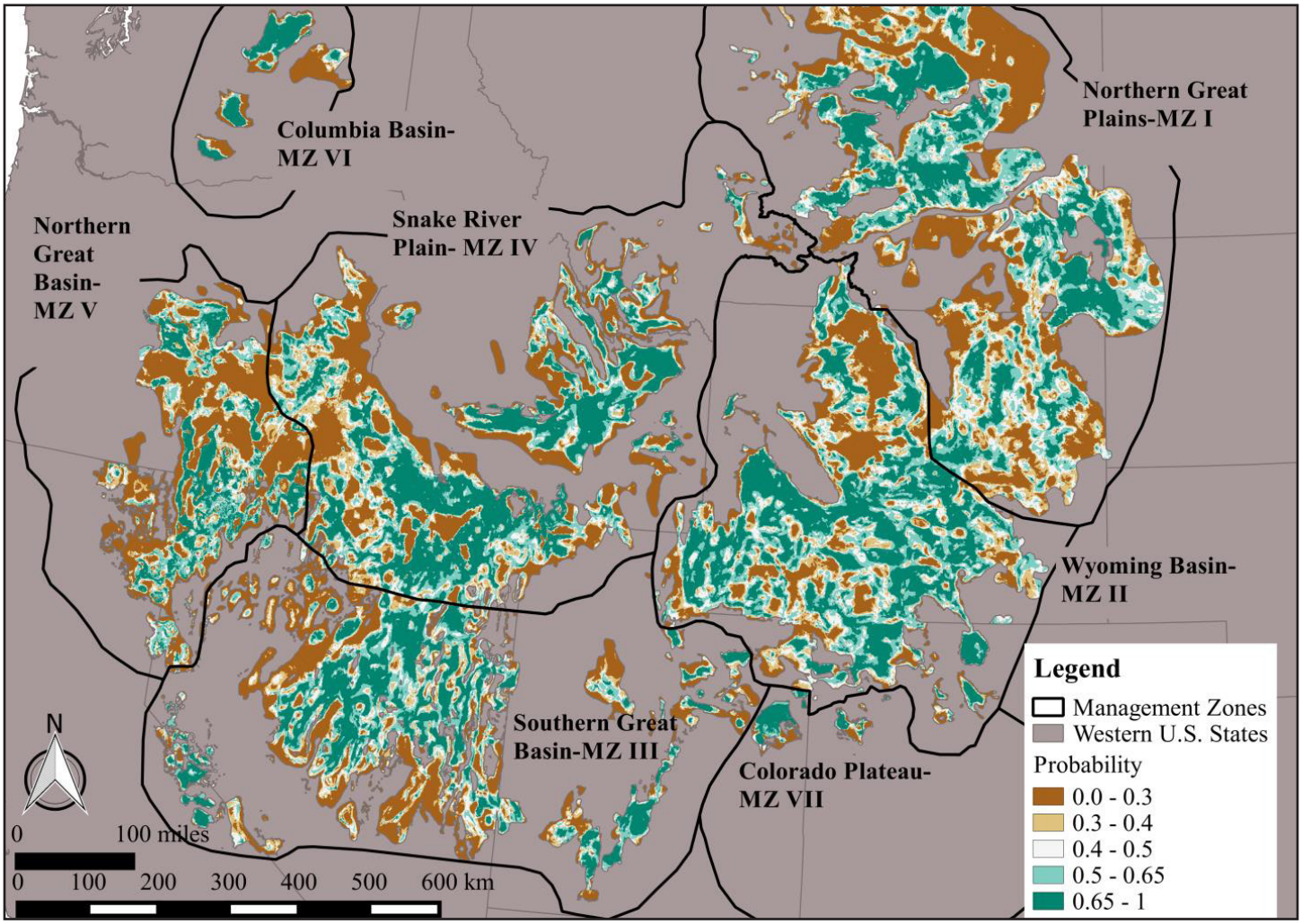
[a] Does not include the Canadian portion of the range.

Table 6. Top 5 variables and their importance values selected for each management zone from 2010 – 2014. Importance values are scaled by management zone, so that the top variable equals 1 and the remaining variables are a proportion derived by dividing by the top variable, and are derived from probability scaled partial plots in the RandomForest package in R.

Management Zone	1st Variable	2nd Variable	3rd Variable	4th Variable	5th Variable
Northern Great Plains (I)	Canopy Cover	All Sagebrush	Roughness	Topographic Wetness Index	Gross Primary Production
	1.00	0.63	0.57	0.55	0.45
Wyoming Basin (II)	All Sagebrush	Canopy Cover	Annual Drought Index	Degree Days > 5°C	Mean Annual Precipitation
	1.00	0.73	0.68	0.59	0.49
Southern Great Basin (III)	All Sagebrush	Degree Days > 5°C	Elevation	Annual Drought Index	Canopy Cover
	1.00	0.79	0.70	0.54	0.48
Snake River Plain (IV)	Canopy Cover	Annual Drought Index	All Sagebrush	Degree Days > 5°C	Gross Primary Production
	1.00	0.60	0.59	0.51	0.50
Northern Great Basin (V)	All Sagebrush	Annual Drought Index	Low Sagebrush	Mean Annual Precipitation	Degree Days > 5°C
	1.00	0.96	0.91	0.79	0.65
Columbia Basin (VI)	Elevation	Degree Days > 5°C	Grassland/ Herbaceous	Annual Drought Index	All Sagebrush
	1.00	0.42	0.41	0.27	0.22
Colorado Plateau (VII)	All Sagebrush	Low Sagebrush	Human Disturbance Index	Oil & Gas Wells	
	1.00	0.67	0.48	0.40	

importance than disturbance variables in predicting occupied breeding habitat (**Table 6, Appendix II**). Not surprisingly, the percent of a landscape dominated by sagebrush within 130.1 km² (50.24 mile²; 32,153 acres) was the top variable in 4 of the 7 models, and was in the top five variables for all models (**Table 6**). We documented variation in habitat selection for sagebrush but also show similar patterns across the range (**Figure 3**). However, functional habitat selection for sagebrush modeled for the Northern Great Plains and Columbia Basin management zones diverged from results for the rest of the management zones, because sage-grouse were modeled to occupy habitats with lower proportions of sagebrush in zones I and VI (**Figure 3**). All sage-grouse breeding habitats showed strong avoidance of tree cover; however strength of avoidance varied between management zones (**Figure 4**). The disturbance index was selected within models for all management zones except the Northern Great Basin with a variable importance range of (0.48 Colorado Plateau – 0.09 Southern Great Basin, **Table 6 and Appendix II**). While threshold values between management zones varied similar to tree canopy cover, models documented clear thresholds in amount of landscape level disturbance tolerated (**Figure 5**). Models show that Northern Great Plains management zone had the lowest threshold for the disturbance index (2.9% when p ~ 0.65, **Figure 5**). They also documented variability and differences in threshold values for the amount of tillage in the landscape, with sage-grouse in management zone I showing the least tolerance for tilled landscapes (**Figure 6**). Despite variability in disturbance and non-habitat thresholds, we found similar patterns in the peaks of probability distributions (p > 0.8) for our two strongest historic climatic predictors (Annual Drought

Figure 2. Breeding habitat model of Greater Sage-grouse developed within each of the 7 Management Zones. The breeding habitat model is a spatially explicit probability prediction that the surrounding landscape will contain enough breeding habitat to support Greater Sage-grouse lek formation. All active leks within the sage-grouse range (2010 – 2014) occurred on probabilities > 0.65.



Index [**Figure 7**] and Degree Days > 5 °C [**Figure 8**]). A current measure of climate as measured by Gross Primary Production had lower variable importance than our historic climate envelopes in model selection (**Table 6**). We documented similar patterns of selection for Gross Primary Production, although peaks varied across the range with the lowest selected range of Gross Primary Production in the Northern Great Basin and the highest in the Northern Great Plains (**Figure 9**). Graphical partial probability plots of all variables are available in alphabetical order by the variable abbreviation in **Appendix III**.

Breeding Population Index Model

We demonstrate distinct clustering in the relative abundance of sage-grouse populations within each management zone (**Figure 10, Figure 11**). On average approximately half of the breeding population is predicted to be within 10% of the occupied range. Across all management zones, all populations visually demonstrated asymptotic properties between the each additional 10% of the population and the area required to contain these populations (**Figure 11**). For example, to go from 80% of the population index to 90% increased the area required by 44% on average (range 41% MZ II – 50% MZ I; **Figure 11**).

DISCUSSION

Even for a species as specialized as sage-grouse, we showed that ecological context mattered in both the strength of habitat selection and values of disturbance thresholds. Our range-wide population and habitat models accounted for regional variation in habitat selection and the relative densities of birds. These models can serve as a consistent and common currency to assess how sage-grouse habitat (**Figure 2**) and populations (**Figure 10**) overlap

with conservation actions or threats over the entire sage-grouse range. Our models demonstrated high statistical model fit (**Table 3, Table 4**). Similar to Coates et al. [14] our models utilize the best information available at a range-wide scale. Our models were built within management zones to account for differences across the entire range.

Our work clearly shows variation in habitat selection and varying thresholds in response to disturbance across the range. Partial probability plots in our analyses are intended to highlight the how ecological gradients (**Appendix I**) across vast occupied range (**Table 5**) can change functional habitat responses, and ultimately predictions of breeding habitat. This is important, because our work implies careful consideration must be given to the validity of extrapolating results of studies in one management zone to others especially if they have vastly different ecological context. The scale of our work was intentionally range-wide, thus we addressed first-order selection of habitat. Therefore threshold values identified in this manuscript are meaningful at the first order scale, however we would strongly caution against implementing our results into management, until multi-scale (1st – 3rd order) and cumulative effects are investigated simultaneously. We also believe formal tests should be implemented to evaluate the differences of the functional response distribution such as a Kolmogorov-Smirnov test, or simulations and randomization tests. Because there are 7 management zones and because of the large number of variables selected within our breeding habitat models, simulations to formalize tests will be computationally expensive, but needed.

Where study boundaries are drawn fundamentally determines what is learned about the ecology of the species. Science and information is developing rapidly and information on first-order functional responses could be used in conjunction with other information to help refine sage-grouse management zones and disturbance thresholds in the future. Analyzing sage-grouse populations by management zones is supported in the literature and our analysis confirms differences between management zones in both habitat selection and thresholds to disturbance. It would also reason that sub-management zone level variation exists and populations close to but on different sides of current management zone boundaries could exhibit similar patterns in habitat selection. Modeling at lower levels of the habitat selection hierarchy could also increase predictive power of models and elucidate biological relationships not evident at the first-order scale. We believe future efforts could complement our analyses by re-running analyses with boundaries defined by range-wide genetic clustering of population units. Once habitat selection models are computed at these smaller extents, these population units could be grouped by niche similarity and habitat responses to refine management zones. Formal ecological and statistical criterion for establishing whether two populations are similar need to be established prior to refining sage-grouse management zones based upon genetic population units, landscape connectivity, and niche responses. units, landscape connectivity, and niche responses.

Acknowledgments

We would like to thank San Stiver (WAFWA) for coordinating the collection of the Greater Sage-grouse lek data and Tom Remington (WAFWA) for providing us clean proofed lek databases which formed the basis of our model. We would also like to thank members of the Sage-grouse technical team for reviewing models and more importantly ensuring spatial predictions were consistent with landscapes they or their staff know about and actively manage. We would like to thank M. Hooten (USGS) and D. Johnson (USGS) for meaningful review of our techniques and suggested ideas which were incorporated in the development of the final breeding population index model. Tom Kimball of the USGS graciously facilitated the USGS peer review process. We would also like to thank Julie Yee (USGS statistician), Erik Blomberg (University of Maine), and David Dalgren (Utah State University) for very detailed and thorough reviews which greatly increased the clarity of this work. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

This product has been peer reviewed and approved for publication consistent with USGS Fundamental Science Practices. Use of trade, product, or firm names does not imply endorsement by the U.S. Government.

Figure 3. Functional habitat response between the percent of ***all sagebrush cover types*** (x-axis) within a 6.4-km buffer (130.1 km2) and the probability (y-axis) a landscape will contain enough breeding habitat to support Greater Sage-grouse lek formation within each Management Zone (2010-2014). Functional response curves were generated using partial probability plots to explore the influence of a given variable on the probability of occurrence, while partialing out the average effects of all other variables in the final model.

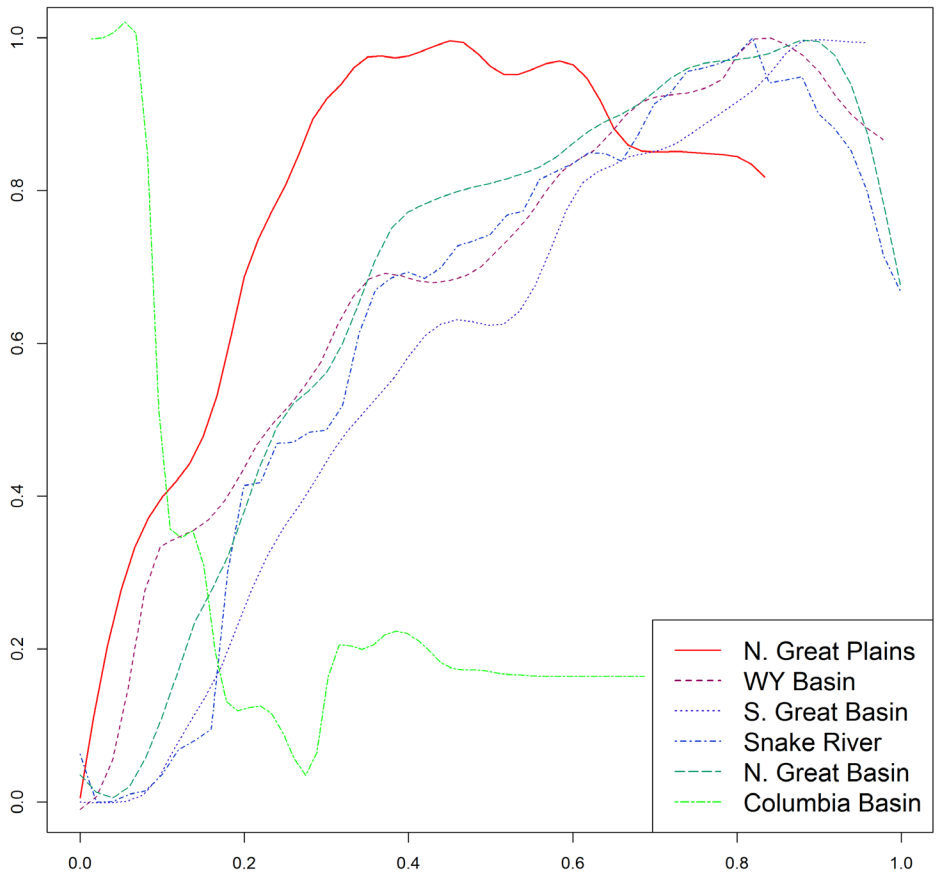


Figure 4. Functional habitat response between ***tree canopy cover*** (x-axis) within a 6.4-km buffer (130.1 km^2) and the probability (y-axis) a landscape will contain enough breeding habitat to support Greater Sage-grouse lek formation within each Management Zone (2010-2014). Functional response curves were generated using partial probability plots to explore the influence of a given variable on the probability of occurrence, while partialing out the average effects of all other variables in the final model.

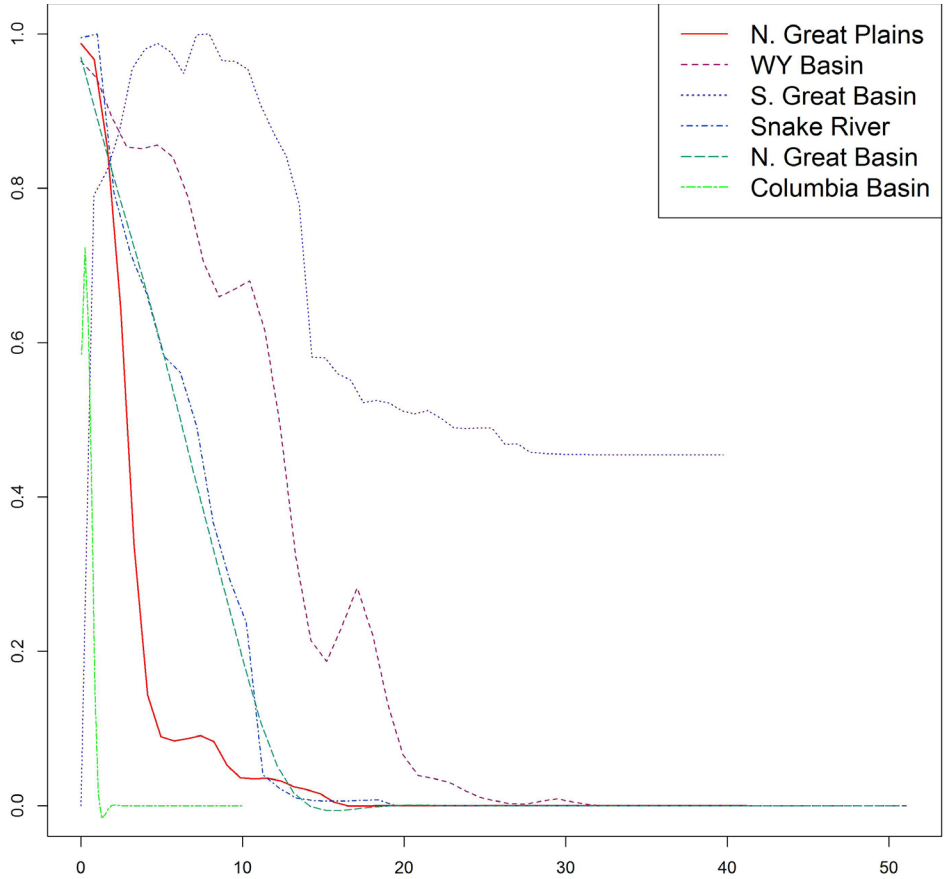


Figure 5. Functional habitat response between the ***amount of human disturbance index*** (x-axis) within a 6.4-km buffer (130.1 km²) and the probability (y-axis) a landscape will contain enough breeding habitat to support Greater Sage-grouse lek formation within each Management Zone (2010-2014). Functional response curves were generated using partial probability plots to explore the influence of a given variable on the probability of occurrence, while partialing out the average effects of all other variables in the final model.

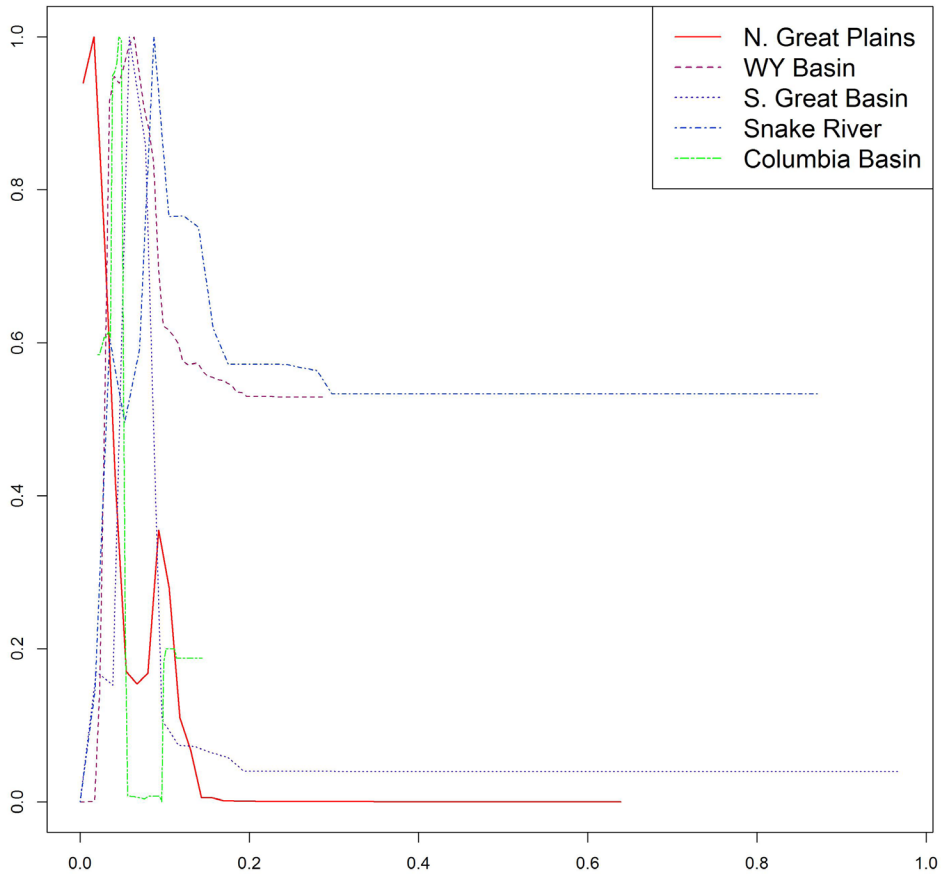


Figure 6. Functional habitat response between the ***amount of tilled cropland*** (x-axis)within a 6.4-km buffer (130.1 km²) and the probability (y-axis) a landscape will contain enough breeding habitat to support Greater Sage-grouse lek formation within each Management Zone (2010-2014). Functional response curves were generated using partial probability plots to explore the influence of a given variable on the probability of occurrence, while partialing out the average effects of all other variables in the final model.

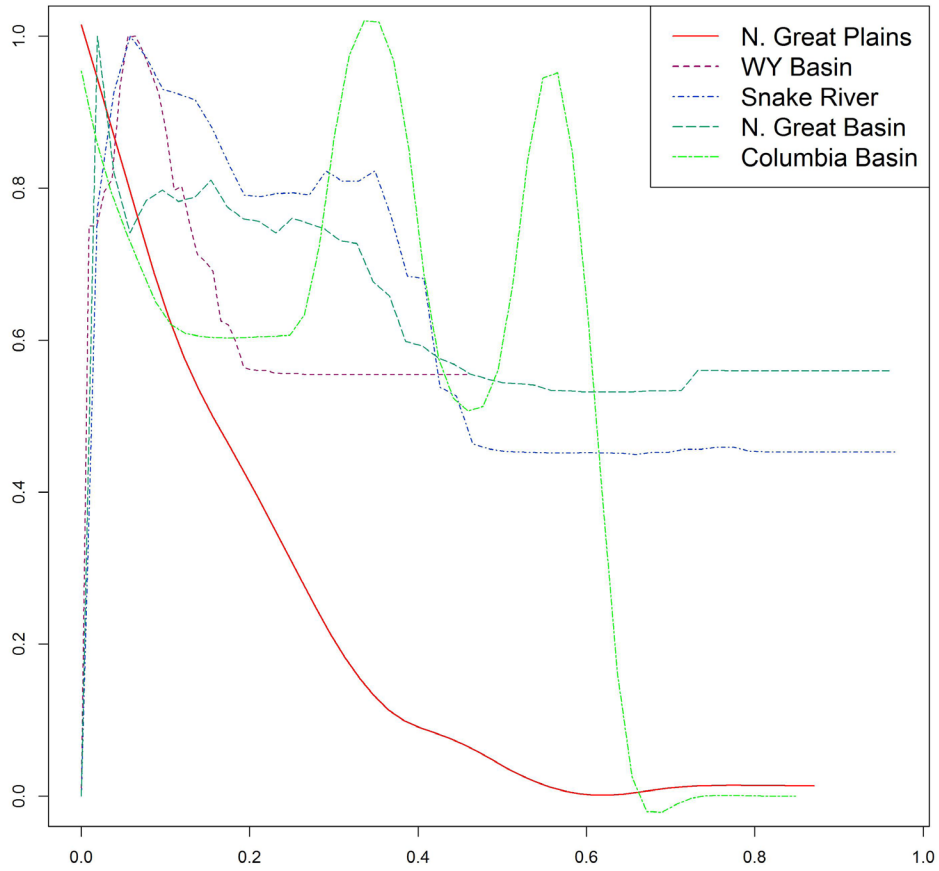


Figure 7. Functional habitat response between the ***average annual drought index*** (x-axis)within a 6.4-km buffer (130.1 km²) and the probability (y-axis) a landscape will contain enough breeding habitat to support Greater Sage-grouse lek formation within each Management Zone (2010-2014). Functional response curves were generated using partial probability plots to explore the influence of a given variable on the probability of occurrence, while partialing out the average effects of all other variables in the final model.

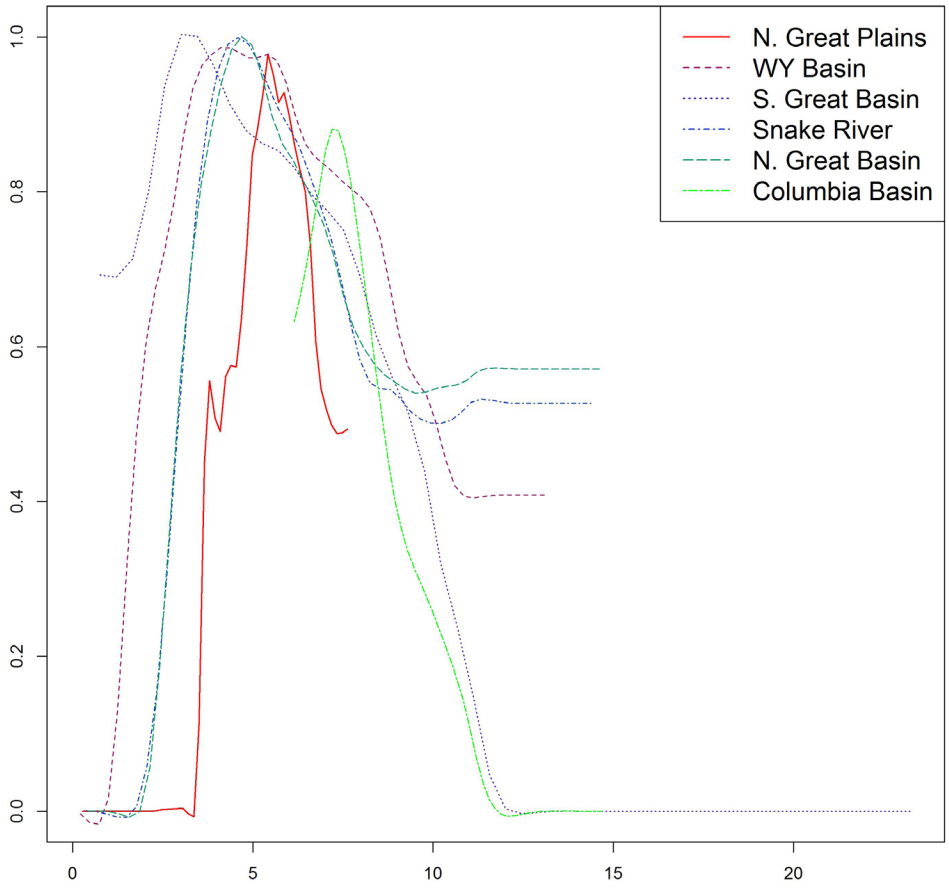


Figure 8. Functional habitat response between the ***average degree day 5°C*** (x-axis)within a 6.4-km buffer (130.1 km²) and the probability (y-axis) a landscape will contain enough breeding habitat to support Greater Sage-grouse lek formation within each Management Zone (2010-2014). Functional response curves were generated using partial probability plots to explore the influence of a given variable on the probability of occurrence, while partialing out the average effects of all other variables in the final model.

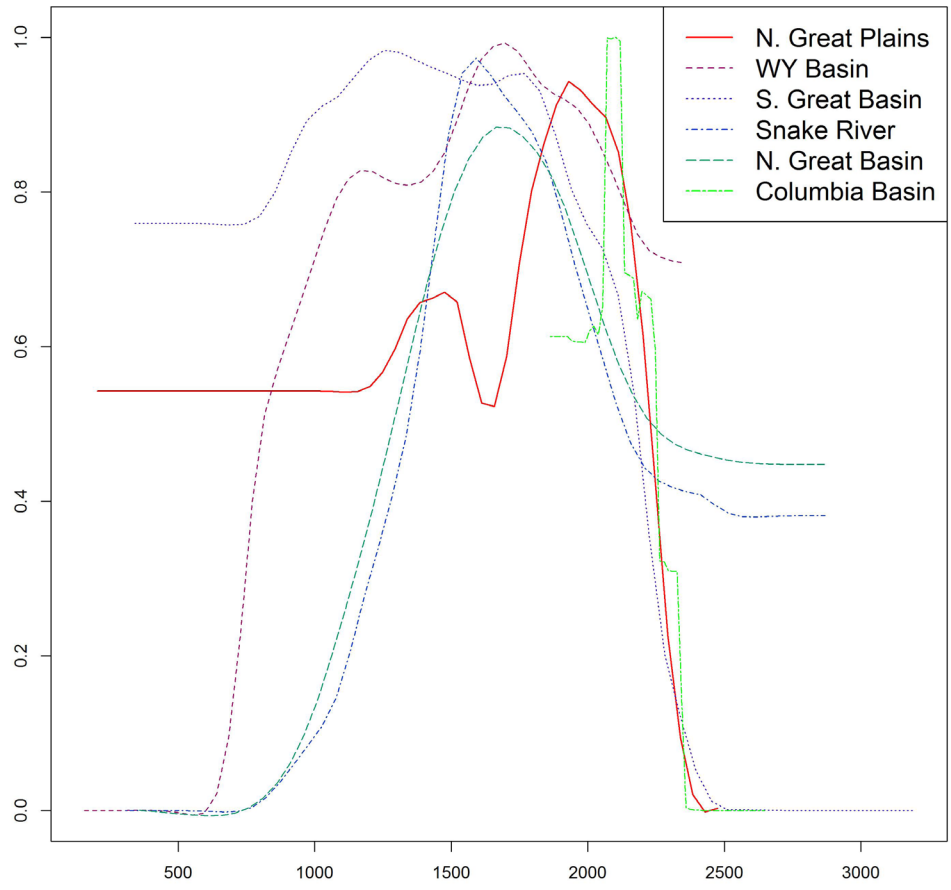
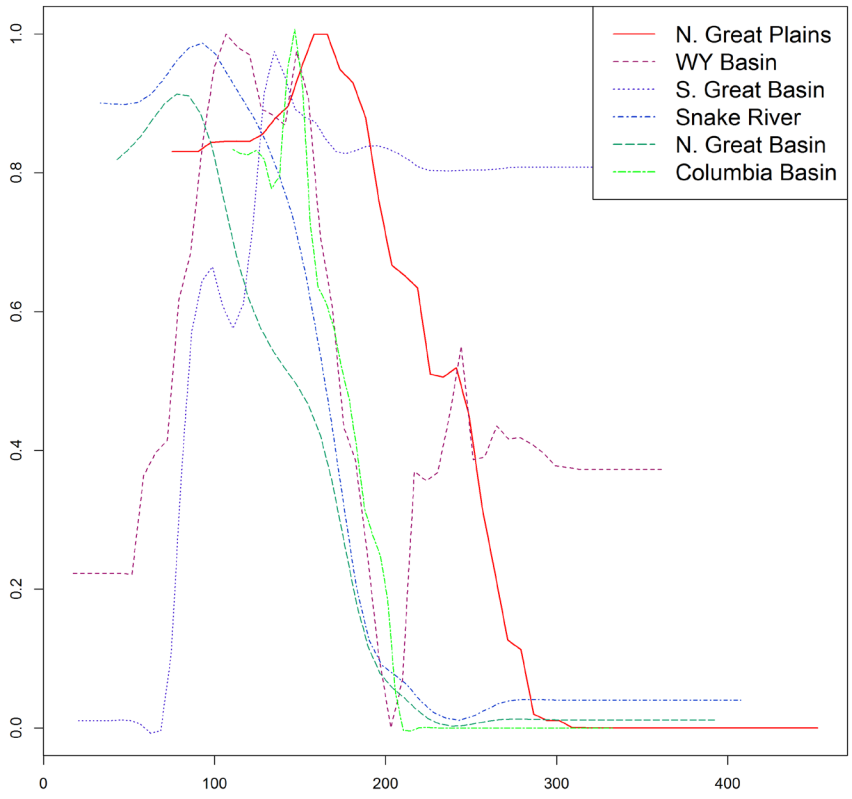


Figure 9. Functional habitat response between the ***gross primary production*** (x-axis)within a 6.4-km buffer (130.1 km²) and the probability (y-axis) a landscape will contain enough breeding habitat to support Greater Sage-grouse lek formation within each Management Zone (2010-2014). Functional response curves were generated using partial probability plots to explore the influence of a given variable on the probability of occurrence, while partialing out the average effects of all other variables in the final model.



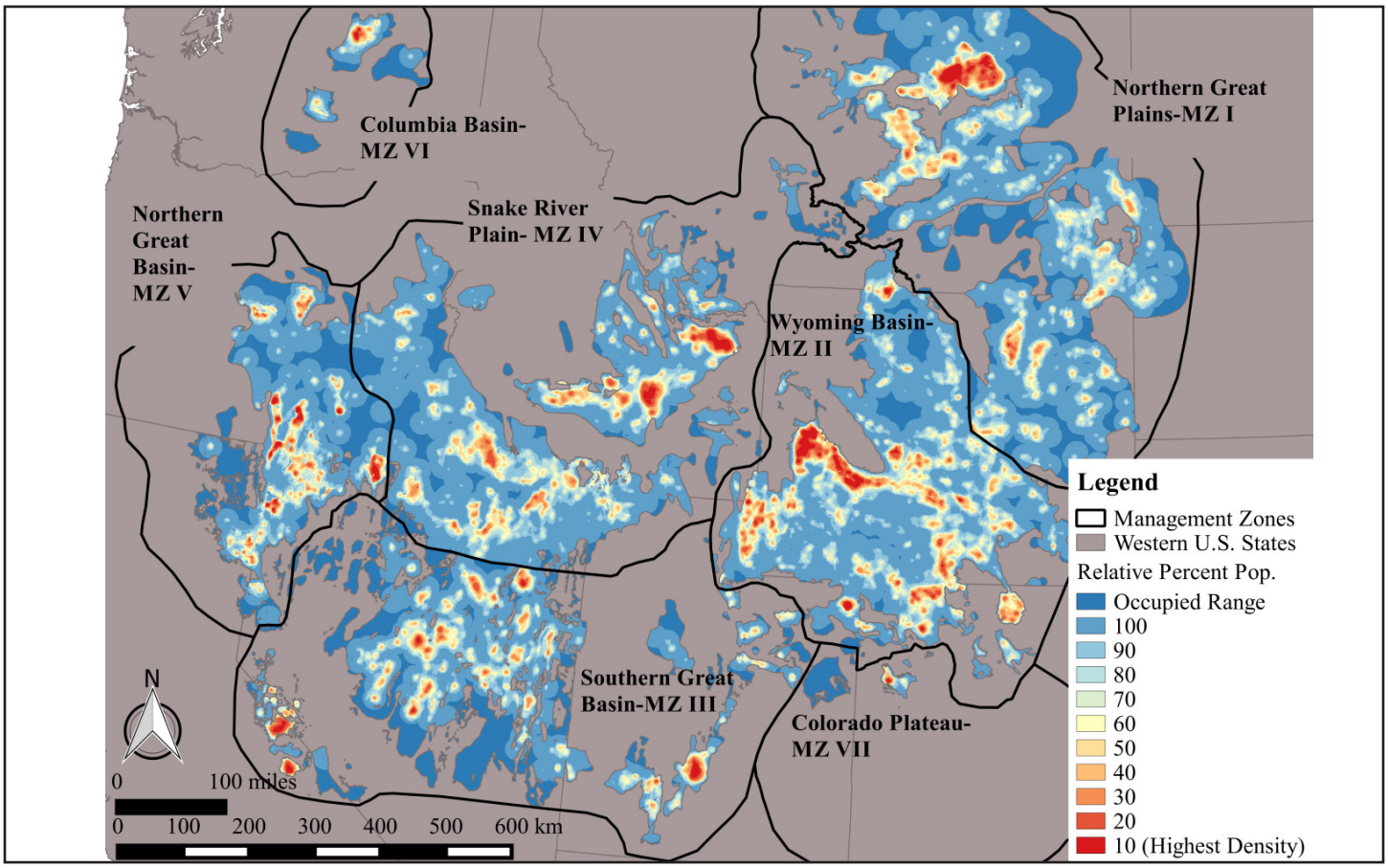


Figure 10. Breeding population index model of Greater Sage-grouse within each of the 7 Management Zones. Our population index model provides spatial insight into the relative importance of specific areas to the overall management zone-wide breeding abundance of Greater Sage-grouse during 2010-1014. Population index values are relative within each management zone. Sage-grouse population index areas represent spatial locations of the known breeding population in 10% bins differentiated by color. The darkest red areas contain 10% of the breeding population. Because bins are additive, red and orange hue areas combined capture 50% of the population, etc.

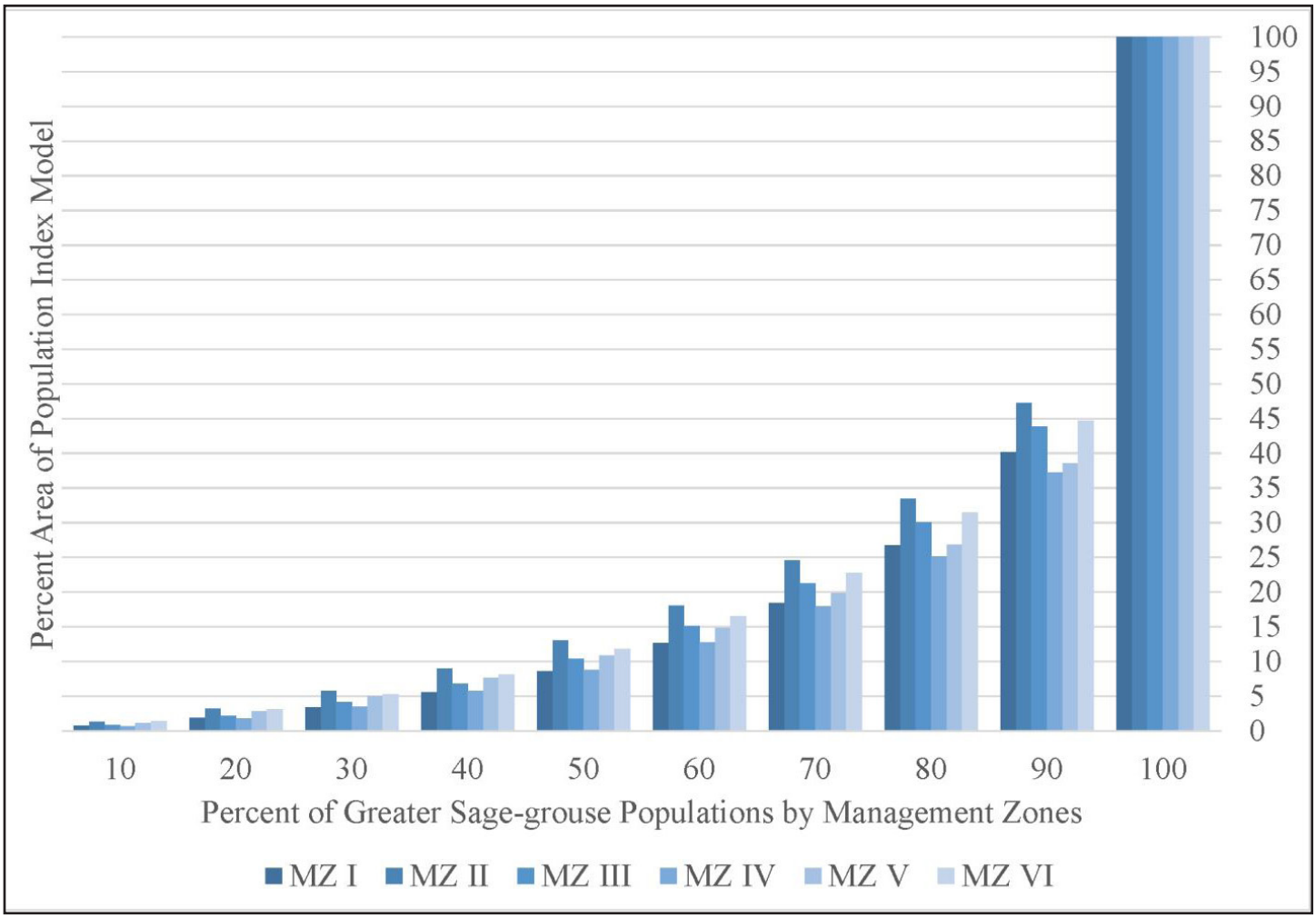


Figure 11. The percent of the population index model and resulting percent area of the entire population index model by management zones during 2010-2014

LITERATURE CITED:

1. Boyce MS, McDonald LL (1999) Relating populations to habitats using resource selection functions. Trends in Ecology & Evolution 14: 268-272.

2. Johnson CJ, Nielsen SE, Merrill EH, McDonald TL, Boyce MS (2006) Resource Selection Functions Based on Use–Availability Data: Theoretical Motivation and Evaluation Methods. Journal of Wildlife Management 70: 347-357.

3. Johnson DS, Hooten MB, Kuhn CE (2013) Estimating animal resource selection from telemetry data using point process models. Journal of Animal Ecology 82: 1155-1164.

4. Randin CF, Dirnböck T, Dullinger S, Zimmermann NE, Zappa M, et al. (2006) Are niche-based species distribution models transferable in space? Journal of Biogeography 33: 1689-1703.

5. Fedy BC, Doherty KE, Aldridge CL, O'Donnell M, Beck JL, et al. (2014) Habitat prioritization across large landscapes, multiple seasons, and novel areas: An example using greater sage- grouse in Wyoming. Wildlife Monographs 190: 1-39.

6. Mysterud A, Ims RA (1998) Functional responses in habitat use: Availability influences relative use in trade-off situations. Ecology 79: 1435-1441.

7. Rhodes JR, Callaghan JG, McAlpine CA, Jong Cd, Bowen ME, et al. (2008) Regional Variation in Habitat-Occupancy Thresholds: A Warning for Conservation Planning. Journal of Applied Ecology 45: 549-557.

8. Carwardine J, Wilson KA, Watts M, Etter A, Klein CJ, et al. (2008) Avoiding Costly Conservation Mistakes: The Importance of Defining Actions and Costs in Spatial Priority Setting. PLoS ONE 3: e2586.

9. van der Hoek Y, Wilson AM, Renfrew R, Walsh J, Rodewald PG, et al. (2015) Regional variability in extinction thresholds for forest birds in the north-eastern United States: an examination of potential drivers using long-term breeding bird atlas datasets. Diversity and Distributions 21: 686-697.

10. Sagarin RD, Gaines SD, Gaylord B (2006) Moving beyond assumptions to understand abundance distributions across the ranges of species. Trends in Ecology & Evolution 21: 524-530.

11. Murphy HT, VanDerWal J, Lovett- Doust J (2006) Distribution of abundance across the range in eastern North American trees. Global Ecology and Biogeography 15: 63-71.

12. Brown JH, Mehlman DW, Stevens GC (1995) Spatial Variation in Abundance. Ecology 76: 2028-2043.

13. Stephens PA, Pettorelli N, Barlow J, Whittingham MJ, Cadotte MW (2015) Management by proxy? The use of indices in applied ecology. Journal of Applied Ecology 52: 1-6.

14. Coates PS, Casazza ML, Brussee BE, Ricca MA, Gustafson KB, et al. (2014) Spatially explicit modeling of greater sage-grouse (Centrocercus urophasianus) habitat in Nevada and northeastern California: a decision-support tool for management. US Geological Survey. 2331-1258 2331-1258.

15. Schroeder MA, Aldridge CL, Apa AD, Bohne JR, Braun CE, et al. (2004) Distribution of sage-grouse in North America. The Condor 106: 363-376.

16. Gibson RM, Langen TA (1996) How do animals choose their mates? Trends in Ecology & Evolution 11: 468-470.

17. Schroeder MA, White GC (1993) Dispersion of greater prairie chicken nests in relation to lek location: evaluation of the hot-spot hypothesis of lek evolution. Behavioral Ecology 4: 266-270.

18. Coates PS, Casazza ML, Blomberg EJ, Gardner SC, Espinosa SP, et al. (2013) Evaluating greater sage-grouse seasonal space use relative to leks: Implications for surface use designations in sagebrush ecosystems. The Journal of Wildlife Management 77: 1598-1609.

19. Holloran MJ, Anderson SH (2005) Spatial distribution of greater sage-grouse nests in relatively contiguous sagebrush habitats. The Condor 107: 742-752.

20. Doherty KE, Naugle DE, Walker BL (2010) Greater sage-grouse nesting habitat: the importance of managing at multiple scales. Journal of Wildlife Management 74: 1544-1553.

21. Doherty KE, Naugle DE, Copeland HE, Pocewicz A, Kiesecker JM (2011) Energy development and conservation tradeoffs: systematic planning for Greater Sage-Grouse in their eastern range. In: Knick ST, Connelly JW, editors. Greater Sage-Grouse: ecology and conservation of a landscape species and its habitats Studies in Avian Biology. Berkeley, CA.: University of California Press,. pp. 505-516.

22. Gibson RM (1996) A re-evaluation of hotspot settlement in lekking sage grouse. Animal Behaviour 52: 993-1005.

23. Knick ST, Dobkin DS, Rotenberry JT, Schroeder MA, Vander Haegen WM, et al. (2003) Teetering on the edge or too late? Conservation and research issues for avifauna of sagebrush habitats. The Condor 105: 611-634.

24. Stiver S, Apa A, Bohne J, Bunnell S, Deibert P, et al. (2006) Greater sage-grouse comprehensive conservation strategy. Western Association of Fish and Wildlife Agencies, Cheyenne, WY.

25. Murphy HT, Lovett-Doust J (2007) Accounting for regional niche variation in habitat suitability models. Oikos 116: 99-110.

26. Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, et al. (2007) Random Forests for Classification in Ecology. Ecology 88: 2783-2792.

27. Evans JS, Murphy MA, Holden ZA, Cushman SA (2011) Modeling species distribution and change using Random Forests in Predictive species and habitat modeling. In: Drew CA, Wiersman YF, Huettmann F, editors. Landscape ecology: concepts and applications Springer, NY.

28. Olden JD, Lawler JJ, Poff NL (2008) Machine Learning Without Tears: A Primer for Ecologist. The Quarterly Review of Biology 83: 171-193.

29. Baruch-Mordo S, Evans JS, Severson JP, Naugle DE, Maestas JD, et al. (2013) Saving sage-grouse from the trees: A proactive solution to reducing a key threat to a candidate species. Biological Conservation 167: 233-241.

30. Gibson RM (1996) Female choice in sage grouse: the roles of attraction and active comparison. Behavioral Ecology and Sociobiology 39: 55-59.

31. Barnett JK, Crawford JA (1994) Pre-Laying Nutrition of Sage Grouse Hens in Oregon. Journal of Range Management 47: 114-118.

32. Connelly JW, Schroeder MA, Sands AR, Braun CE (2000) Guidelines to manage sage grouse populations and their habitats. Wildlife Society Bulletin: 967-985.

33. Doherty KE, Naugle DE, Copeland HE, Pocewicz A, Kiesecker JM (2011) Energy development and conservation tradeoffs: systematic planning for Greater Sage-grouse in their eastern range. In: Knick ST, Connelly JW, editors. Greater Sage-grouse: ecology and conservation of a landscape species and its habitats. Berkeley, CA, USA.: Studies in Avian Biology Vol. 38, University of California Press. pp. 505 - 516.

34. Coates PS, Casazza ML, Blomberg EJ, Gardner SC, Espinosa SP, et al. (2013) Evaluating greater sage-grouse seasonal space use relative to leks: Implications for surface use designations in sagebrush ecosystems. The Journal of Wildlife Management 77: 1598-1609.

35. Meyer CB, Thuiller W (2006) Accuracy of resource selection functions across spatial scales. Diversity and Distributions 12: 288-297.

36. Johnson DH (1980) The Comparison of Usage and Availability Measurements for Evaluating Resource Preference. Ecology 61: 65-71.

37. DeCesare NJ, Hebblewhite M, Schmiegelow F, Hervieux D, McDermid GJ, et al. (2012) Transcending scale dependence in identifying habitat with resource selection functions. Ecological Applications 22: 1068-1083.

38. Coates PS, Halstead BJ, Blomberg EJ, Brussee B, Howe KB, et al. (2014) A hierarchical integrated population model for greater sage-grouse (Centrocercus urophasianus) in the Bi-State Distinct Population Segment, California and Nevada. US Geological Survey. 2331-1258 2331-1258.

39. Donovan TM, Jones PW, Annand EM, Thompson FR (1997) Variation in local-scale edge effects: Mechanisms and landscape context. Ecology 78: 2064-2075.

40. Diggle P (1985) A kernel method for smoothing point process data. Applied statistics: 138-147.

41. Evans JS (2015) spatialEco: An R package for spatial analysis and modeling. R package version 01-1 http://cranr-projectorg/package=spatialEco.

42. Evans JS, Cushman SA (2009) Gradient modeling of conifer species using random forests. Landscape Ecology 24: 673-683.

43. Fedy BC, Aldridge CL, Doherty KE, O'Donnell M, Beck JL, et al. (2012) Interseasonal movements of greater sage-grouse, migratory behavior, and an assessment of the core regions concept in Wyoming. The Journal of Wildlife Management 76: 1062-1071.

44. Northrup JM, Hooten MB, Anderson CR, Wittemyer G (2013) Practical guidance on characterizing availability in resource selection functions under a use–availability design. Ecology 94: 1456-1463.

45. Murphy MA, Evans JS, Storfer A (2010) Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics. Ecology 91: 252-261.

46. Hastie T, Tibshirani R, Friedman J (2008) The Elements of Statistical Learning: Second Edition. Springer-Verlag.

47. Cressie N (1991) Statistics for spatial data. New York: Wiley-Interscience.

48. Team RC (2012) R: A language and environment for statistical computing: R. Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

49. Bivand RS, Keitt T, Rowlingson B (2013) rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 08-9 http://CRANR-projectorg/package=rgdal

50. Bivand RS, Pebesma EJ, Gomez-Rubio V (2008) Applied spatial data analysis with R: Springer, NY. http://www.asdar-book.org/.

51. Hijmans RJ, Etten Jv (2013) raster: Geographic data analysis and modeling: R package version 2.1-25. http://CRAN.R-project.org/package=raster

52. Breiman L (2001) Random Forests. Machine Learning 45: 5-32.

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

54. Evans JS, Murphy MA (2014) rfUtilities. R package version 10-0, http://CRANR-projectorg/package=rfUtilities.

55. Shmueli G (2010) To explain or to predict? Statistical science: 289-310.

56. Groves CR (2003) Drafting a Conservation Blueprint: A Practitioner's Guide to Planning for Biodiversity: Island Press, Washington DC, USA.

57. Sanderson EW, Redford KH, Vedder A, Coppolillo PB, Ward SE (2002) A conceptual model for conservation planning based on landscape species requirements. Landscape and urban planning 58: 41-56.

58. Connelly JW, Knick ST, Schroeder MA, Stiver SJ. Conservation assessment of greater sage-grouse and sagebrush habitats; 2004. Western Association of Fish and Wildlife Agencies Cheyenne, WY.

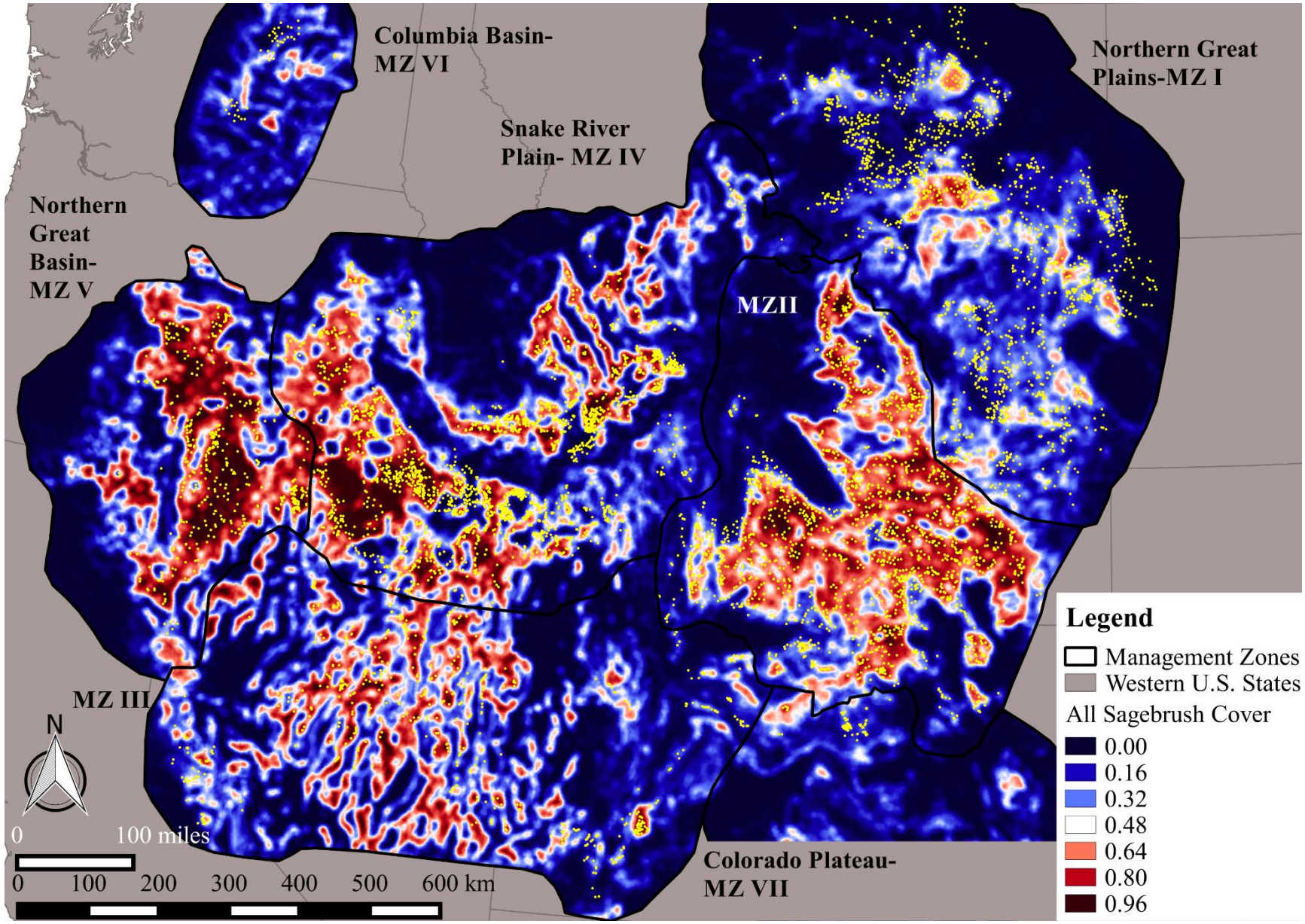
59. Silverman BW (1986) Density estimation for statistics and data analysis: CRC press.

60. Worton BJ (1989) Kernel methods for estimating the utilization distribution in home-range studies. Ecology 70: 164-168.

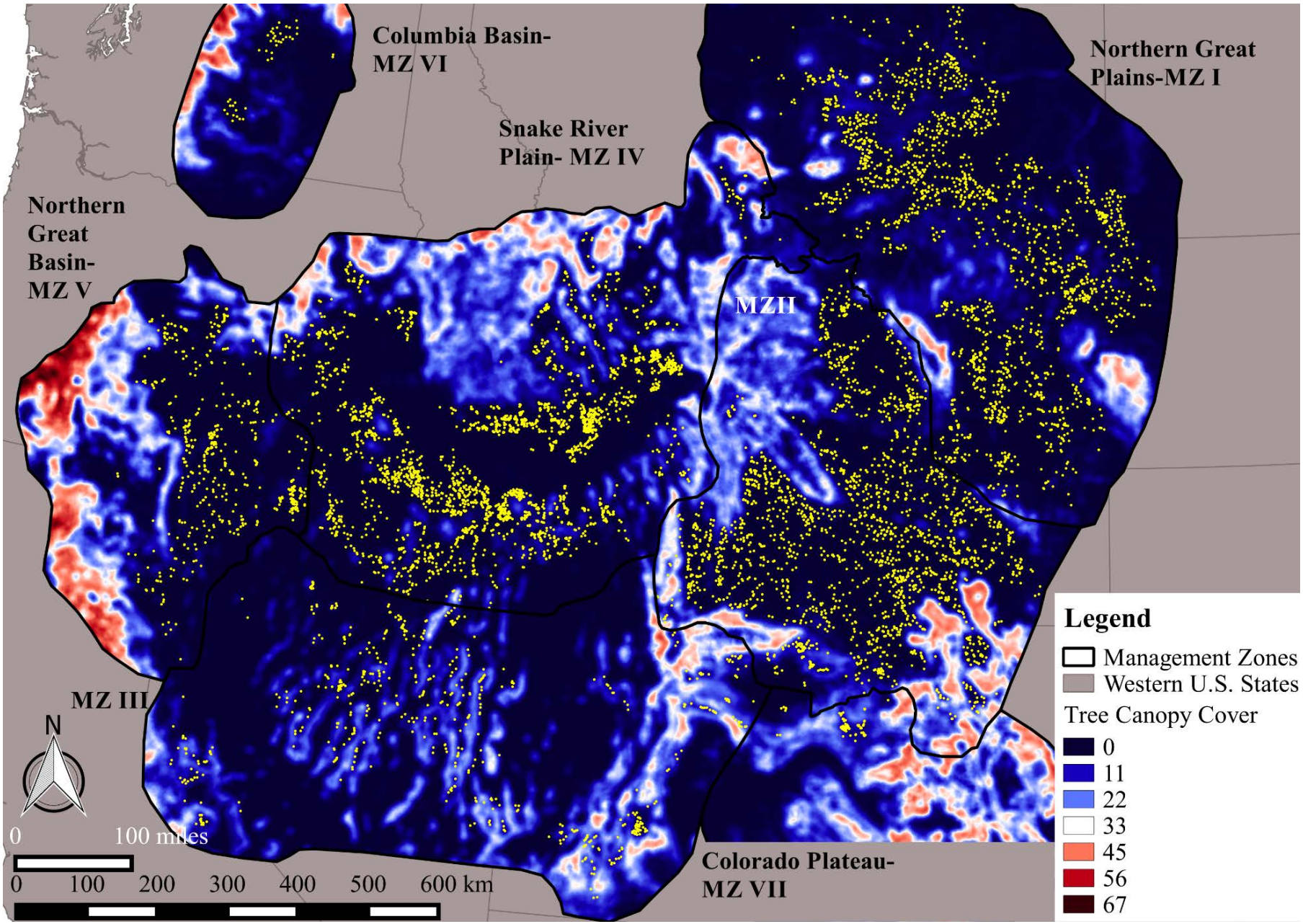
61. SAGA-GIS (2015) version 2.1.0 . http://www.saga-gis.org/.

APPENDIX I. Visual display of the ecological gradients derived from partial probability plots from Figures 3-9. Habitat variables were grouped in 7 equal interval bins for display purposes.

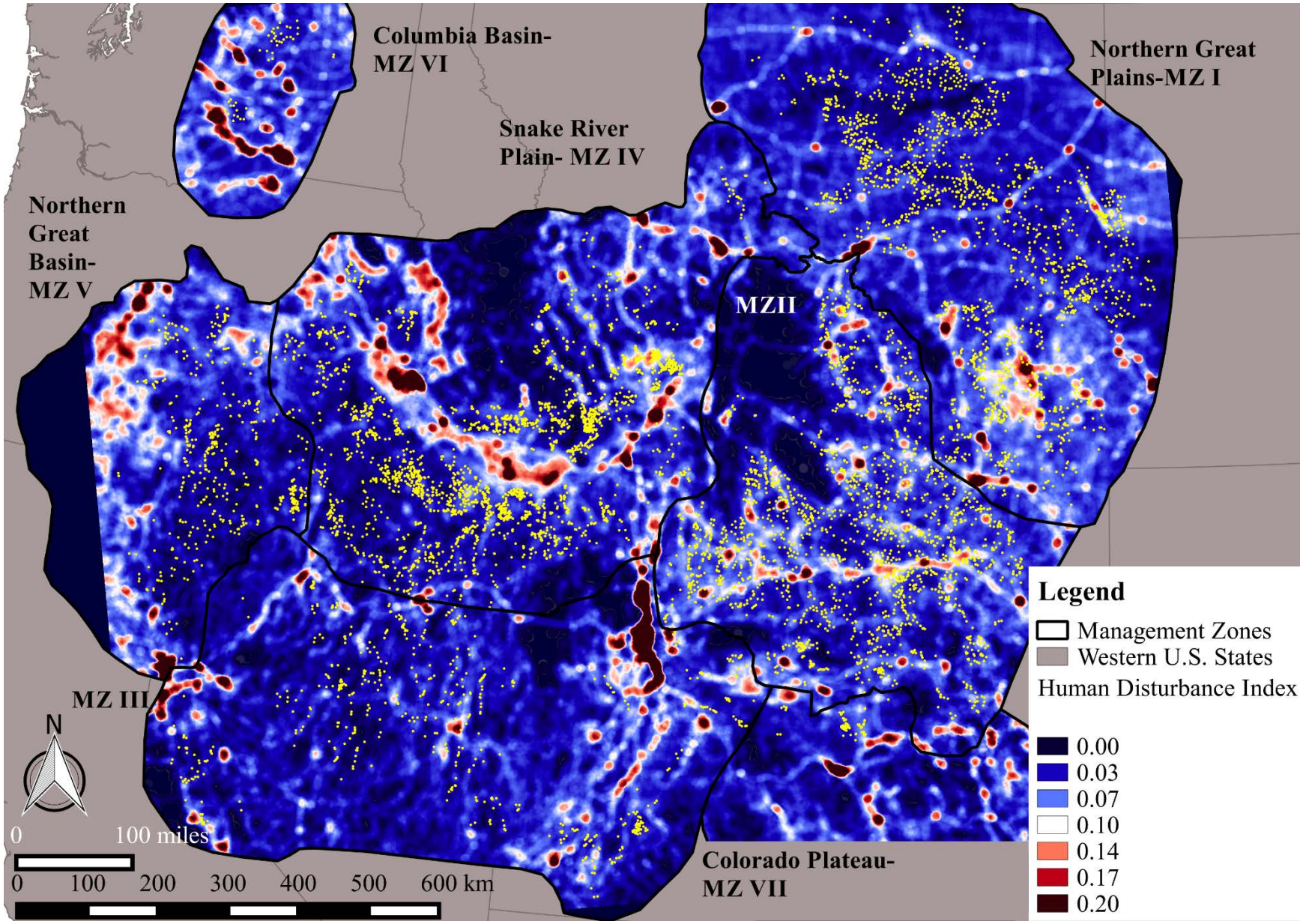
Appendix I-All Sagebrush Cover. Ecological variation in the percent of ***all sagebrush cover types***, across the seven greater sage-grouse management zones. Small yellow dots are locations of active lek locations during 2010-2014.



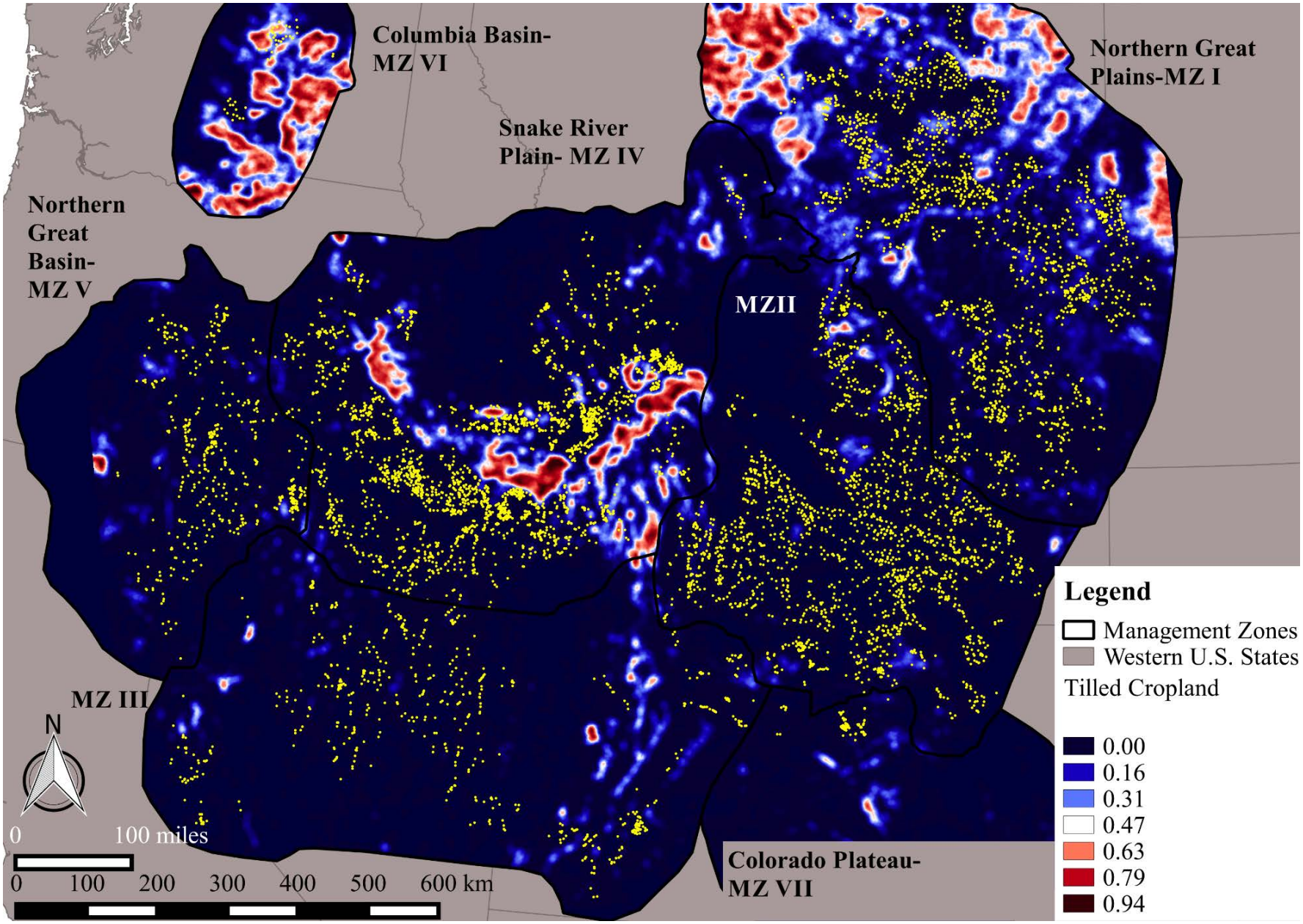
Appendix I-Tree Canopy cover. Ecological variation in ***tree canopy cover***, across the seven greater sage-grouse management zones. Small yellow dots are locations of active lek locations during 2010-2014.



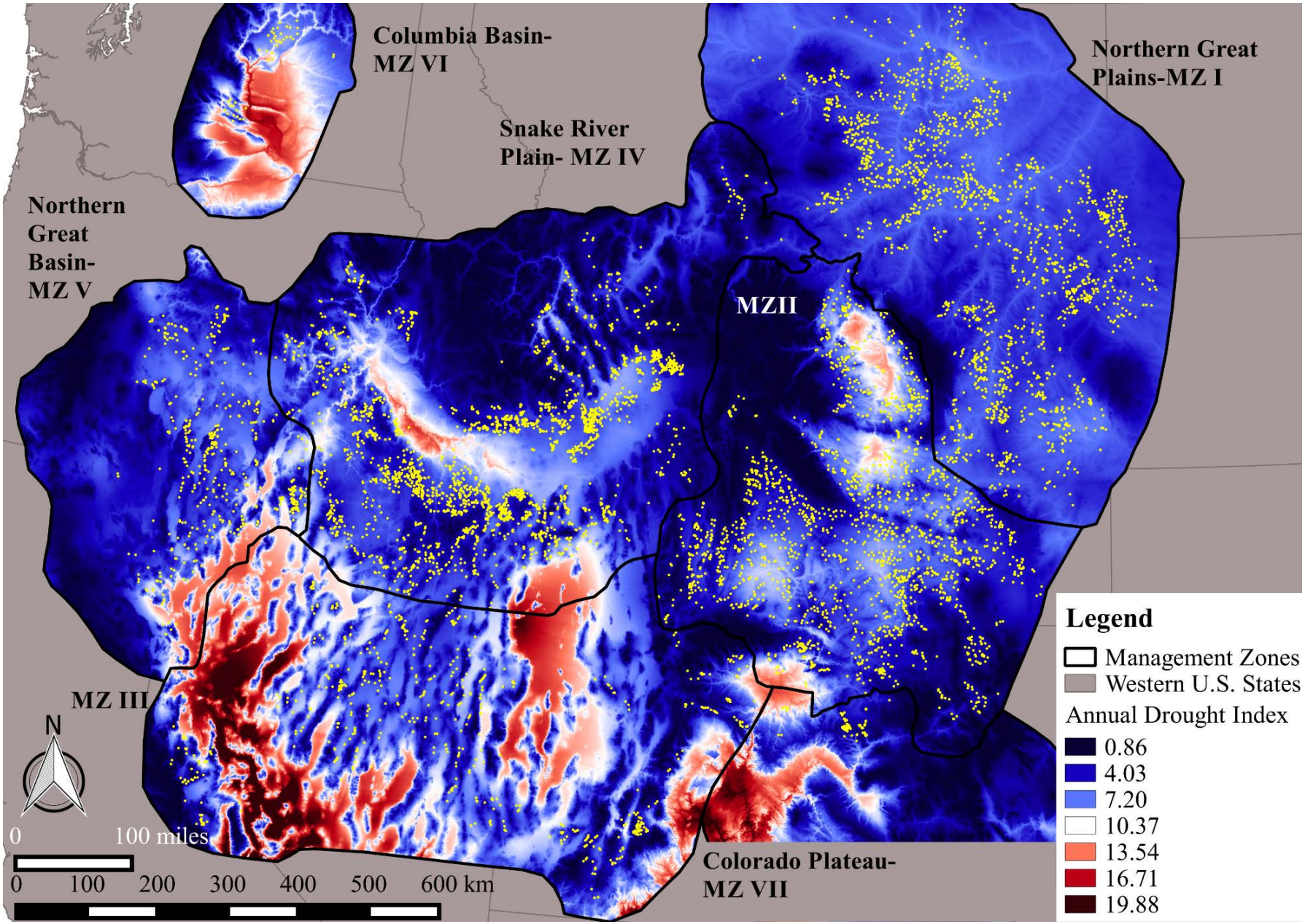
Appendix I-Human Disturbance Index. Ecological variation in the ***amount of human disturbance index***, across the seven greater sage-grouse management zones. Small yellow dots are locations of active lek locations during 2010-2014. Note sharp line of dark blue in western MZ V are a result of NoData.



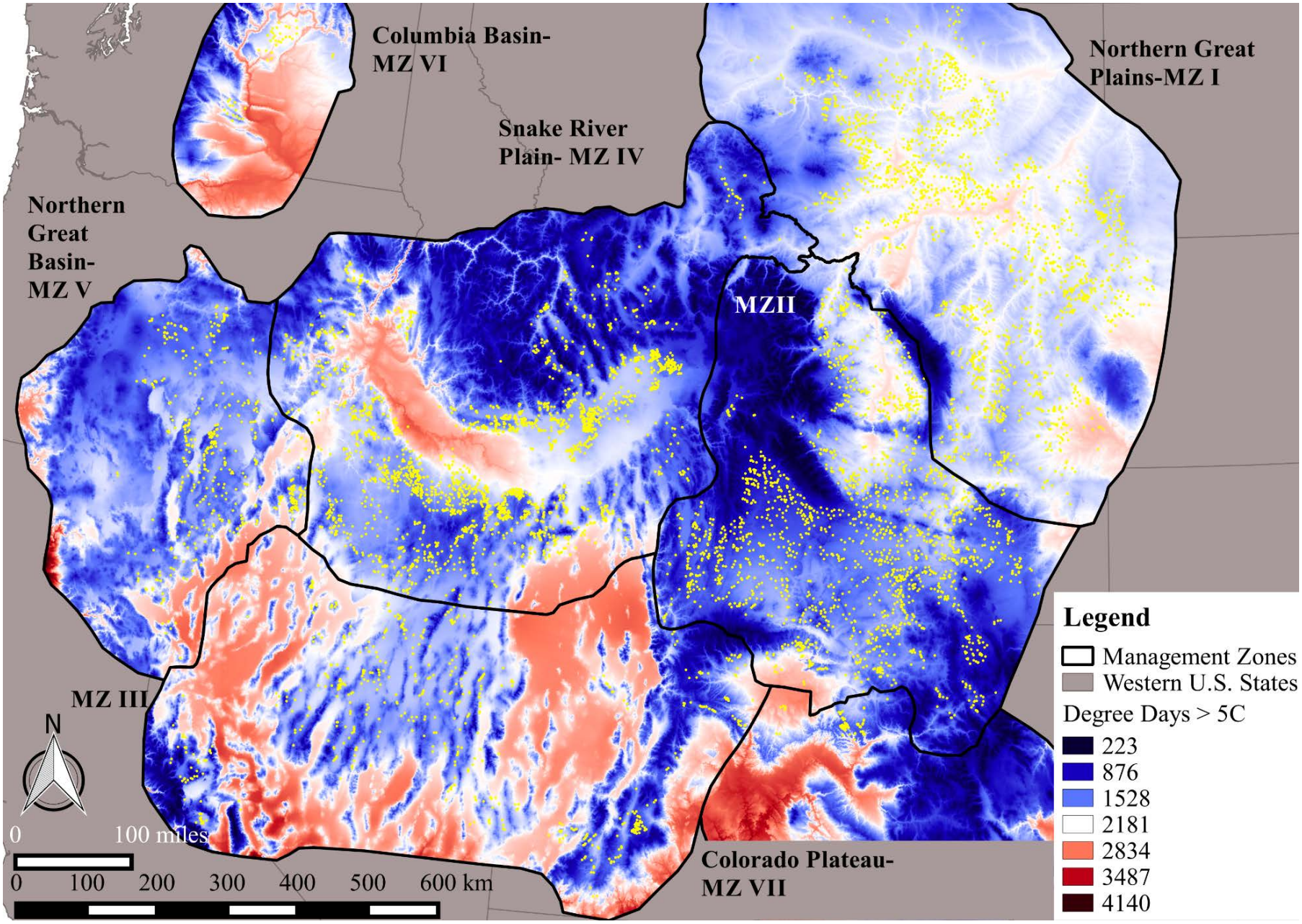
Appendix I-Tillage. Ecological variation in the ***amount of tilled cropland***, across the seven greater sage-grouse management zones. Small yellow dots are locations of active lek locations during 2010-2014. Note sharp line of dark blue in western MZ V are a result of NoData.



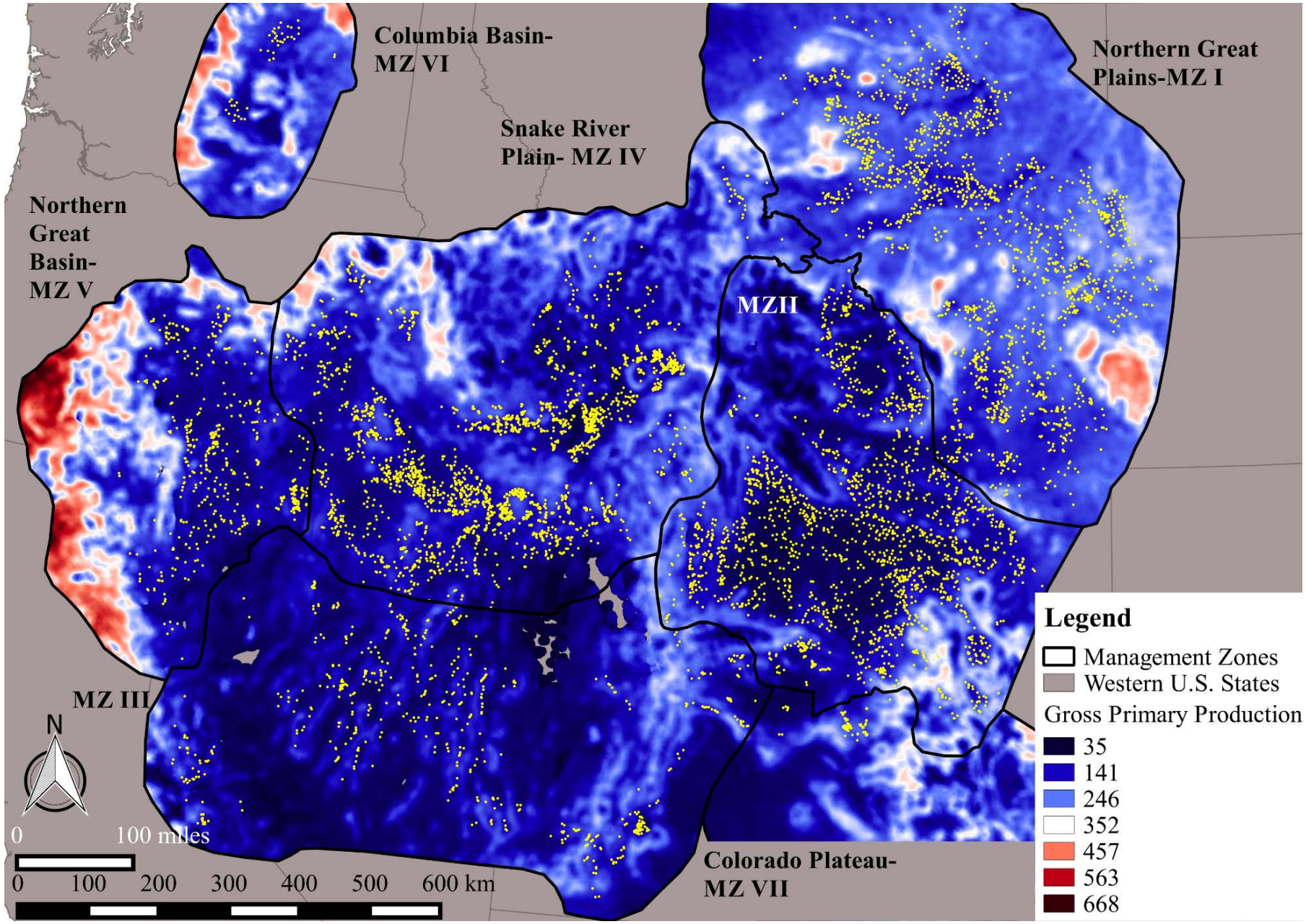
Appendix I-Annual Drought Index. Ecological variation in the ***average annual drought index***, across the seven greater sage-grouse management zones. Small yellow dots are locations of active lek locations during 2010-2014.



Appendix I-Degree Days > 5 °c. Ecological variation in the ***average degree day 5°C***, across the seven greater sage-grouse management zones. Small yellow dots are locations of active lek locations during 2010-2014.

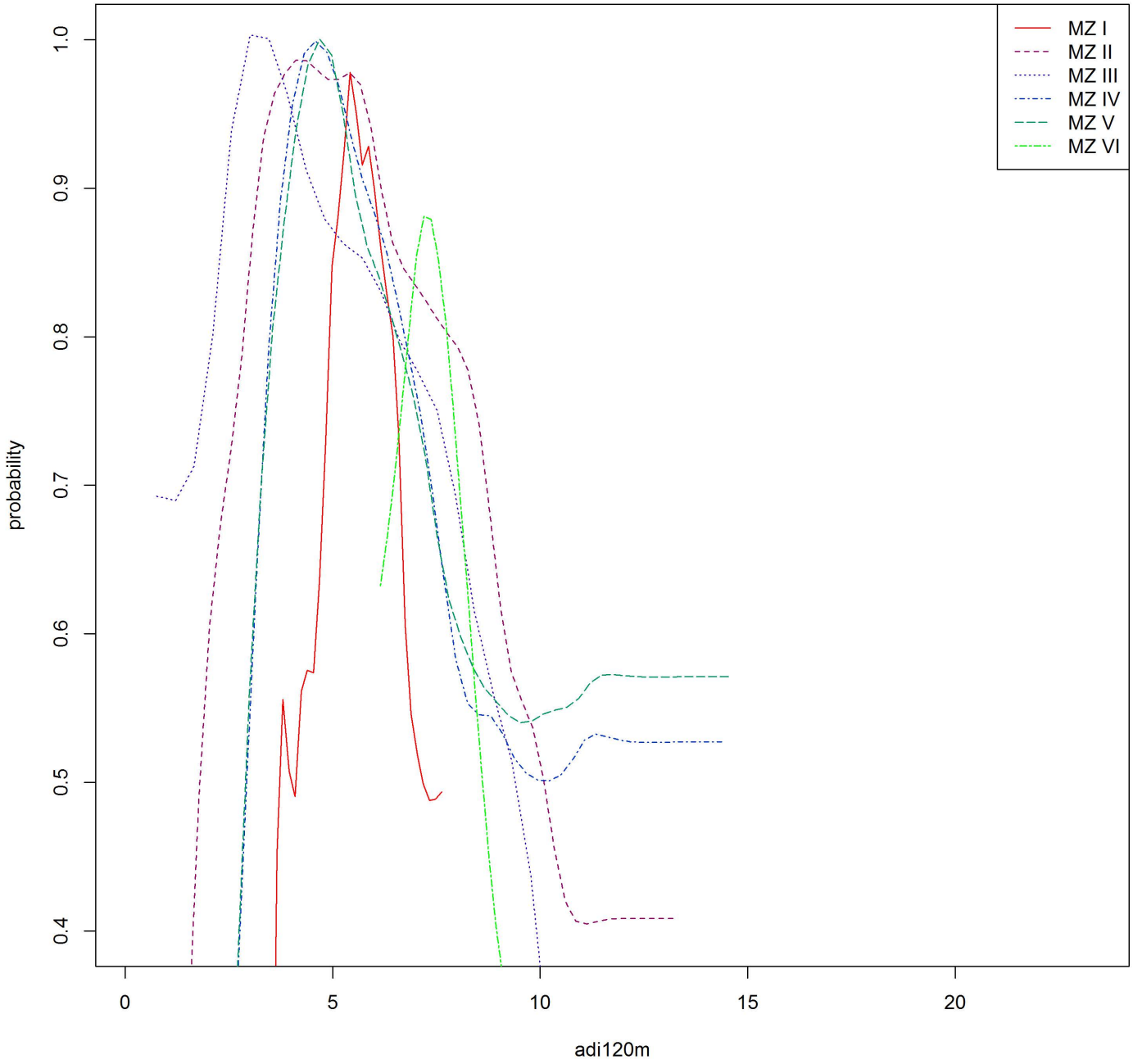


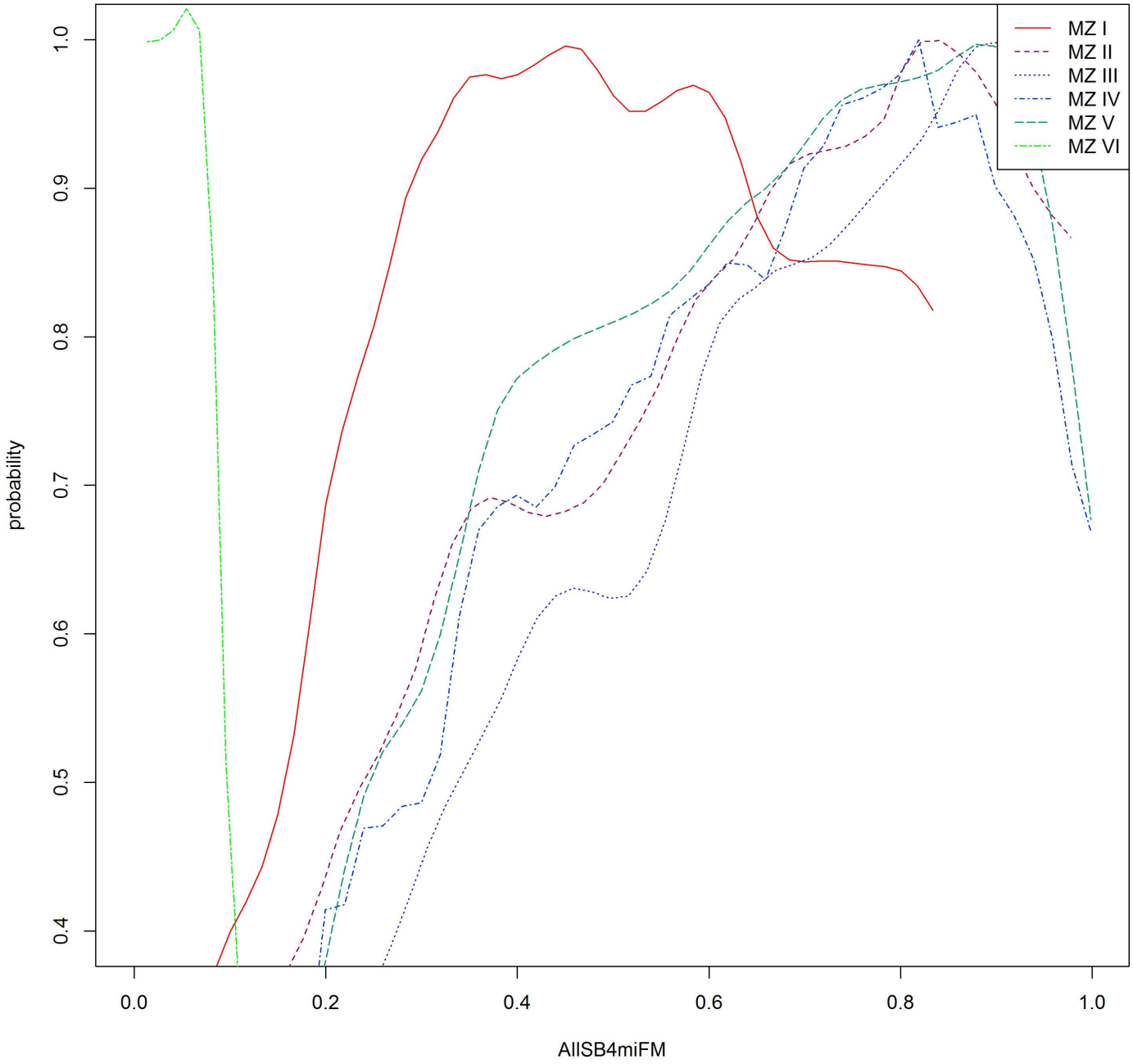
Appendix I-Gross Primary Production. Ecological variation in the ***gross primary production***, across the seven greater sage-grouse management zones. Small yellow dots are locations of active lek locations during 2010-2014.

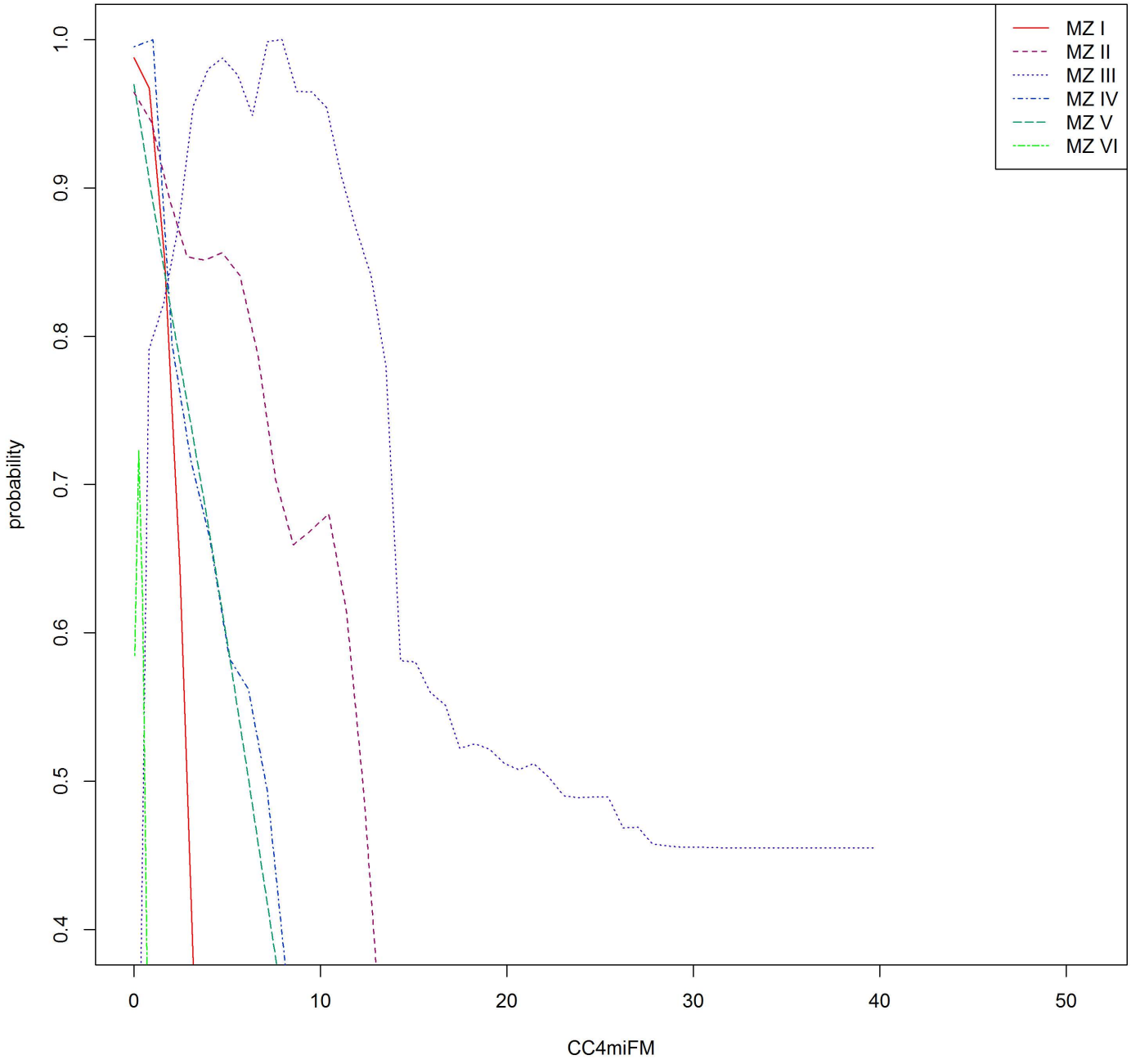


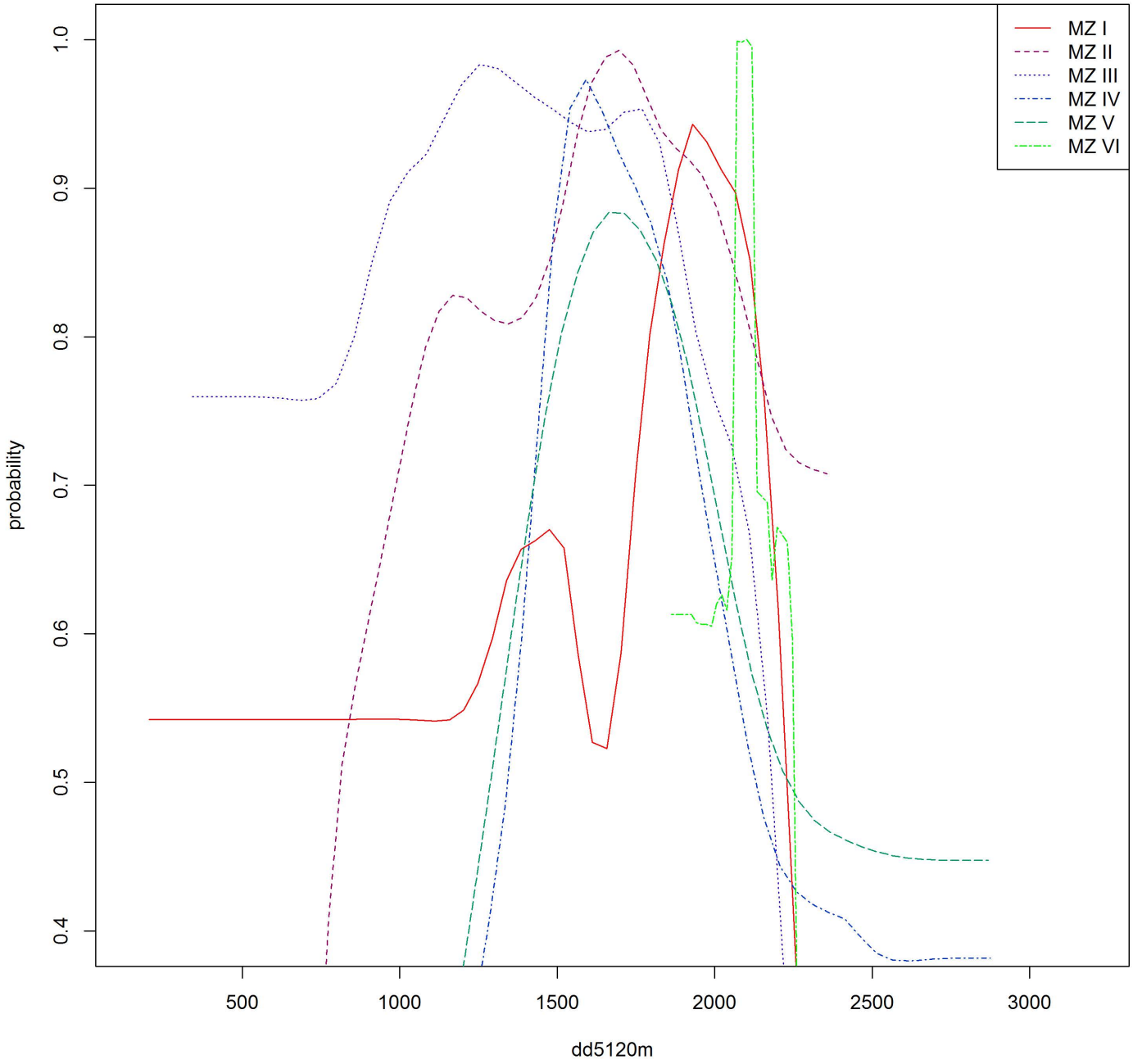
Appendix II. Remaining variables retained (variable 6th – 15th) and importance values from the model selection process for each management zone from 2010 – 2014. Importance values are scaled by management zone, so that the top variable equals 1 and the remaining variables are a proportion derived by dividing by the top variable, and are derived from probability scaled partial plots in the RandomForest package in R.

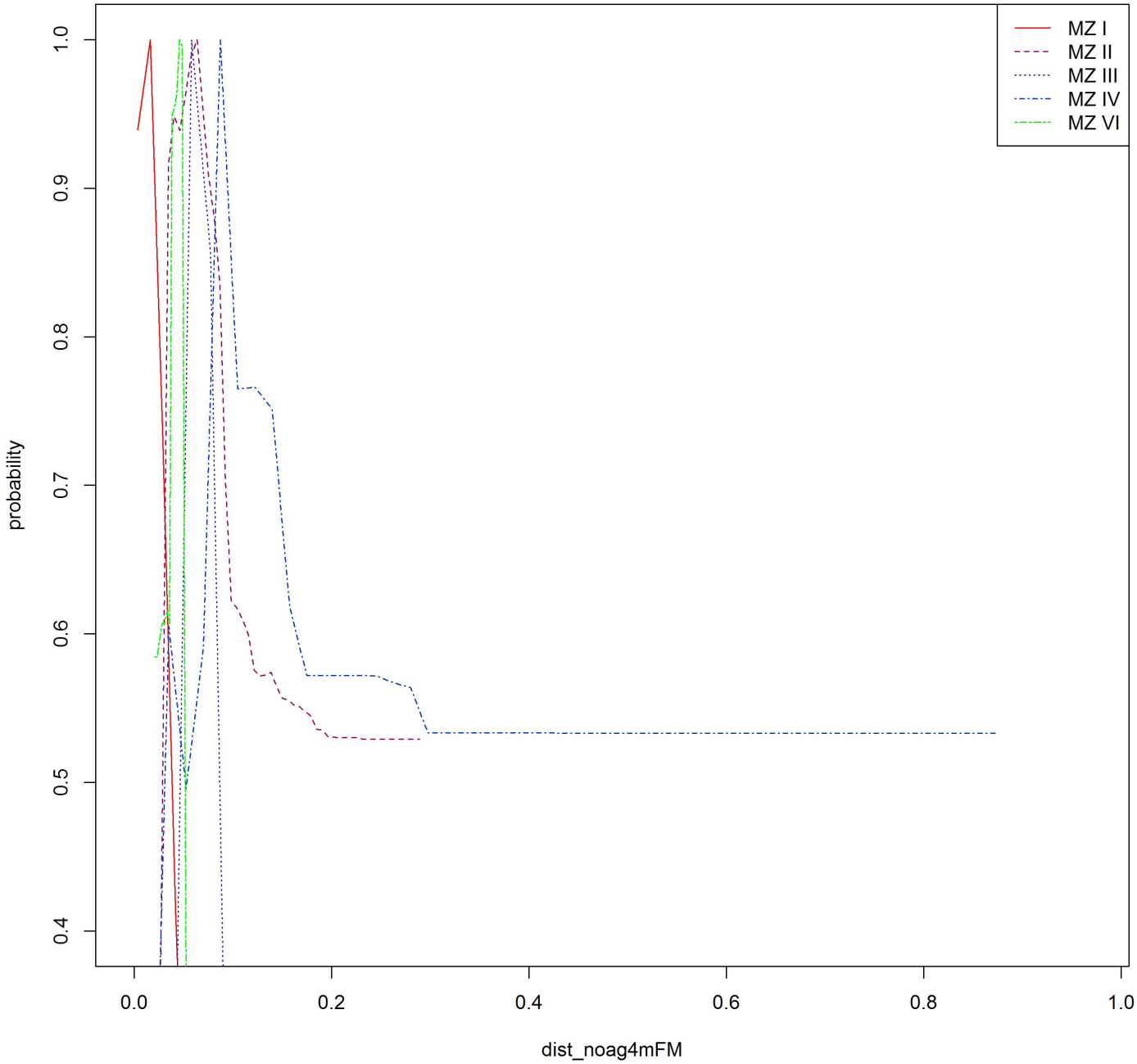
Management Zone	Variable Importance (Name & relative Rank)									
	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th
MZ I	Elev4miFM	tillage4miFM	Steep4miFM	map120m	NHD_Intermittent4miFM	dd5120m	dist_noag4mFM	GH4miFM	adi120m	ogptdsm4mifm
	0.44	0.41	0.38	0.37	0.30	0.27	0.27	0.26	0.26	0.19
MZ II	Elev4miFM	Steep4miFM	TWI4miFM	Rough4miSD	GPP0913FM	dist_noag4mFM	GH4miFM	NHD_Perennial4miFM	tillage4miFM	LowSB4miFM
	0.47	0.46	0.44	0.42	0.42	0.41	0.36	0.29	0.21	0.19
MZ III	NHD_Perennial4miFM	GPP0913FM	LowSB4miFM	map120m	Steep4miFM	TWI4miFM	SpringsandSeepsDensity4miFM	Rough4miSD	NHD_Intermittent4miFM	dist_noag4mFM
	0.36	0.35	0.31	0.31	0.30	0.25	0.22	0.16	0.10	0.09
MZ IV	Elev4miFM	Rough4miSD	LowSB4miFM	NHD_Intermittent4miFM	map120m	Steep4miFM	TWI4miFM	tillage4miFM	dist_noag4mFM	SpringsandSeepsDensity4miFM
	0.50	0.45	0.43	0.39	0.36	0.35	0.31	0.25	0.20	0.19
MZ V	GPP0913FM	CC4miFM	TWI4miFM	Elev4miFM						
	0.58	0.56	0.41	0.38						
MZ VI	GPP0913FM	dist_noag4mFM	map120m	NHD_Perennial4miFM	NHD_Intermittent4miFM	TWI4miFM	CC4miFM	tillage4miFM		
	0.21	0.20	0.13	0.07	0.05	0.04	0.04	0.03		

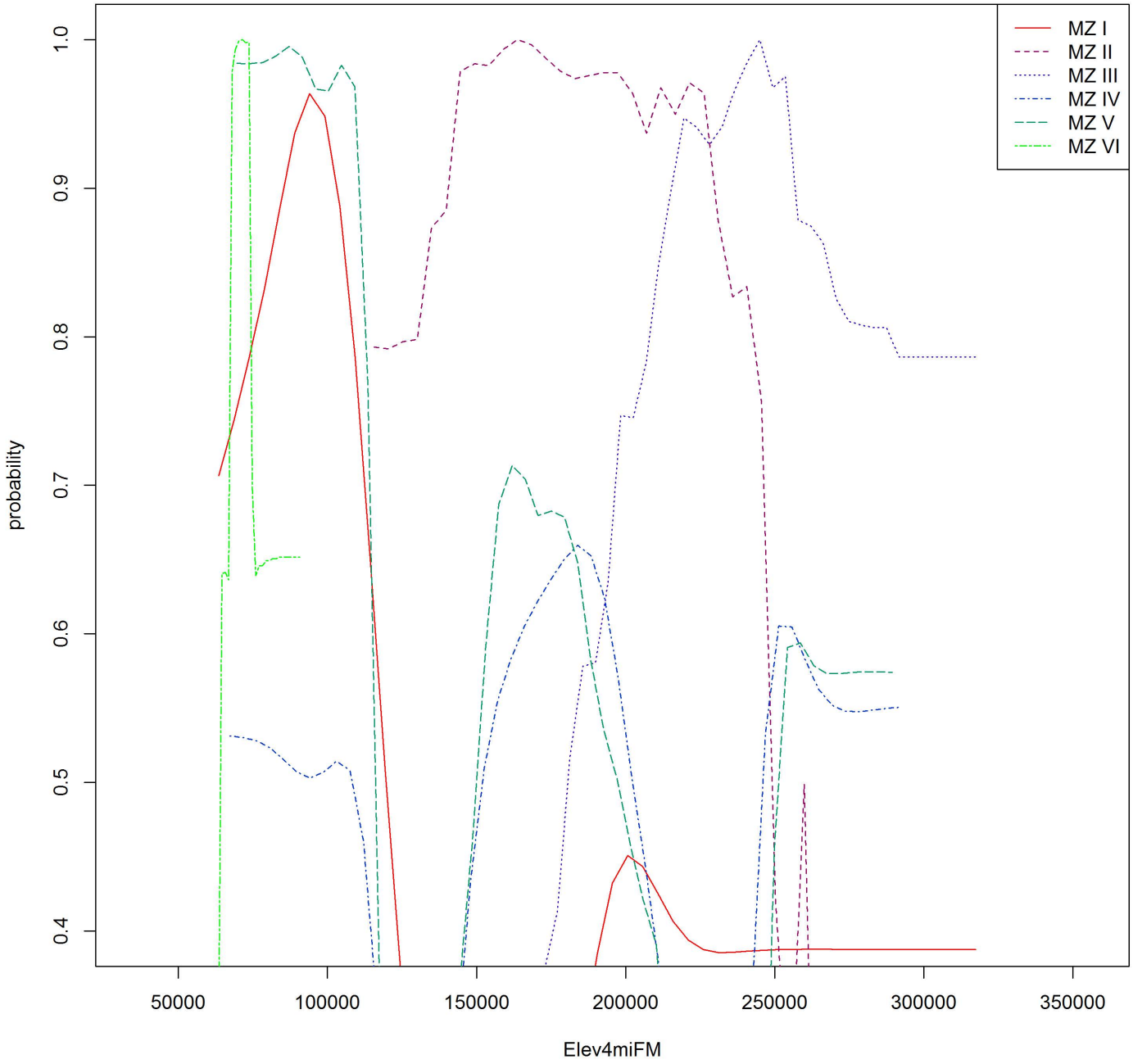

MZ I
MZ II
MZ III
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
5
10
15
20
adi120m

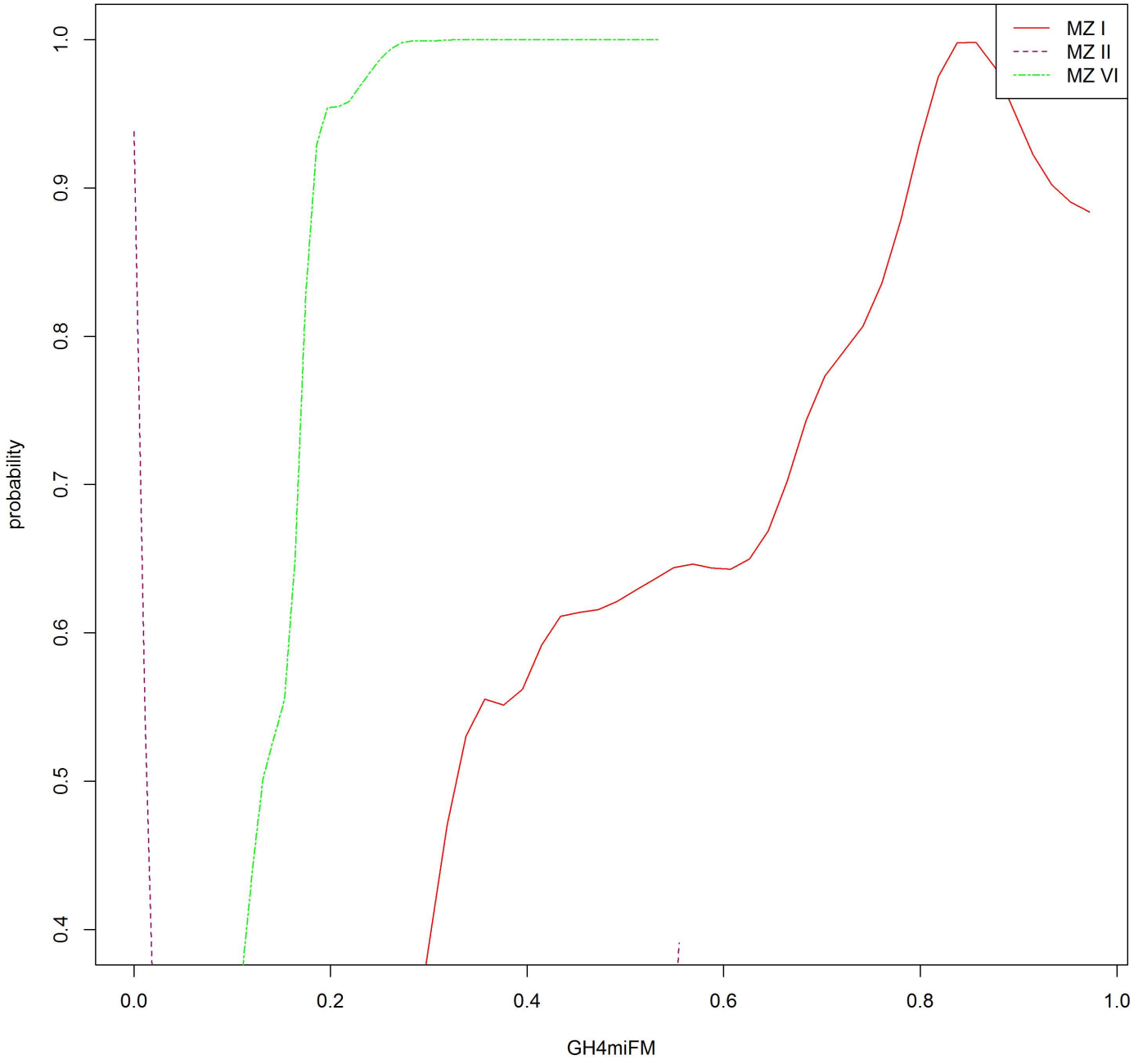

MZ I
MZ II
MZ III
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.2
0.4
0.6
0.8
1.0
AllSB4miFM

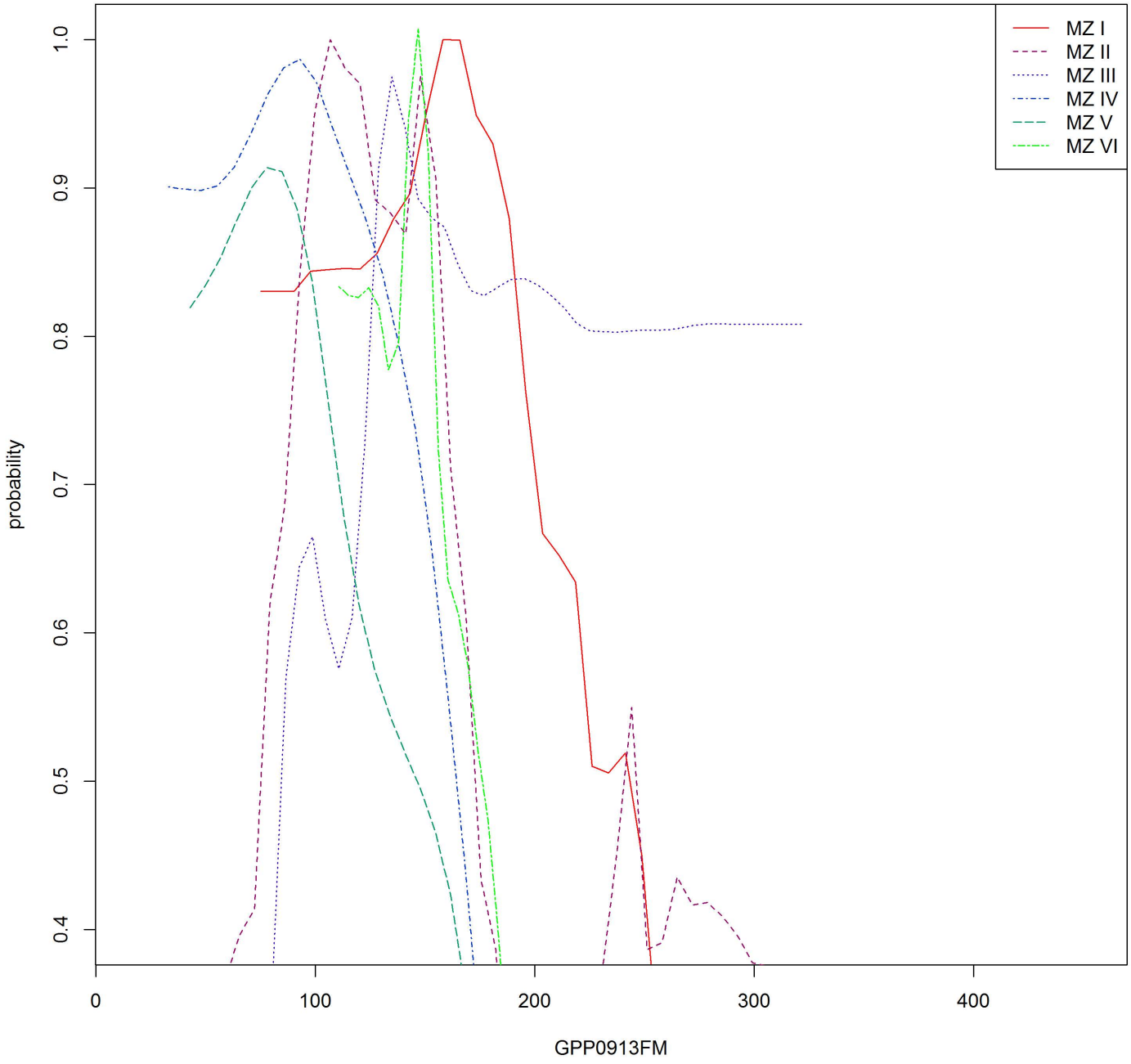

MZ I
MZ II
MZ III
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
10
20
30
40
50
CC4miFM

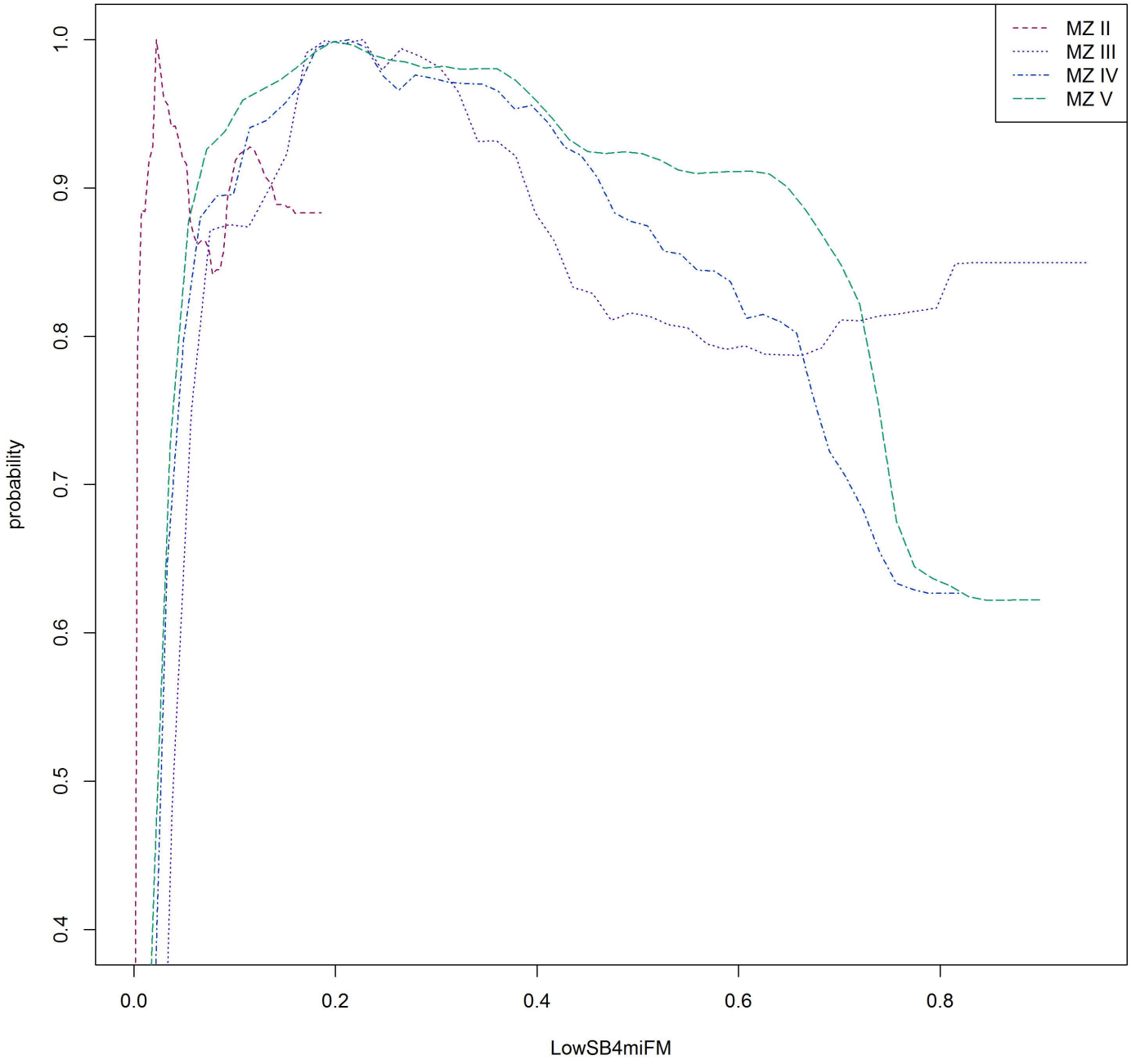

MZ I
MZ II
MZ III
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
500
1000
1500
2000
2500
3000
dd5120m

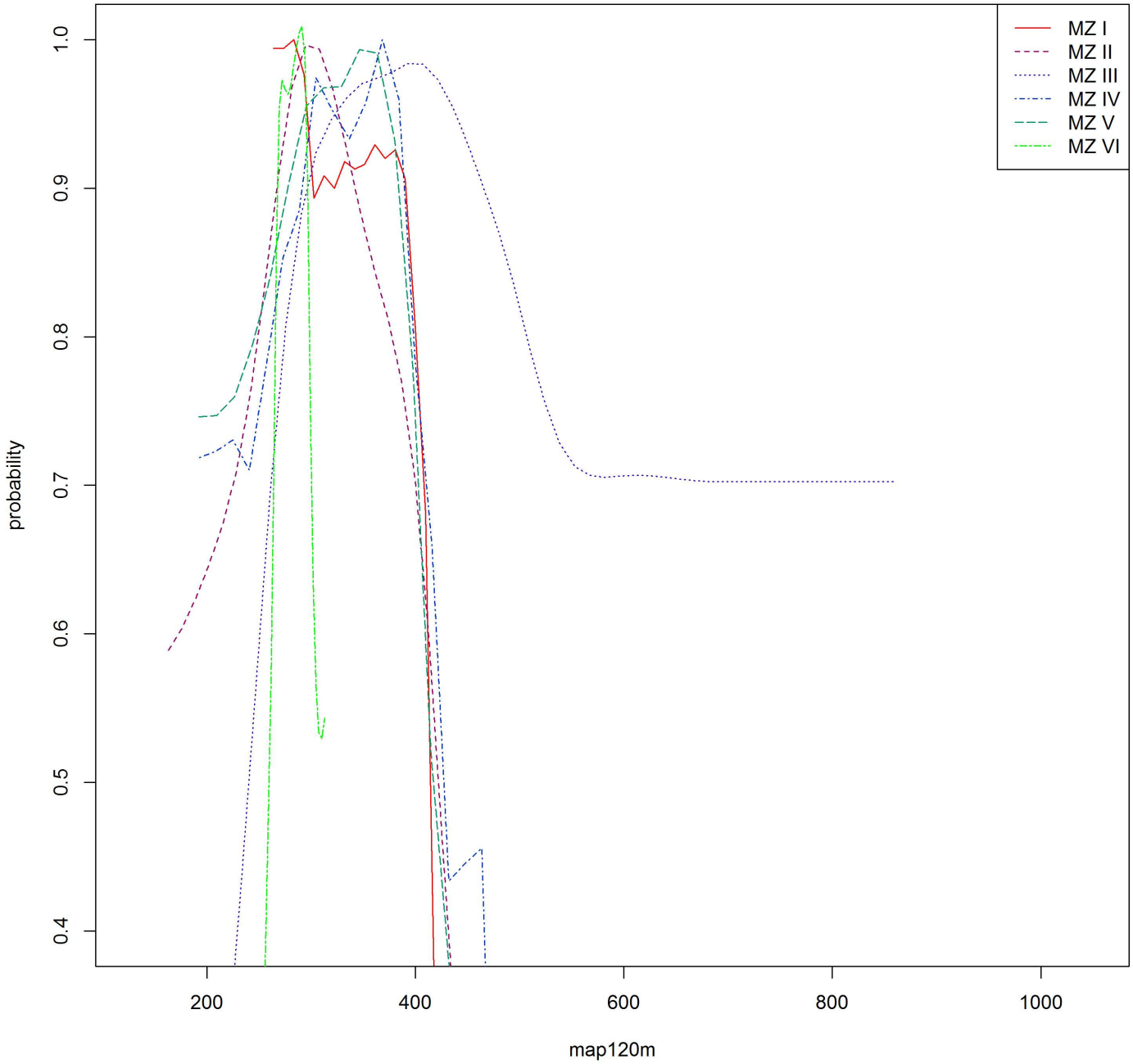

MZ I
MZ II
MZ III
MZ IV
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.2
0.4
0.6
0.8
1.0
dist_noag4mFM

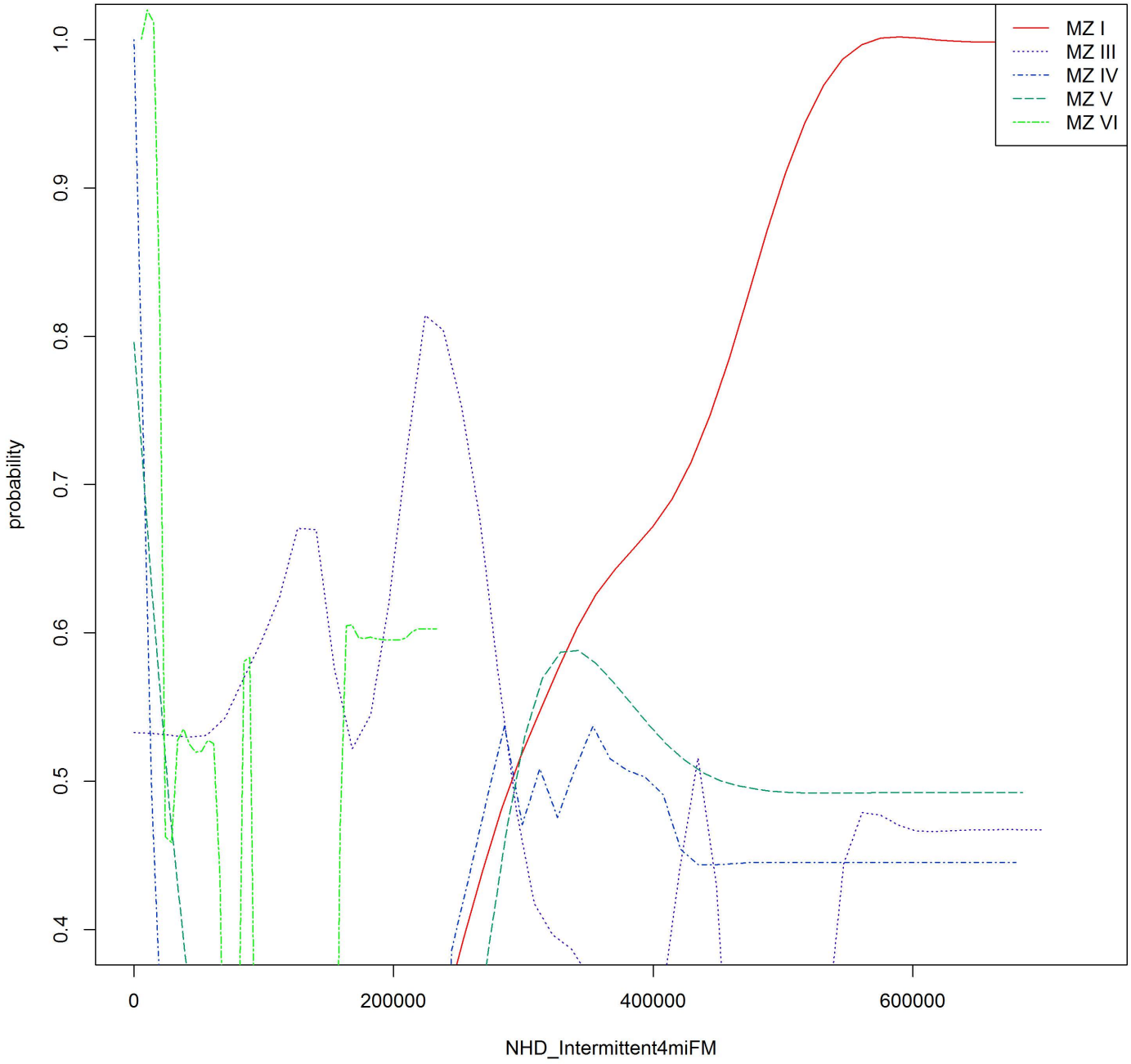

MZ I
MZ II
MZ III
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
50000
100000
150000
200000
250000
300000
350000
Elev4miFM

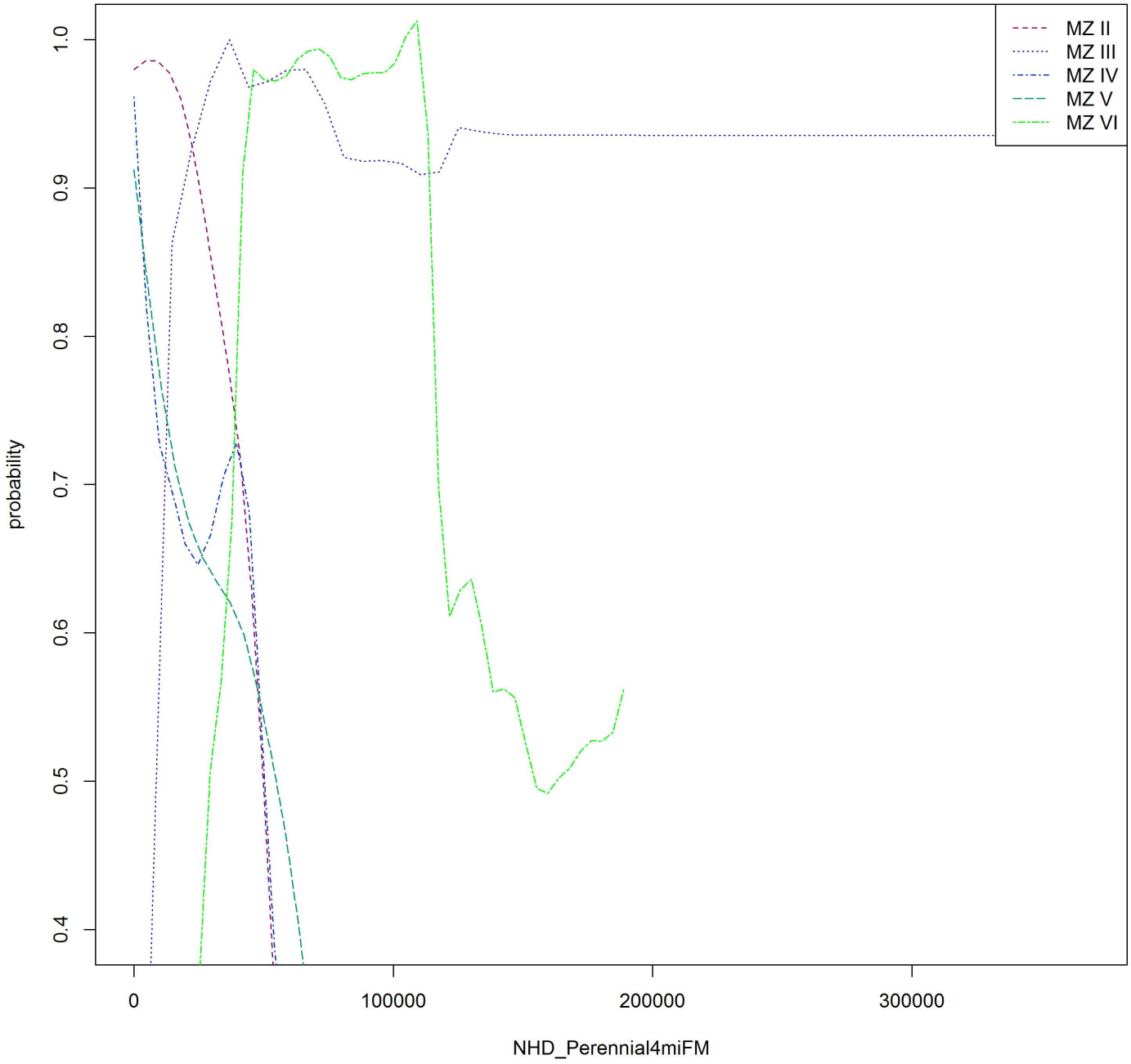

MZ I
MZ II
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.2
0.4
0.6
0.8
1.0
GH4miFM

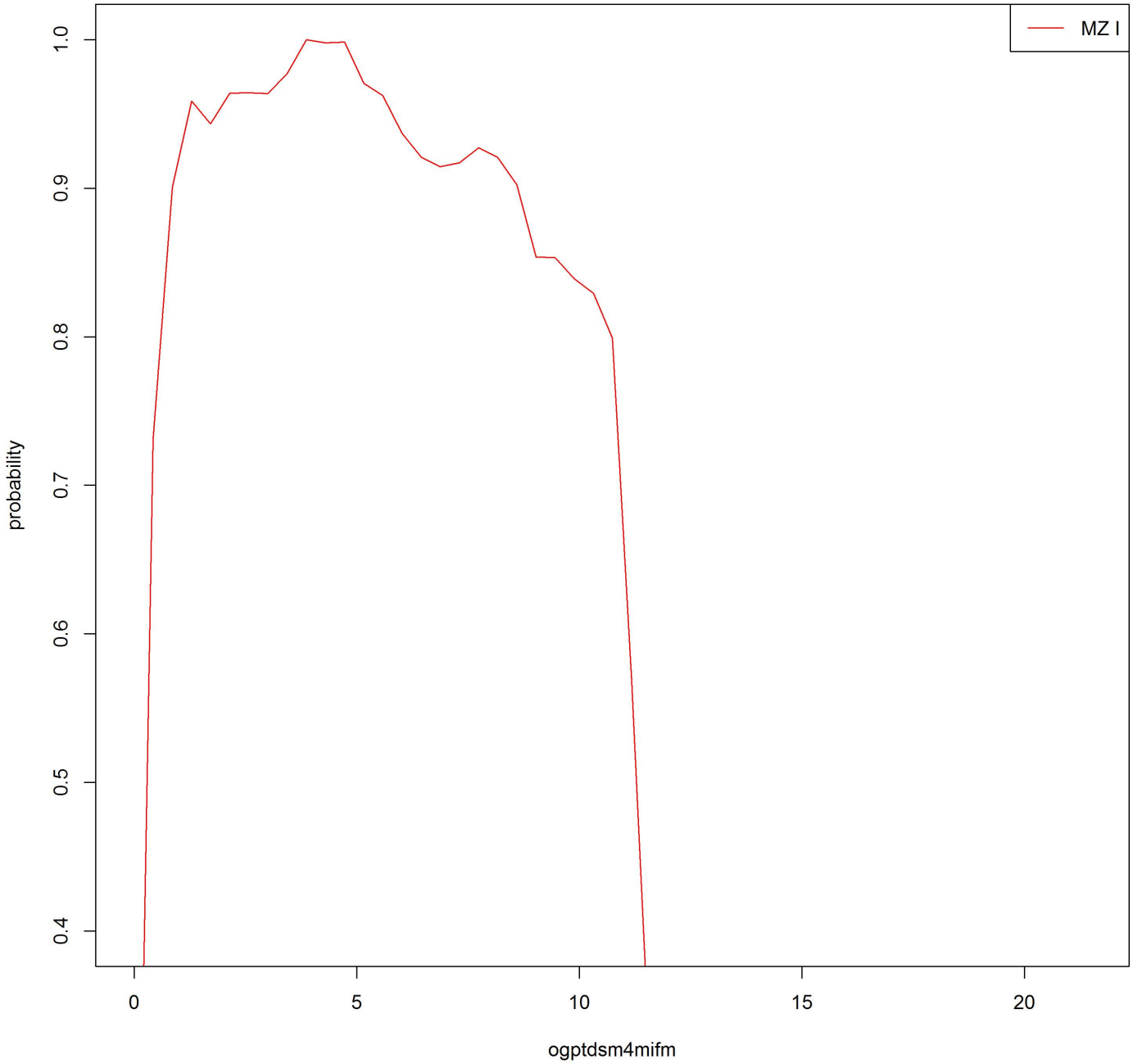

MZ I
MZ II
MZ III
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
100
200
300
400
GPP0913FM

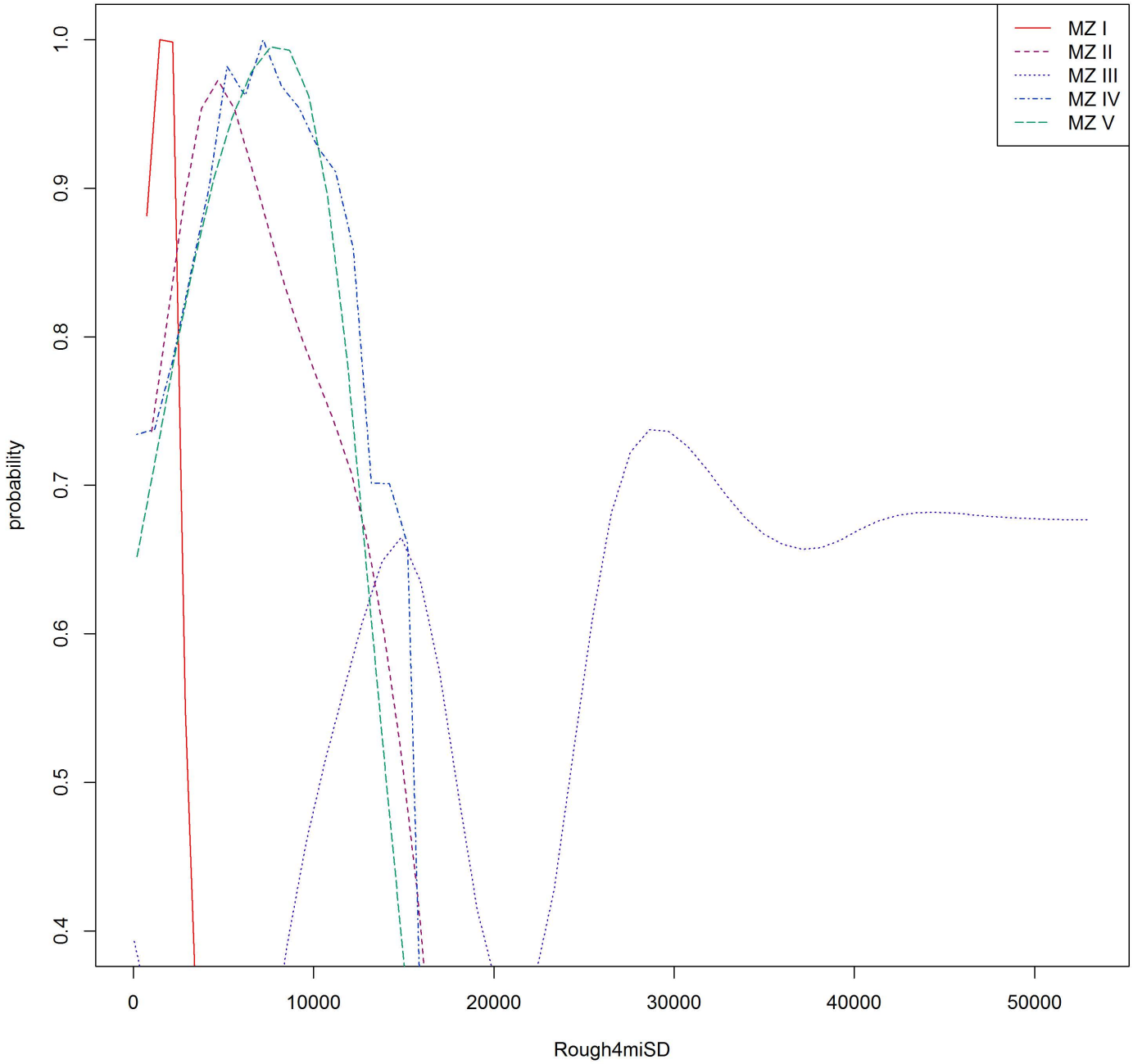

MZ II
MZ III
MZ IV
MZ V
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.2
0.4
0.6
0.8
LowSB4miFM

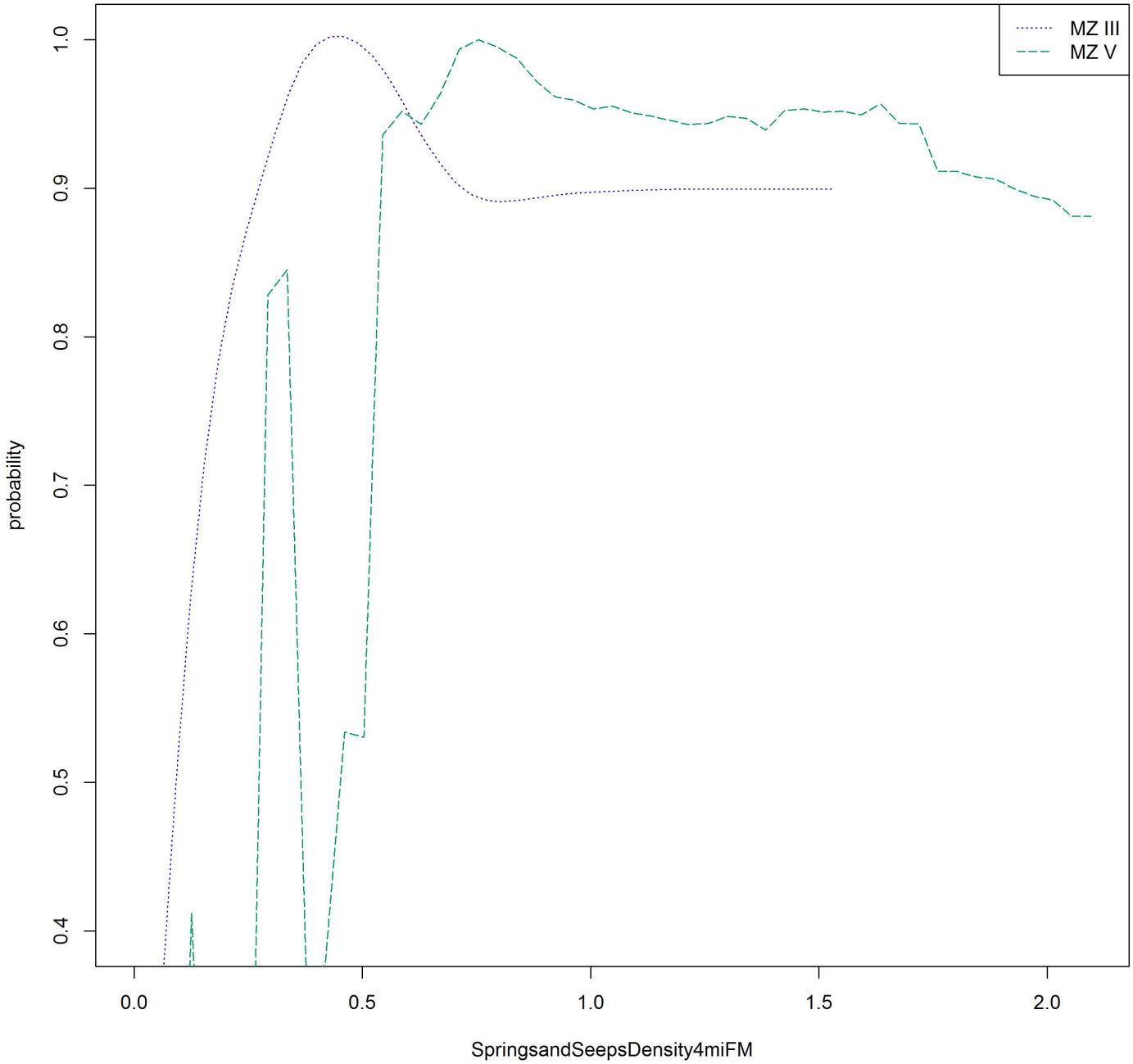

MZ I
MZ II
MZ III
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
200
400
600
800
1000
map120m

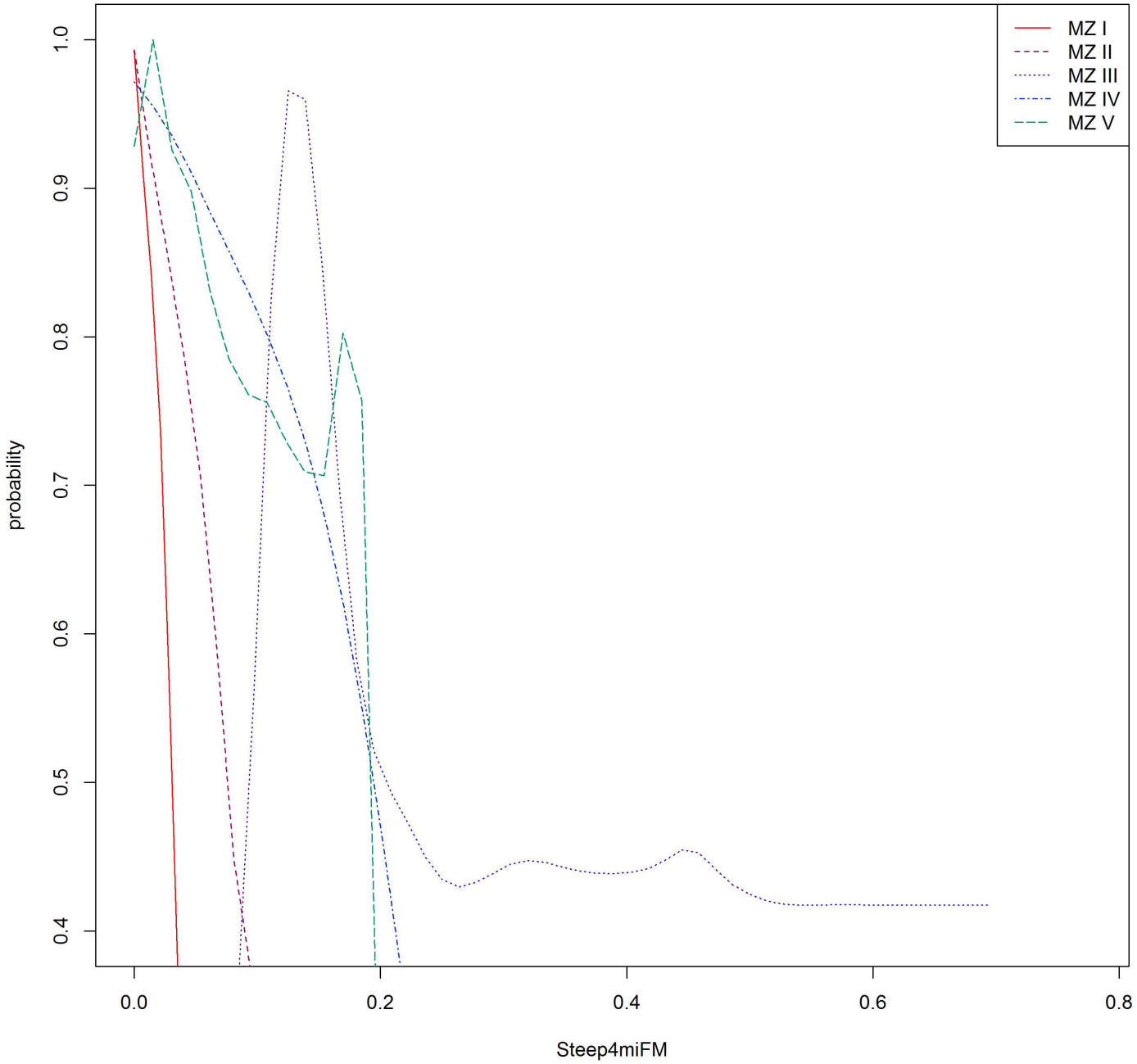

MZ I
MZ III
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
200000
400000
600000
NHD_Intermittent4miFM

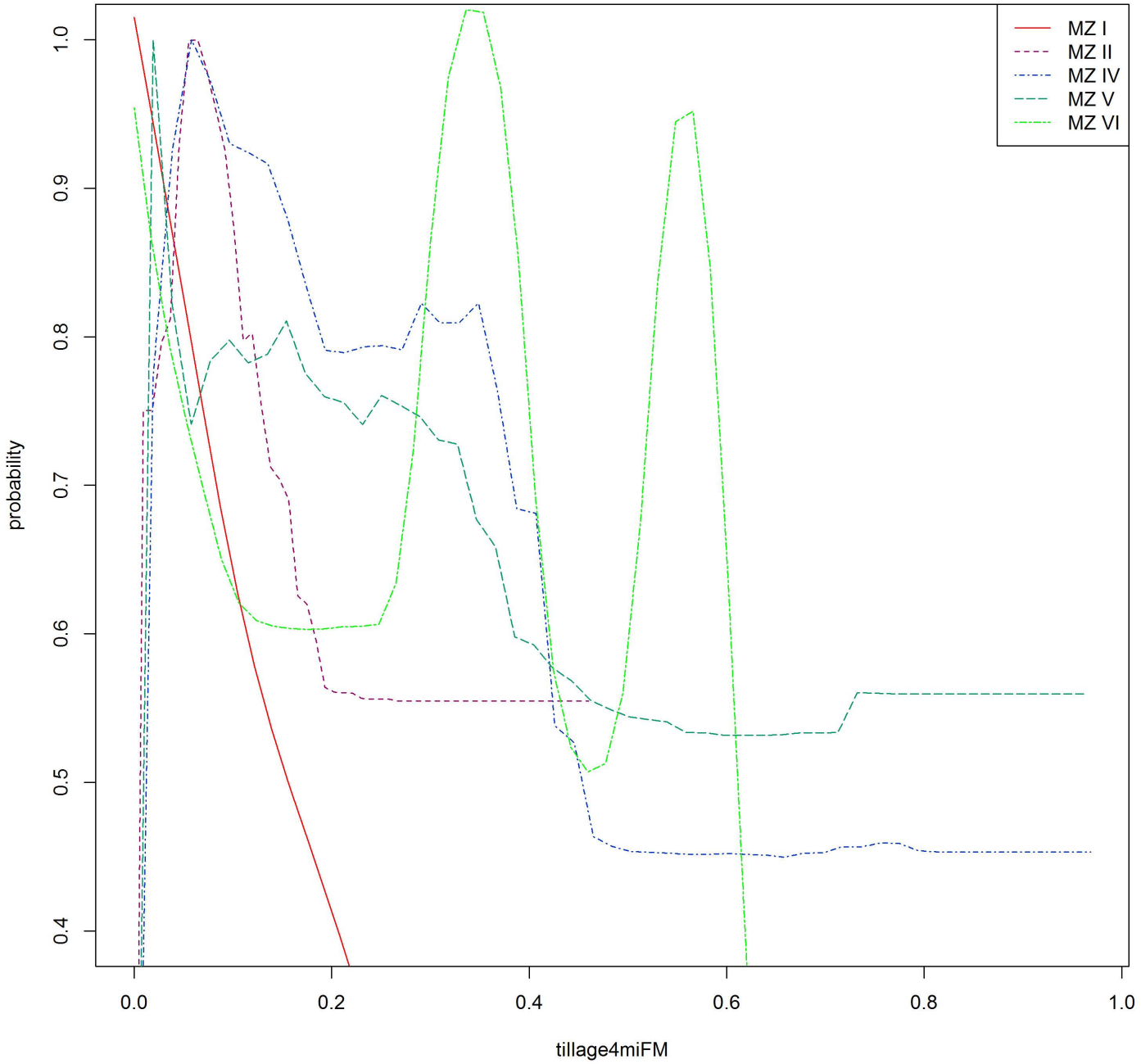

MZ II
MZ III
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
100000
200000
300000
NHD_Perennial4miFM

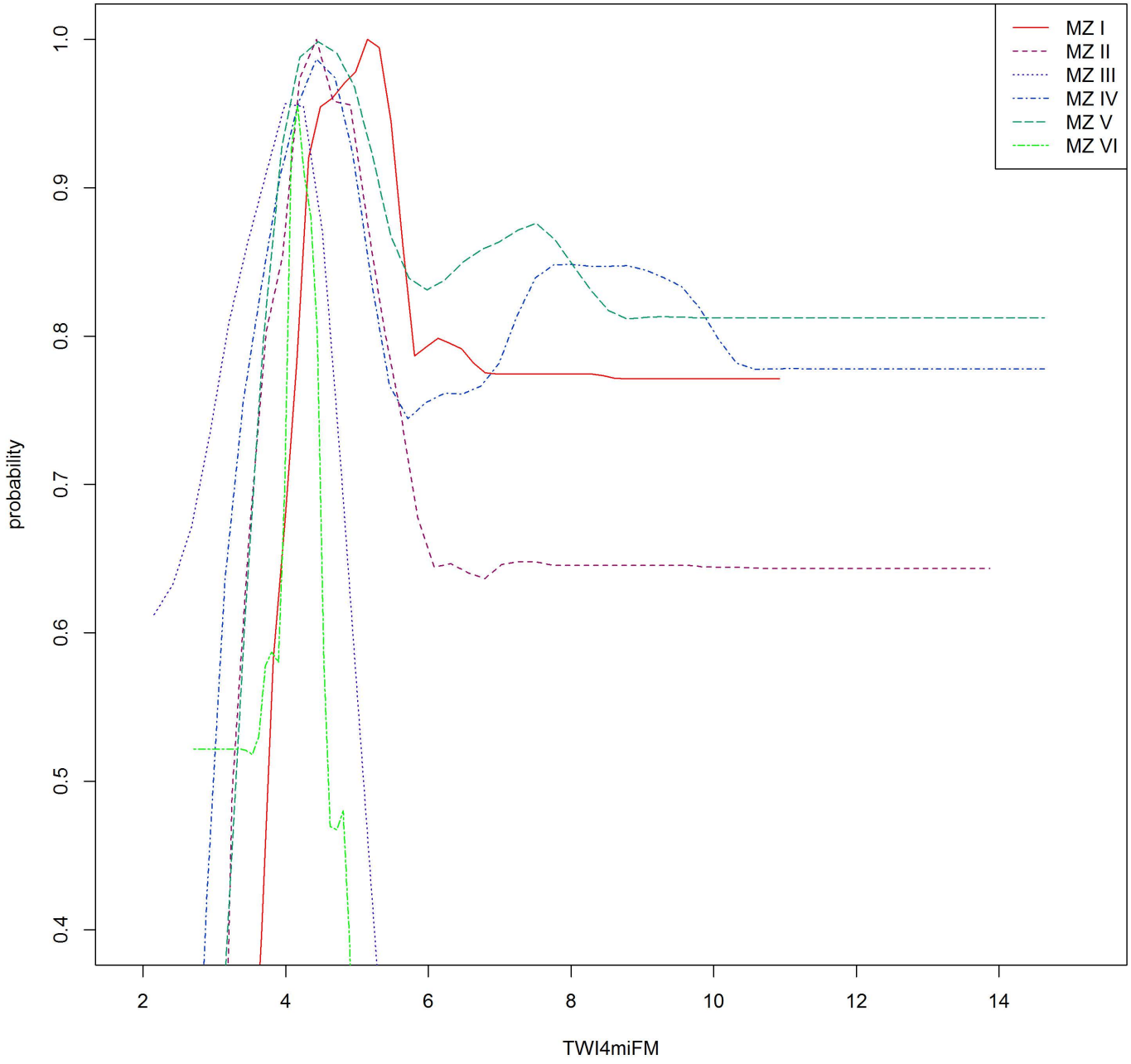

MZ I
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
5
10
15
20
ogptdsm4mifm


MZ I
MZ II
MZ III
MZ IV
MZ V
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
10000
20000
30000
40000
50000
Rough4miSD


MZ III
MZ V
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.5
1.0
1.5
2.0
SpringsandSeepsDensity4miFM


MZ I
MZ II
MZ III
MZ IV
MZ V
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.2
0.4
0.6
0.8
Steep4miFM


MZ I
MZ II
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.2
0.4
0.6
0.8
1.0
tillage4miFM


MZ I
MZ II
MZ III
MZ IV
MZ V
MZ VI
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
2
4
6
8
10
12
14
TWI4miFM