July 21, 2017

Terry A. Messmer
UMC 5230
Utah State University
Logan, Utah 84322-5230
terry.messmer@usu.edu

Guidelines for managing greater sage-grouse habitats in Utah: An empirical

Approach

Ben A. Crabb[1], Michel Kohl[2] Terry A. Messmer[2]†, David K. Dahlgren[2], Randy T. Larsen[3,4]

Shandra N. Frey[2], Eric Thacker[2], and Rick J. Baxter[4]

[1]Remote Sensing/GIS Laboratory, Quinney College of Natural Resources, Utah State University, Logan, Utah, United State of America,[2]Jack H. Berryman Institute, Department of Wildland Resources, Utah State University, Logan, Utah, United States of America, [3]Department of Plant and Wildlife Sciences, Brigham Young University, Provo, Utah, United States of America. [4]The Monte L. Bean Life Sciences Museum, Brigham Young University, Provo, Utah, United States of America,

†Corresponding author (E-mail) – terry.messmer@usu.edu (Ph) 1.435.797.3975 (fax)

Abstract

Keywords: brood habitat, *Centrocercus urophasianus*, greater sage-grouse, habitat clusters, habitat

guidelines, nesting habitat, radio-telemetry, random forest, Utah

The US Fish and Wildlife Service (USFWS) concluded in 2015 that listing the greater sage-

grouse (*Centrocercus urophasianus*) for protection under the Endangered Species Act of 1973

(USFWS 2015) was not warranted (USFWS 2015). In making this decision, the USFWS cited

that the primary species conservation threats, habitat loss and fragmentation, and lack of regulatory mechanisms had been mitigated by range wide conservation actions implemented by the states, and federal land and resources management agencies (USFWS 2013).

In response to guidance provided by the USFWS, the Bureau of Land Management (BLM) and U.S. Forest Service (USFS) revised their resource management (RMP) and land use plans (LUP) to address the sage-grouse conservation threats identified by the USFWS on federal lands (USFWS 2013). In the record of decision (ROD), the USFS and BLM cited Connelly et al. (2000) as the best available science when they established standards that will be used evaluate sage-grouse habitat conditions on federal lands (BLM 2015). To implement the RMP and LUP, the BLM and the USFS have committed to re-evaluate the existing management actions by 2020, to include livestock grazing allotments, to determine if the vegetation conditions and the landscapes approximate the ROD sage-grouse habitat guidelines.

Connelly et al. (2000) recommended that each state and province develop and implement sage-grouse conservation plans that address local differences in landscape conditions. The guidelines in addition to providing habitat recommendations, suggested actions that could be taken to mitigate the effects of anthropogenic activities on sage-grouse and their habitats.

Since the guidelines were published in 2000, new information suggests additional sources of sampling bias (Smith et al. under review). For ground-nesting birds, vegetation surrounding the nest may play an important role in mediating nest success by providing concealment from predators. Height of grasses surrounding the nest has been reported to be a driver of sage-grouse nest survival in greater sage-grouse (Gregg 1991, Gregg et al. 1994, Doherty et al. 2014). However, Gibson et al. (2016) reported that widely-used field methods can produce misleading inference on the relationship between grass height and nest success. Specifically measuring

vegetation concealment following nest fate (failure or hatch) introduces a temporal bias whereby successful nests are measured later in the season, on average, than failed nests. This sampling bias can produce inference suggesting a positive effect of grass height on nest survival, though the relationship arises due to the confounding effect of plant phenology, not an effect on predation risk (Smith et al. in press).

Sage-grouse population vital rates (i.e., nest initiation rates, nest success, clutch size, breeding success, fledging success, and survival probability of breeding aged birds) may also be affected by environmental variability, population structure and past reproductive efforts (Guttery et al. 2013, Caudill et al. 2014, Caudill et al. 2016a, Caudill et al. 2016b). Guttery et al. (2013), Caudill et al. (2014), and Dahlgren et al. (2016) demonstrated that the effect of environmental factors such as weather on chick survival and the role of juvenile and adult females and on annual herbaceous cover are paramount to the long-term conservation of the species. Thus, focusing on just one aspect of the sage-grouse ecology such as the effect of grass height on nest success may overlook the importance of seasonal habitats in sustaining viable populations and in particular the impact of seasonal variation in weather on sage-grouse and their habitats.

Connelly et al. (2000) provided habitat vegetation guidelines for nesting, breeding, and winter habitats. Their broad guidelines were gleaned habitat data recorded at sage-grouse use locations and reported in student thesis, dissertations, and peer-reviewed papers published prior to 2000. As such, the guidelines represent a range of estimated conditions which could be extrapolated to determine overall habitat suitability. However, because the Connelly et al. (2000) habitat guidelines were developed for sage-grouse primarily inhabiting sagebrush-steppe (*Artemisia* spp.), they may not be applicable to the Great Basin and desert shrub areas of the Intermountain West (Messmer 2013, Dahlgren et al. 2016). Thus, achieving the ROD habitat

vegetation standards based strictly on Connelly et al. (2000) in areas exhibiting inherent differences in precipitation and ecological site potentials may be problematic.

To address these limitations we developed an empirical approach that incorporates spatially continuous habitat vegetation, climatic, and elevation data recorded at known nest and brood locations from 1998-2014 across the state of Utah to model micro-site 4th order habitat guidelines (Stiver et al. 2015). We used a non-parametric ensemble classifier and a clustering algorithm to identify distinct sage-grouse nesting and breeding habitats across and within sage-grouse management areas (SGMAs; Utah 2013) and BLM priority (PHMA) and general habitat areas (GHMAs; BLM 2015). The relative importance of these clusters to 4th order habitat conditions were assessed using field observations recorded at radio-telemetry locations statewide (Dahlgren et al. 2016). We then generated a set of guidelines for nesting and brood-rearing habitat for the clusters that differed in 4th order habitat conditions.

Study Area

Sage-grouse location data were compiled from sage-grouse populations at 13 study areas in Utah from 1998 to 2014 (Fig. 1). These study areas represented most of the sage-grouse populations in Utah.

Populations in northern Utah inhabited sagebrush-steppe, while populations in central and southern Utah primarily used sagebrush semi-desert (West 1983). Both were shrub dominated sagebrush systems differentiated by an increased herbaceous component in higher latitude sagebrush-steppe

systems compared with lower latitude sagebrush semi-desert.

Generally, big sagebrush (*A. tridentata*) varieties dominated most landscapes within occupied habitats with Wyoming (*A. t. wyomingensis*), basin (*A. t. tridentata*), and mountain (*A. t. vaseyana*) big sagebrush at lower, mid, and high elevations, respectively. Shallow soils supported inclusions of low (*A. arbuscula*) and black (*A. nova*) sagebrush across the state. Silver sagebrush (A. cana) was present at high-elevation mesic areas and there was limited distribution of three-tip sagebrush (*A. tripartita*) in northern Utah.

Beck et al. (2003) mapped the current distribution of sage-grouse within Utah. Most of the sage-grouse populations in the state were relatively small and inhabited isolated and remote landscapes (Beck et al. 2003). The largest sage-grouse populations in the state were associated with larger, contiguous sagebrush landscapes (Dahlgren et al. 2016).

Methods

Sage-grouse telemetry database

We used 1,043 sage-grouse nest and 5,853 brood locations recorded by USU and BYU researchers using very high frequency radio-telemetry necklace-style radio transmitters from 1998-2013 to describe habitat-use areas in Utah (Dahlgren et al. 2016; Fig. 1). Field variables were measured at many of these locations following standard procedures (Connelly et al. 2003). We focused on seven variables available in the telemetry database that would be most applicable to management (Connelly et al. 2000). The included percent shrub cover (*shrub.cvr*), shrub height (*shrub.ht*), forb cover (*forb.cvr*), forb height (*forb.ht*), grass cover (*grass.cvr*), grass height (*grass.ht*), and percent sagebrush composition (*sage.comp.pct*).

Climate, elevation, and vegetation data

We obtained climatic, topographic, and vegetation cover values from freely available, nationally consistent datasets. Climatic conditions were described at 800m spatial resolution by the average annual temperature, precipitation, and minimum and maximum temperatures, over the period 1981-2010, as measured by the PRISM Climate Group. Elevation and vegetation data were taken from the LANDFIRE project (www.landfire.gov). LANDFIRE is a consortium of federal agencies that produces consistent, continuous geo-spatial raster data products at 30m resolution describing topography, vegetation, and land cover nationwide. At each nest and brood location, we tabulated the prevalence of vegetation classes at the nearest 100 pixels using the Existing Vegetation Type (EVT) 1.3.0 dataset. The EVT classifications reflect ecological systems (Comer et al. 2003), and are the finest thematic resolution of land cover data in the LANDFIRE suite of products. To spatially project the results of the clustering technique, we also collected climate and elevation values, and tabulated vegetation classes, at a grid of regularly spaced points with 1-km spacing across all locations within SGMAs, PHMAs, GHMAs, and any location within 20-km of a recorded sage-grouse location in the telemetry database. Tabulating the nearest 100 vegetation pixels implies a buffer distance of ~175 m around each telemetry and grid point location, or an area of about 9.6 ha. This local scale should provide a reasonable reflection of the vegetation structure and composition for the daily forage and cover needs of sage-grouse.

Cluster development and analysis

Our analysis of the sage-grouse telemetry data to develop the habitat clusters is summarized below.

1. We used random forest (RF) clustering (Shi and Horvath 2006) to classify k clusters in the telemetry data, for values of k ranging from 2 to 6. We assessed the stability of the clusters using the Jaccard coefficient (Hennig 2007).

2. We projected cluster classes spatially to sage-grouse radio-telemetry locations statewide by fitting a second RF trained on clusters identified in step 1.

3. We assessed cluster significance by summarizing and statistically comparing 4th order habitat conditions across clusters.

4. We selected the optimal number of clusters k based on the cluster stability and significance assessed in steps 2 and 3. We then generated habitat guidelines based on distribution of 4th order habitat conditions across clusters.

Random forest clustering

A major output of clustering methods is a dissimilarity measure. We completed an unsupervised RF to estimate dissimilarities between the sage-grouse telemetry locations using an RF predictor (Breiman 2001). A RF predictor is a collection of individual classification trees, each of which is trained on a random subset of the data. Model predictions were made by aggregating across the individual trees. The RF procedure generated a measure of proximity between two samples based on the proportion of times the trees in the forest that placed them in the same terminal node. When used as an unsupervised classifier, RF separates the observed data from a synthetic dataset created by sampling randomly from the univariate distributions of the original data. Restricting the resulting proximity matrix to the observed data yields a measure of similarity between unlabeled data.

We performed unsupervised RF on a sample of 1467 telemetry locations drawn such that not more than one data point was represented per 1 km^2 region across the state. This approach provided a spatially balanced representation of sage-grouse space use in the state relative to the resolution of the climate data, and expedited processing times. After fitting an unsupervised RF

predictor to these data, we defined a dissimilarity matrix as the square root of one minus the proximity matrix (Kaufman and Rousseeuw 1990).

We used partitioning around mediods (PAM; Kaufman and Rousseeuw 1987) to identify clusters of similar observations in the dissimilarity matrix. The PAM is similar to the commonly used kmeans classifier, which assigns observations to clusters based on minimizing the distance from each observation to the cluster centroid. In PAM, these centroids are constrained to be one of the observations in the clustering algorithm. This produced interpretable cluster centroids that were guaranteed to be members of the training data set.

We assessed the stability of the clustering solutions using the Jaccard similarity coefficient on 30 bootstrap samples of the nest and brood data (Hennig 2007). The Jaccard similarity coefficient is a statistic used for comparing the similarity of two sets. It is defined as the size of the intersection of the sets divided by the size of the union of sets; higher values indicate more stable clusters. Generally, a valid, stable cluster should exhibit a Jaccard coefficient of greater than 0.75; values from 0.6 to 0.75 indicate patterns in the data but exactly which points should belong to which cluster is doubtful; Jaccard values below 0.6 should not be trusted (Hennig 2007). We used the *clusterboot* function in the FPC package in R to estimate the average bootstrapped Jaccard values for each cluster.

Spatial projection of clusters

We used a supervised random forest model to spatially project the clustering results. In this model, PAM cluster labels were predicted using the suite of climatic, elevation, and vegetation variables at sage-grouse nest and brood telemetry locations. The ability of the random forest model to accurately predict cluster values provided an additional indication of whether the

clusters reflected real structure in the data. After fitting the model to PAM cluster labels at telemetry locations, it was used to predict cluster labels at the grid of regular points covering all areas within 20-km of SGMAs, PHMAs, GHMAs, and sage-grouse telemetry locations. The grid of points was converted to raster format, with pixel values indicating predicted cluster memberships. Finally, we ran a 3x3-pixel window over the raster surface that assigned the modal value to the center pixel. This had the effect of smoothing the raster surface by removing the salt-and-pepper effect that arises from individual pixels surrounded by neighbors of a differing class. This resulted in a more spatially consistent, less fragmented map of habitat clusters.

Assessing cluster significance

We used cluster values extracted from the smoothed raster surface for each telemetry location to assess cluster significance. We used pairwise Wilcoxon rank sum tests, utilizing a Holm p-value adjustment to control for type I error risk to determine if the utility of the k clusters for 4th order habitat characteristics differed. Alpha values of less than 0.05 were considered significant.

Results

Cluster stability

Cluster partitions exhibited greater stability for lower numbers of k (Table 1). Across 30 bootstrap iterations, the average minimum Jaccard similarity coefficient value for $k = 2$ was 0.91, and the mean was 0.94, indicating two highly stable clusters. When $k = 3$, the average minimum Jaccard was 0.81, and the mean was 0.84, again indicating stable clusters. At $k = 4$, bootstrapped Jaccard values were 0.52, 0.73, 0.82, and 0.83, indicating that three of the four clusters were stable, with the fourth capturing a dubious pattern in the data. The average bootstrapped Jaccard value for $k = 4$ was 0.73, indicating that on average the clustering solution captured quite stable

patterns in the data. When k equaled five or six, average Jaccard values were 0.68 and 0.64 respectively, indicating that these clustering solutions were progressively less stable than those found for smaller values of k.

Significance of clusters

The supervised RF used to predict clusters performed well for all values of k, with monotonically increasing OOB error rates of 1.9% when $k = 2$ to 6.1% when $k = 6$, providing additional evidence that the clusters reflect real structure in the climatic, topographic, and vegetation data. Pairwise comparisons of seven field variables in the nest data using Wilcoxon rank sum tests indicated a monotonic negative relationship between the number of clusters and the rate of significant pairwise differences (Table 2). For values of k from 2 through 6, 71%, 67%, 60%, 47%, and 40% of all pairwise comparisons differed, respectively. Notably, none of the pairwise cluster comparisons showed a significant difference in shrub cover at nest locations. The other six 4th order habitat variables exhibited a generally declining rate of significant differences with increases in k. Average rates of significant pairwise differences across clusters for values of k from 2 to 6 were 0 for shrub cover, 0.28 for forb height, 0.55 for shrub height, 0.73 for sage.comp.pct, 0.77 for grass cover, 0.78 for forb cover, and 0.87 for grass height.

Pairwise comparisons using the brood data show a similar decreasing relationship between k and the rate of significant pairwise differences (Table 3), although it is not strictly monotonic. For values of k from 2 through 6 in the brood data, 71%, 95%, 81%, 77%, and 67% of all pairwise comparisons differed, respectively. The clusters were able to distinguish significant differences in the shrub cover variable for the majority of pairwise comparisons with the exception of the $k = 2$ solution, in which shrub cover did not differ. Average rates of significant pairwise differences across clusters for values of k from 2 to 6 were 0.4 for

sage.comp.pct, 0.53 for grass cover, 0.60 for shrub cover, 0.67 for forb height and shrub height, 0.87 for grass height, and 0.93 for forb cover.

The distributions of the seven 4th order habitat variables, elevation, and precipitation, are shown for $k = 4$ clusters for nest and brood data in Figures 1 and 2, respectively. The clusters quite effectively separated elevation and precipitation, and performed variably on the seven field variables, with minimal differences between shrub cover across clusters, particularly in the nest data. Visual assessment of the density curves suggests general agreement with the findings of the pairwise Wilcoxon tests.

Spatial projection of clusters

When the $k = 4$ clusters are spatially projected across the state of Utah, two of the clusters cover very large extents, while two cover much more localized areas. The two clusters covering localized areas are located in the Parker Mountain area in south-central Utah. The 'Parker 2' area is the highest-elevation cluster in the state, followed by 'Parker 1.'

Selection of optimal number of clusters

We selected $k = 4$ to represent classes of sage-grouse nesting and brooding habitats across the state. With four clusters, 60% and 81% of pairwise comparisons of the field variables differed significantly in the nest and brood data, respectively. Cluster stability was reasonably high as assessed using the Jaccard coefficient, although the 'Parker 2' cluster exhibited a bootstrapped Jaccard values of 0.52, suggesting that it may not reflect meaningful structure in the data. The other clusters exhibited bootstrapped Jaccard values of 0.73 ('Parker 1'), 0.82 ('Med-high'), and 0.83 ('Low'). Hence, three of the four clusters exhibited good stability.

While the $k = 2$ and $k = 3$ solutions had more desirable Jaccard and pairwise Wilcoxon results than the $k = 4$ solution, we decided to use the higher k value because we are trying to identify meaningful local habitat guidelines, and a cluster solution that is too coarse is at risk of containing information that does not sufficiently reflect local conditions. Indeed, a higher value of k may prove more desirable than the four cluster solution we present here.

While we had expected more differentiation of local habitat types across the state, future refinement of this approach may improve the clustering results. For example, performing the clustering technique presented here on subsets of telemetry locations defined by temporal windows may produce improved results. The $k = 4$ solution presented here nonetheless retained satisfactory, if not exemplary, quantitative indicators (Jaccard, Wilcoxon values) while classifying the state into habitat areas that generally agree with our knowledge of conditions on the ground. Because not all of the 4th order habitat conditions differ across clusters, we suggest that habitat guidelines be adapted from Tables 4 and 5, perhaps by taking averages across non-differing clusters identified in the "same_as" column or by visual assessment of density curves in Figures 1 and 2.

Discussion
Management Implications

Literature Cited

Breiman, L. 2001. Random forests. Machine Learning 45: 5-32.

Comer, P., D. Faber-Langendoen, R. Evans, S. Gawler, C. Josse, G. Kittel, S. Menard, M. Pyne, M. Reid, K. Schulz, K. Snow, and J. Teague. 2003. Ecological Systems of the United States: A Working Classification of U.S. Terrestrial Systems. NatureServe, Arlington, Virginia

Hennig, C. 2015. fpc: Flexible procedures for clustering. R package version 2.1-10. https://CRAN.R-project.org/package=fpc

Hennig, C. 2007. Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis* 52: 258-271.

Kaufman, L, PJ Rousseeuw. 1990. Finding groups in data: an introduction to cluster analysis. New York, Wiley.

Kaufman, L. and Rousseeuw, P.J. 1987. Clustering by means of Medoids. In: Statistical Data Analysis Based on the L1-Norm and Related Methods, edited by Y. Dodge, North-Holland, 405–416.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Shi, T., S. Horvath. 2006. Unsupervised learning with Random Forest predictors. Journal of Computational and Graphical Statistics 15(1): 118-138.

Table 1. Cluster stability assessments using the Jaccard similarity coefficient. K = 3 clusters was selected as a reasonable representation of stable clusters, in which the minimum Jaccard value exceeded 0.75. We did not select the 2 cluster solution to enable greater localization of habitat guidelines.
[note: dissim = sqrt(1 – proximity); all 1467 telemetry locations (1 per 1 km pixel) were used.]

Number of clusters (k)	Minimum Jaccard	Mean Jaccard
2	0.91	0.94
3	0.81	0.88
4	0.52	0.73
5	0.55	0.68
6	0.26	0.64

Table 3. Proportion of pairwise comparisons of 4th order habitat characteristics in nest data that exhibited significant differences at a 95% confidence level, as assessed using Wilcoxon rank sum tests, using the Holm p-value adjustment method to account for familywise Type I error. The last column indicates the overall rate of occurrence of significant differences.

# of clusters (k)	Max # of signif diffs = k(k-1)/2	*shrub.cvr*	*shrub.ht*	*forb.cvr*	*forb.ht*	*grass.cvr*	*grass.ht*	*sage.comp.pct*	Signif diff rate
2	1	0.00	1.00	1.00	0.00	1.00	1.00	1.00	0.71
3	3	0.00	0.33	0.67	0.67	1.00	1.00	1.00	0.67
4	6	0.00	0.67	1.00	0.33	0.67	0.83	0.67	0.60
5	10	0.00	0.40	0.70	0.20	0.60	0.80	0.60	0.47
6	15	0.00	0.33	0.53	0.20	0.60	0.73	0.40	0.40
Average rate		0.00	0.55	0.78	0.28	0.77	0.87	0.73	

Table 4. Proportion of pairwise comparisons of 4th order habitat characteristics in brood data that exhibited significant differences at a 95% confidence level, as assessed using Wilcoxon rank sum tests, using the Holm p-value adjustment method to account for familywise Type I error. The last column indicates the overall rate of occurrence of significant differences.

# of clusters (k)	Max # of signif diffs = k(k-1)/2	*shrub.cvr*	*shrub.ht*	*forb.cvr*	*forb.ht*	*grass.cvr*	*grass.ht*	*sage.comp.pct*	Signif diff rate
2	1	0.00	1.00	0.00	1.00	1.00	1.00	1.00	0.71
3	3	0.67	1.00	1.00	1.00	1.00	1.00	1.00	0.95
4	6	0.67	0.67	1.00	0.67	1.00	1.00	0.67	0.81
5	10	0.70	0.90	0.80	0.80	0.70	1.00	0.50	0.77
6	15	0.60	0.67	0.93	0.67	0.53	0.87	0.40	0.67
Average rate		0.53	0.85	0.75	0.83	0.85	0.97	0.71	

Table 4. Summaries of 4th order brooding habitat characteristics per cluster. Column name: *cl_no*: numeric cluster identifier; cluster: descriptive cluster name; *n_sg*: number of brood telemetry locations; *n_obs*: number of telemetry locations with observations of variable; *n_diff*: number of significantly differing clusters based on pairwise Wilcoxon test at 95% confidence level with Holm p-value correction; *same_as*: indicates *cl_no* values of clusters that do not differ; *q5... q95*: quantiles of observations.

Brood Variable	cl_no	cluster	n_sg	n_obs	mean	n_diff	same_as	q5	q10	q15	q20	q30	q50	q95
shrub.cvr	2	Low	849	266	19.65	***		0.0	5.3	7.4	9.0	12.1	19.0	39.1
	1	Med-high	3271	850	26.98	**	4	3.8	7.4	10.7	12.7	16.7	24.3	60.2
	3	Parker 2	1060	370	22.84	***		7.8	10.6	12.7	14.9	17.4	22.1	39.9
	4	Parker 1	663	246	25.45	**	1	10.8	14.0	15.8	17.5	20.2	25.1	40.8
shrub.ht	2	Low	849	162	39.22	*	1, 3	18.6	19.9	21.5	23.4	26.6	32.6	83.7
	1	Med-high	3271	405	42.13	*	2, 3	13.5	17.1	19.3	22.0	25.7	33.7	72.5
	3	Parker 0	1060	100	34.09	*	1, 2	11.4	14.2	16.0	17.5	22.7	33.7	61.2
	4	Parker 1	663	97	22.75	***		8.7	9.4	10.3	11.0	12.2	14.6	52.9
forb.cvr	2	Low	849	267	6.42	**	4	0.0	0.3	0.5	0.8	1.3	3.2	24.3
	1	Med-high	3271	843	13.83	***		1.3	2.5	3.3	4.4	6.7	12.4	33.0
	3	Parker 2	1060	370	9.96	***		1.8	2.3	2.9	3.5	5.3	8.6	22.8
	4	Parker 3	663	246	4.87	**	2	0.5	1.0	1.2	1.5	2.1	3.5	14.0
forb.ht	2	Low	849	157	11.86	***		2.8	4.0	4.6	5.1	6.0	9.3	26.9
	1	Med-high	3271	402	12.1	***		5.3	6.3	7.2	7.8	8.6	10.4	24.2
	3	Parker 4	1060	100	6.03	**	4	3.3	3.5	4.0	4.1	4.5	5.3	10.7
	4	Parker 5	663	97	6.81	**	3	3.1	3.7	4.2	4.4	5.0	6.2	13.3
grass.cvr	2	Low	849	267	12.47	*	3, 4	0.1	2.0	3.2	3.9	5.0	8.6	35.5
	1	Med-high	3271	845	18.03	***		4.4	5.5	7.1	8.4	11.1	15.8	40.8
	3	Parker 6	1060	370	10.75	*	2, 4	4.1	5.1	5.9	6.6	8.0	10.1	19.6
	4	Parker 7	663	246	10.19	*	2, 3	2.5	3.9	4.8	5.5	7.1	9.4	22.6
grass.ht	2	Low	849	171	26.04	**	1	12.0	16.4	18.5	19.5	20.4	25.0	42.9
	1	Med-high	3271	407	25.94	**	2	10.6	12.4	13.8	15.1	17.7	23.1	47.1
	3	Parker 8	1060	100	10.11	***		5.8	6.5	7.1	7.6	8.0	9.5	15.4

Nest Variable	cl_no	cluster	n_sg	n_obs	mean	n_diff	same_as	q5	q10	q15	q20	q30	q50	q95
	4	Parker 9	663	97	11.03	***		6.3	6.9	7.6	8.5	9.2	10.8	17.2
	2	Low	849	85	57.55	**	1	0.0	0.0	0.0	15.1	28.2	67.3	100.0
sage.comp.pct	1	Med-high Parker	3271	133	68.32	*	2, 4	5.8	20.1	36.3	46.9	59.8	76.8	100.0
	3	10 Parker	1060	26	96.89	***		89.1	92.6	94.8	96.3	99.0	100.0	100.0
	4	11	663	22	83.27	**	1	45.3	62.2	65.7	73.5	79.4	87.8	99.9

Table 5. Summaries of 4th order nest habitat characteristics per cluster. Column name: *cl_no*: numeric cluster identifier; cluster: descriptive cluster name; *n_sg*: number of brood telemetry locations; *n_obs*: number of telemetry locations with observations of variable; *n_diff*: number of significantly differing clusters based on pairwise Wilcoxon test at 95% confidence level with Holm p-value correction; *same_as*: indicates *cl_no* values of clusters that do not differ; *q5… q95*: quantiles of observations.

Nest Variable	cl_no	cluster	n_sg	n_obs	mean	n_diff	same_as	q5	q10	q15	q20	q30	q50	q95
	2	Low	144	114	24.93	**	3	5.84	11.11	13.61	15.5	19.09	25.82	44.38
shrub.cvr	1	Med-high	662	251	29.47	*	3, 4	13.17	14.5	17.44	19.33	21.75	27.55	51.83
	3	Parker 2	165	117	28.95	*	1, 3	6.96	10.91	18.81	21.31	24.03	29.23	45.38
	4	Parker 1	65	37	26.06		1, 2, 4	8.08	8.08	8.66	11.47	16.9	25.93	45.91
	2	Low	144	113	44.92	**	3	20.56	26.07	27.61	29.94	32.65	41.76	82.02
shrub.ht	1	Med-high	662	204	35.59	**	3	16.78	19.73	21.57	23.73	26.8	33.65	63.61
	3	Parker 2	165	51	27.3	***		12.34	12.94	13.99	14.37	18.98	25.47	47.75
	4	Parker 1	65	13	38.23	*	1, 2	19.73	23.91	24.81	26.48	32.45	39.77	54.65
	2	Low	144	113	7.99	**	3	0	0.35	0.8	1.22	2.04	4.1	27.81
forb.cvr	1	Med-high	662	271	13.51	***		1.06	2.6	3.73	4.5	6.35	12.05	31
	3	Parker 2	165	117	2.96	***		0.16	0.3	0.42	0.59	0.85	1.28	9.21
	4	Parker 1	65	37	5.95	**	2	0.72	1.05	1.2	1.26	1.97	4.35	19.49
	2	Low	144	108	9.81	*	1, 3	2.18	4.04	4.63	5.4	6.92	9.78	17.9
forb.ht	1	Med-high	662	204	9.14	*	2, 3	3.35	4.31	5.01	5.92	6.89	8.58	16.12
	3	Parker 2	165	51	7.07	**	3	3.64	4	4.28	4.4	4.91	6.33	13.37

	4	Parker 1	65	13	7.1		1, 2, 4	2.87	3.07	3.62	3.82	4.27	6.81	12.4
	2	Low	144	113	17.1	**	1	2.58	3.05	3.49	4.25	6.55	14.5	41
grass.cvr	1	Med-high	662	250	15.14	**	2	3.27	4.31	6.4	7.4	9.6	13.85	32.26
	3	Parker 2	165	117	6.96	**	3	2.11	2.53	2.7	3.25	4.15	6.18	13.98
	4	Parker 1	65	37	6.7	**	4	4.05	4.39	4.46	4.57	5.25	6.13	12.18
	2	Low	144	113	21.03	***		9	9.91	12.83	14.36	15.58	20	37.51
grass.ht	1	Med-high	662	204	17.36	***		8.57	10.01	11.17	12.11	13.27	16.05	31.87
	3	Parker 2	165	51	11.13	**	3	7.44	7.78	8.12	8.59	9.45	10.73	19.13
	4	Parker 1	65	13	10.14	**	4	6.64	6.95	7.9	8.32	8.82	10.15	13.86
	2	Low	144	52	61.7	***		17.64	27.34	34.71	37.47	44.43	62.89	98.54
sage.comp.pct	1	Med-high	662	55	75.07	**	4	2.76	27.45	45.65	50.42	71.47	87.83	100
	3	Parker 2	165	38	82.28	**	1	51.95	56.66	67.15	69.45	78.36	88.03	98.45
	4	Parker 1	65	9	97.11	***		88.16	91.28	93.94	96.13	98.38	99.35	100
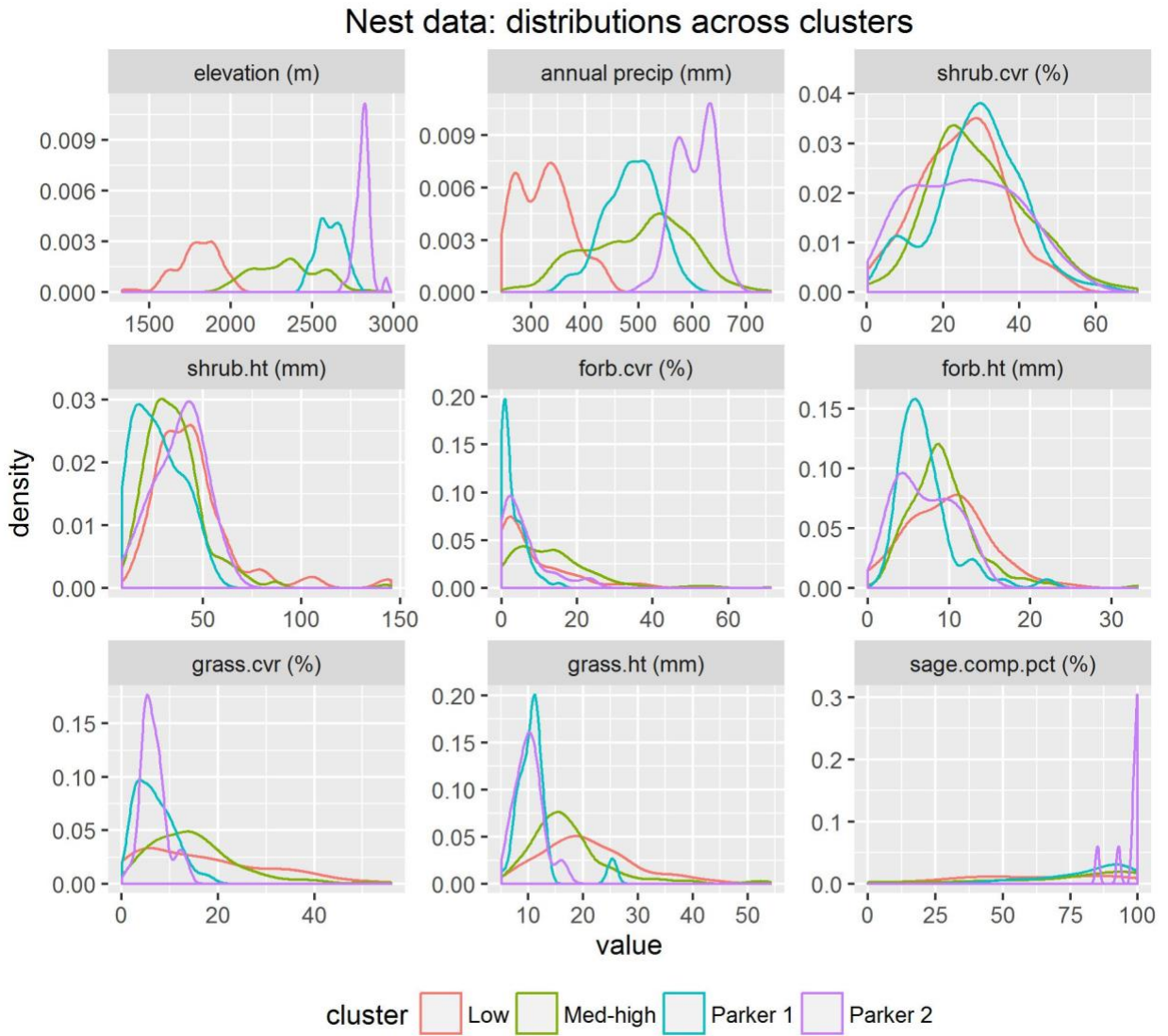
Figure 1: Distributions of variables across k=4 clusters in the nest data.
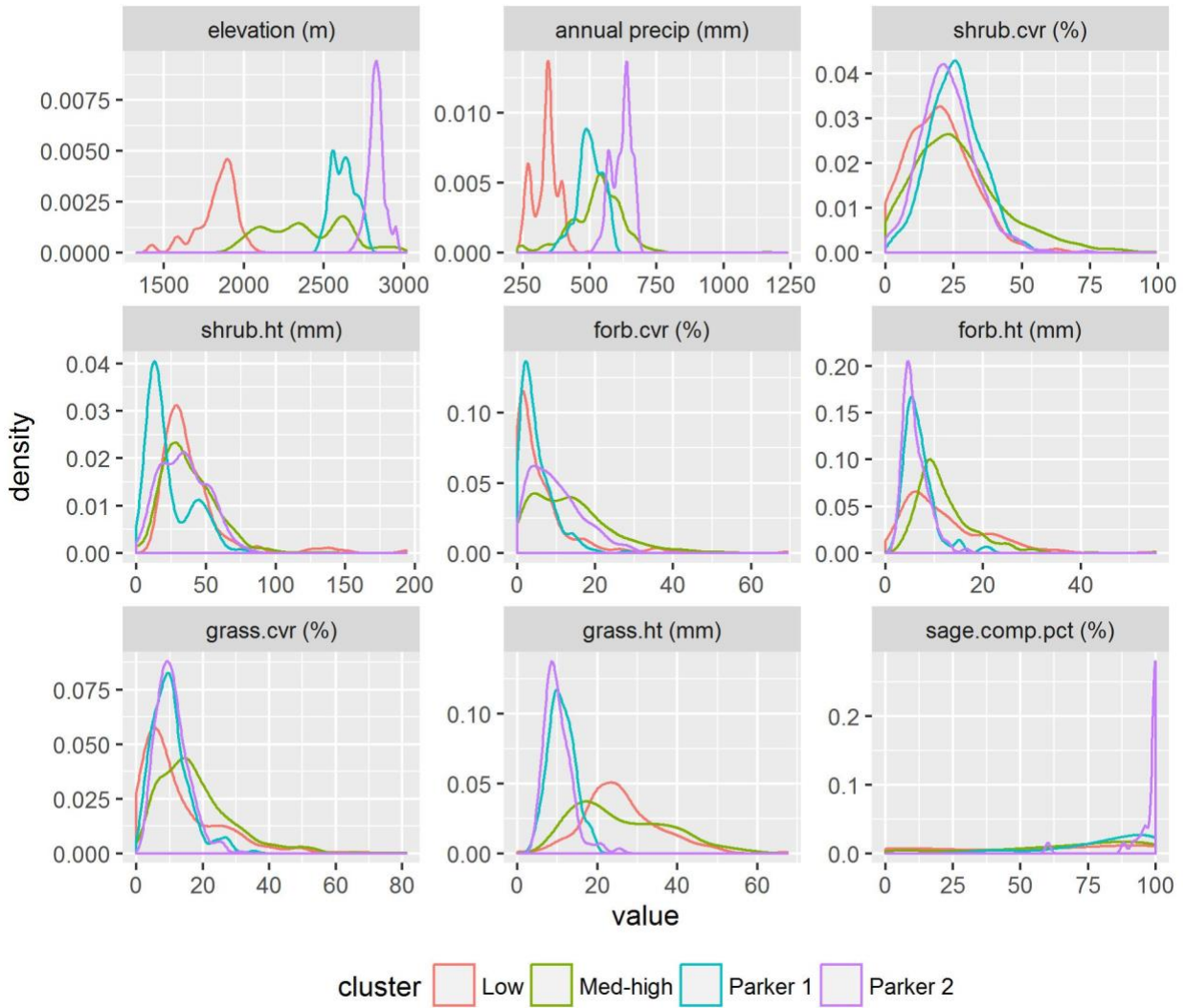
Figure 1. Distributions of variables across k=4 clusters in the brood data.

Legend

SGMAs16d_dissolved

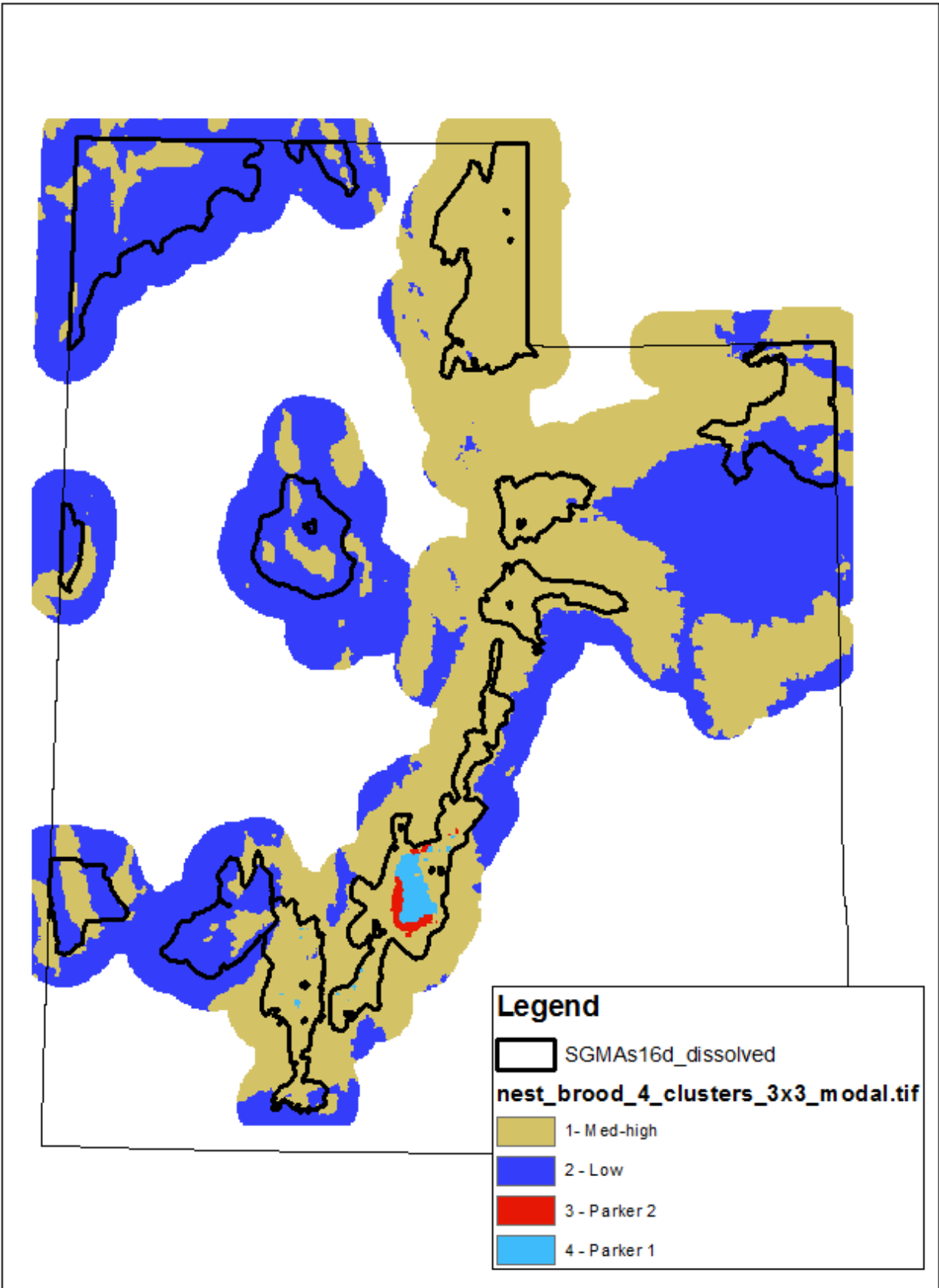**nest_brood_4_clusters_3x3_modal.tif**

1- Med-high

2 - Low

3 - Parker 2

4 - Parker 1

Figure 3. Spatial projection of k = 4 clusters