

## Chapter 4: A Sampling and Analytical Approach to Develop Spatial Distribution Models for Sagebrush-Associated Species

Matthias Leu, Steven E. Hanser, Cameron L. Aldridge, Scott. E. Nielsen, Brian S. Cade, and Steven T. Knick

*Abstract.* Understanding multi-scale floral and faunal responses to human land use is crucial for informing natural resource management and conservation planning. However, our knowledge on how land use influences sagebrush (*Artemisia* spp.) ecosystems is limited primarily to site-specific studies. To fill this void, studies across large regions are needed that address how species are distributed relative to type, extent, and intensity of land use. We present a study design for the Wyoming Basin Ecoregional Assessment (WBEA) to sample sagebrush-associated flora and fauna along a land cover-human land use gradient. To minimize field costs, we sampled various taxonomic groups simultaneously on transects (ungulates and lagomorphs), point counts (song birds), and area-searches of 7.29-ha survey blocks (pellet counts, burrow counts, reptile surveys, medium-sized mammals, ant mounds, rodent trapping, and vegetation sampling of native and exotic plants). We then present an exploratory approach to develop species occurrence and abundance models when *a priori* model building is not an option. Our study design has broad applications for large-scale evaluations of arid ecosystems.

*Key words:* anthropogenic disturbance, data collection, ecoregional assessment, habitat, hierarchical multi-stage modeling, land use, model evaluation, species distribution model.

Ecoregional assessments have become common tools for researchers to evaluate ecosystem health across large extents (Freilich et al. 2001, Groves et al. 2000, The

Nature Conservancy 2000, McMahon et al. 2001, Neely et al. 2001, Noss et al. 2001, Weller et al. 2002, Wisdom et al. 2005). The recognized value of such assessments in addressing the functioning of entire ecosystems has resulted in multiple agency initiatives to conduct landscape-scale assessments, such as the recently developed U.S. Bureau of Land Management Rapid Ecoregional Assessments and U.S. Fish and Wildlife Service Landscape Conservation Cooperatives. Crucial management actions will rest on the guidance provided by ecoregional assessments. However, most input parameters and understanding of habitat or species responses used to develop previous assessments stem from data collected from different spatial and temporal locations or scales and frequently from ecosystems not represented within the assessment region. Responses of species to anthropogenic disturbances and the underlying mechanisms or processes may be applicable across different ecosystems, but the generality of these responses should be evaluated (Lobo et al. 2008). In addition, evaluations are rarely conducted to assess model fit (Freilich et al. 2001) resulting in large uncertainty in the confidence of assessment results and subsequent management recommendations.

We present methods for developing spatial models driven by empirical data allowing for inferences to be made based on relationships directly assessed between species of interest, land cover composition and configuration, abiotic factors, and potential anthropogenic drivers. Complete faunal and floral inventories are logistically difficult and prohibitively costly (for

TABLE 4.1. Distances used to delineate effect zones surrounding anthropogenic features to define the ecological human footprint gradient for the Wyoming Basins Ecoregional Assessment.

Anthropogenic feature	Range of reported empirical distances <sup>a</sup>	Effect zone distance (m)
Agricultural land	≈260 m surrounding pivot fields	135
Communication towers, including associated infrastructure <sup>b</sup>	≈113 m (10 acres, assuming circular shape)	90
Human impact zone	≈610 m	405
Interstate highways	365-1,200 m	855
Irrigation channels	No empirical support	0
Oil/gas wells abandoned/inactive <sup>b</sup>	0.5-1 ha for well pad 0.7 ha/km for roads	90 <sup>c</sup>
Oil/gas wells active, including associated infrastructure <sup>b</sup>	0.5-2 ha for well pad 0.7-2.2 ha/km for roads 3.2 km: Distance avoided by greater sage-grouse	225 <sup>d</sup>
Power lines	300-4,000 m	135
Railroads	0-500 m	135
Secondary roads	100–600 m	135
State/federal highways	100–600 m	405

<sup>a</sup> See Appendix 4.1 for detailed information on effect zone delineation.

<sup>b</sup> Because we only had point locations for these anthropogenic features, we included surface disturbance associated with infrastructure such as roads, condensation tanks (oil and gas wells only), and power lines.

<sup>c</sup> 90 m: 4 cells surrounding center cell (5-cell pattern), area = 4.05 ha.

<sup>d</sup> 225 m: 20 cells surrounding center cell (21-cell pattern), area = 17.01 ha.

discussion see Mac Nally and Fleishman 2004). We therefore developed a sampling design that incorporated data collection across various taxonomic groups, including birds, mammals, reptiles and plants, while minimizing overall sampling costs and ensuring that modeled relationships would be applicable to the entire ecoregion.

We describe the design and analytical approaches developed for the Wyoming Basin Ecoregional Assessment (WBEA) that combined traditional field methods integrated within a Geographical Information System (GIS). We also present an exploratory approach to develop species occurrence and abundance models when *a priori* model building is not an option, and illustrate how these models can be predicted spatially for management purposes and

evaluated for their strengths and weaknesses. Finally, we discuss implications and limitations of our sampling design, providing insights for future ecoregional assessments.

## FIELD SAMPLING METHODS

### Defining the Sampling Space

A challenge in land management is to identify thresholds at which land-use patterns influence the distribution of flora and fauna. This challenge exists because species occurrence and abundance models are often based either on land cover or human land-use gradients but rarely incorporate both (but see e.g., Sawyer et al. 2005, Aldridge and Boyce 2007, Walker et al. 2007, Doherty et al. 2008, Avila-Flores

et al. 2010). To account for potential synergistic species responses to anthropogenic as well as land cover-based drivers, we developed a stratified sampling design across the WBEA according to two gradients: (1) land use, based on a human footprint analysis and (2) land cover, based on Normalized Difference Vegetation Index (NDVI).

#### *Land use: ecological human footprint*

We used 11 anthropogenic features to delineate land use across the WBEA (Table 4.1). We selected these anthropogenic features because they influence species distribution, demography, or both, for one or more species of interest (Appendix 4.1, Leu et al. 2008, Leu and Hanser 2011). We delineated land use based on the ecological human footprint (Leu et al. 2008) represented by a cumulative map of land-use intensity and influence on ecological processes.

We derived the ecological human footprint based on three point features (communication towers, oil/gas wells abandoned/inactive, and oil/gas wells active), six linear features (interstate highways, irrigation channels, power lines, railroads, secondary roads, and state/federal highway), and two polygonal features (agricultural land and human impact zone [industrial areas, urban, exurban, and rural]). For each anthropogenic feature, we delineated its effect zone (the extent at which an anthropogenic feature influences ecological processes) based on a comprehensive literature review to understand the extent of anthropogenic impacts on wildlife and their habitats (Appendix 4.1). We took a conservative approach in delineating effect zones by employing the reported effect distances or areas (Table 4.1) adjusted to fit multiples of the 90-m resolution of our spatial data.

We delineated effect zones for each of 11 anthropogenic features in ArcMap 9.2 (ESRI 2006) by first creating proximity grids for each feature (Euclidian distance). We then used these proximity grids to derive effect zones surrounding anthropo-

genic features based on distances summarized from existing literature (Table 4.1). The resulting map consisted of a binary surface where cells within the effect zone received a value of one, and all other cells were coded as zero. For oil and gas wells, we used two approaches to model effect zones: (1) for abandoned/inactive wells, we used a distance of 90 m from the pixel containing the point location, which resulted in the selection of the four adjacent pixels in the cardinal directions (area = 4.05 ha); and (2) for active wells, we used a distance of 135 m from center point of pixel, which resulted in the selection of eight pixels surrounding the center pixel (area = 7.29 ha). This captured the larger disturbance associated with active wells. Once the effect zones were delineated, we merged the 11 individual anthropogenic layers (maximum cell value = 11) and reclassified this layer to a binary layer with cell values zero or one. We did not incorporate cumulative anthropogenic effects because empirical data to weight individual anthropogenic features were not available. Rather, we focused on whether an area overlapped with the effect zone of at least one anthropogenic feature.

We then put the ecological human footprint in the context of sagebrush (*Artemisia* spp.)-associated vertebrate responses. First, we calculated the relative extent of the ecological human footprint, using moving window analyses (circular shape) on the binary ecological human footprint. Sizes of moving windows were based on seven “model” home ranges that captured published results for 38 of the 40 vertebrate species of concern (Appendix 4.2). We could not find any empirical data on home range size for the Great Basin spadefoot toad (*Scaphiopus intermontanus*) and omitted home range estimates for the spotted bat (*Euderma maculatum*), given the enormous estimated foraging distances of this species (Rabe et al. 1998). Spatial extents used included: 0.8 ha (raw data, no moving window analysis), 2.5 ha (1-cell radius window extent), 41 ha

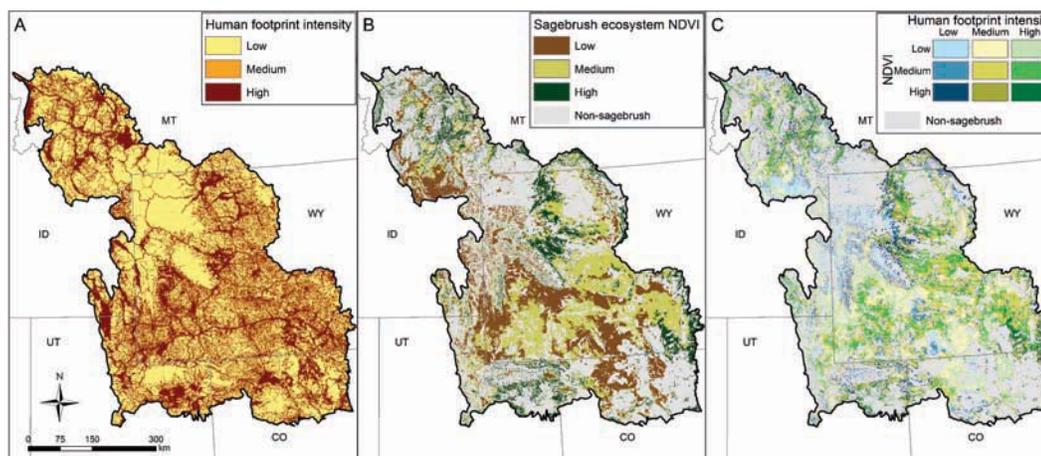


FIG. 4.1. Spatial representation of (A) human footprint intensity, (B) sagebrush ecosystem productivity (NDVI), and (C) sampling matrix (combined human footprint and NDVI gradients) across the Wyoming Basin Ecoregional Assessment area. Human footprint intensity and NDVI were used to stratify sampling locations.

(4-cell radius), 125 ha (7-cell radius), 430 ha (13-cell radius), 2,771 ha (33-cell radius), and 6,361 ha (50-cell radius). Last, we averaged the seven layers to create an ecological human footprint within the average home range of sagebrush-associated vertebrates (Fig. 4.1A).

#### *Sagebrush ecosystem productivity*

The primary land-cover map available for this region in 2004, the “Sagestitch Map” (Comer et al. 2002), did not distinguish sagebrush taxa at the subspecies (variety) level; therefore productivity of sagebrush ecosystems (mesic versus xeric sagebrush ecosystems) could not be differentiated. As a result, we defined sagebrush ecosystem productivity using the Normalized Difference Vegetation Index (NDVI) derived from MODIS (Moderate Resolution Imaging Spectroradiometer, Carroll et al. 2006) classifications from May to August of 2004. We clipped the NDVI layer to the extent of the combined shrub-grassland land cover identified in the “Sagestitch Map” (Comer et al. 2002) (Fig. 4.1B).

#### *Sampling design spatial data set*

We allocated equal sampling effort across gradients of the ecological human

footprint and NDVI by using a 3 x 3 matrix. We reclassified the mean ecological human footprint value within a 33-cell radius according to three ordinal categories containing equal areas ranging from low (0–0.20), moderate (>0.20–0.38), to high (>0.38–1). The 33-cell radius dataset was used to facilitate placement of sample locations by generalizing the ecological human footprint over a broader area than the surface created from the average home range size. Similarly, we reclassified the NDVI layer into three ordinal categories of equal area ranging from low (-1–0.37), moderate (>0.37–0.53), to high (>0.53–1). We combined the reclassified gradients spatially to produce a spatial data set consisting of nine sampling strata (Fig. 4.1C).

#### **Sampling Location Selection**

We used a hierarchical-spatial sampling design to survey flora and fauna across the WBEA area (Ch. 2) during spring/summer of 2005 and 2006. We restricted our surveys to WBEA areas consisting of shrub-grassland land cover within Wyoming and Colorado, given the focus of the assessment on the sagebrush ecosystem. To increase sampling efficiency, we first randomly placed 49 non-overlapping circles of 30-km radius

throughout the WBEA within Wyoming in 2005 (29 circles), and Wyoming and Colorado in 2006 (20 circles). We selected center locations of circles using the RANDOM POINT GENERATOR in ARCVIEW (Version 1.1, Utah State University). We populated the area within each 30-km circle overlapping the combined gradients of the ecological human footprint and shrub-grassland land cover productivity (i.e., area covered by nine sampling stratum of the 3 x 3 matrix) with as many random points (1-km apart) we could fit. We restricted potential random points within each circle to areas with <25% slope, based on 90-m Digital Elevation Models (DEM; National Elevation Dataset, USGS EROS, <http://seamless.usgs.gov/>), such that observers were able to walk to random points while collecting data. These random points represented the center of two types of points in relation to roads (Fig. 4.1): near-road = 0-750 m from road, and far-road = >750-3,000 m from road. We then selected a third set of on-road points using COSTPATH in ARC/INFO (ESRI 2006) (Fig. 4.1). These on-road points were located at the road end of the least-cost path in terms of pixel-based elevation change (using DEM) between the far-road points and the road network.

We then selected a preliminary set of points from this pool to ensure equal replication within each of nine sampling stratum; consequently, not all 30-km circles contained the same number of points because the area covered by each of nine sampling stratum varied among 30-km circles. In the field, we first attempted to sample the original set of points. However, this was not always possible due to access issues (mainly private land). In such cases, we selected the next nearest point within the same disturbance-productivity class. We were unable to get access to replacement points in some 30-km circles, resulting in slightly unbalanced sampling across strata and in relation to roads (n = 330; 162 in 2005 and 168 in 2006; on-road n = 104, near-road n = 125, far-road = 101). Nearest

neighbor distance among all points averaged 2.36 km (SD  $\pm$  2.27 km, range = 0.69–19.6 km), among far-road points averaged 4.98 km (SD  $\pm$  3.15 km, range = 1.20–19.6 km), and among on-road and near-road points averaged 4.82 km (SD  $\pm$  3.07 km, range = 1.20–20.79 km) apart. Selected points were converted to 270 m x 270 m survey blocks (7.29 ha) centered on points and oriented on cardinal axes, with corners facing northeast, southeast, southwest, and northwest (Fig. 4.2).

We surveyed larger-sized vertebrates on 145 transects that extended between the center points of survey blocks (Figs. 4.1 and 4.2). The combined transect/survey block sampling design allowed us to sample multiple vertebrate species, thereby decreasing travel time and sampling cost. We used two types of transects: (1) short transects, starting at roads (mainly gravel roads), and ending at centers of near-road survey blocks; and (2) long transects, starting at centers of on-road survey blocks and ending at centers of paired far-road survey blocks. Transects between on-road and far-road survey blocks were identified by the least-cost path used to select the on-road survey blocks. Least-cost paths were also developed between near-road survey blocks and the closest point on the road using the same analysis procedure. For field application, transects were converted from the COSTPATH raster output into line shapefiles and uploaded into GPS units (Garmin E-trex Venture) using the Minnesota Department of Natural Resources Garmin software (Version 4.41, <http://www.dnr.state.mn.us/mis/gis/tools/arcview/extensions/DNRGarmin/DNRGarmin.html>) to aid field navigation. We recorded track logs of altered transects for subsequent sampling if observers deviated from predetermined transects due to obstacles encountered during the first sampling bout of the season.

#### **Floral and Faunal Sampling Protocol**

Our surveys incorporated multiple techniques designed to detect the full suite of

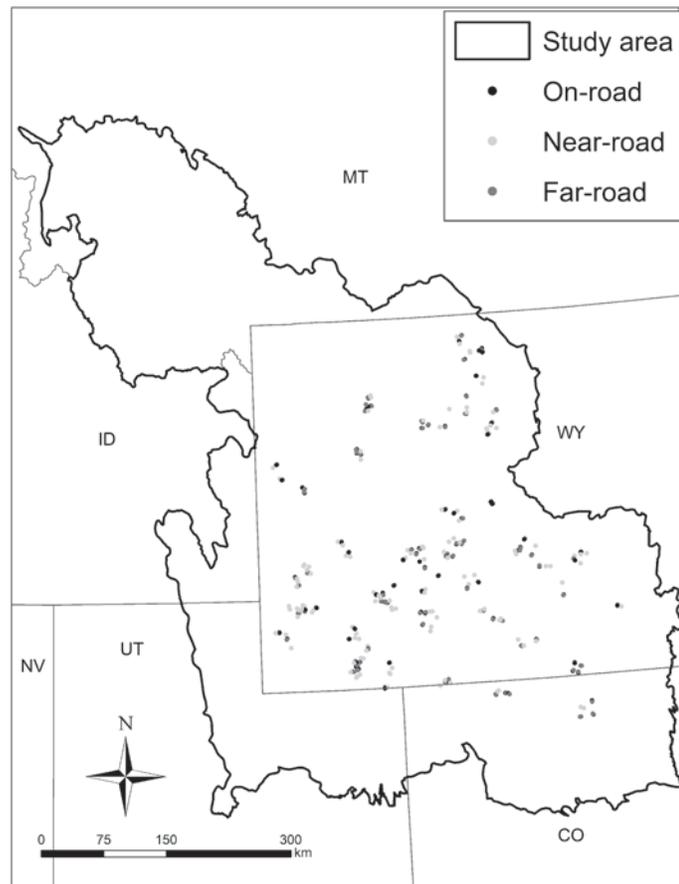


FIG. 4.2. Distribution of survey blocks and transect across the Wyoming Basin Ecoregional Assessment area. Shown are locations for survey block for on-road = directly adjacent to road ( $n = 104$ ), near-road = 0-750 m ( $n = 125$ ), and far-road = >750-3,000 m ( $n = 101$ ). Transects ( $n = 145$ ), not shown, occur between near-road and far-road survey blocks ( $n = 101$ ) and between roads and near-road survey blocks ( $n = 44$ ; transects > 100-m long) (see Fig. 4.3).

sagebrush steppe-associated fauna as well as information on plant community composition. Our survey protocols were applied as follows: (1) surveys conducted on transects while navigating between the on-road and far-road survey blocks or roads and near-road survey blocks and (2) surveys conducted within each survey block (Fig. 4.3).

On short and long transects, we applied distance sampling (Buckland et al. 2001, 2004) to enable density estimation for medium to large-sized mammals. For each detected individual or group, we recorded location of observer (latitude and longitude), azimuth using a compass, and

distance between observer and object using a rangefinder (Bushnell Yardage Pro Legend).

Within survey blocks, we used variable-width point counts (Bibby et al. 1992) centered on survey blocks (Fig. 4.3) to survey sagebrush-associated songbirds. We estimated distance between observer and birds using a rangefinder. We used area-searches based on within survey block transects of 2.16-km length (Fig 4.3) to survey medium-sized mammals (lagomorphs and larger rodents), pygmy rabbit (*Brachylagus idahoensis*) burrows, reptiles, ant mounds, and greater sage-grouse (*Centrocercus urophasianus*) pellets. We sur-

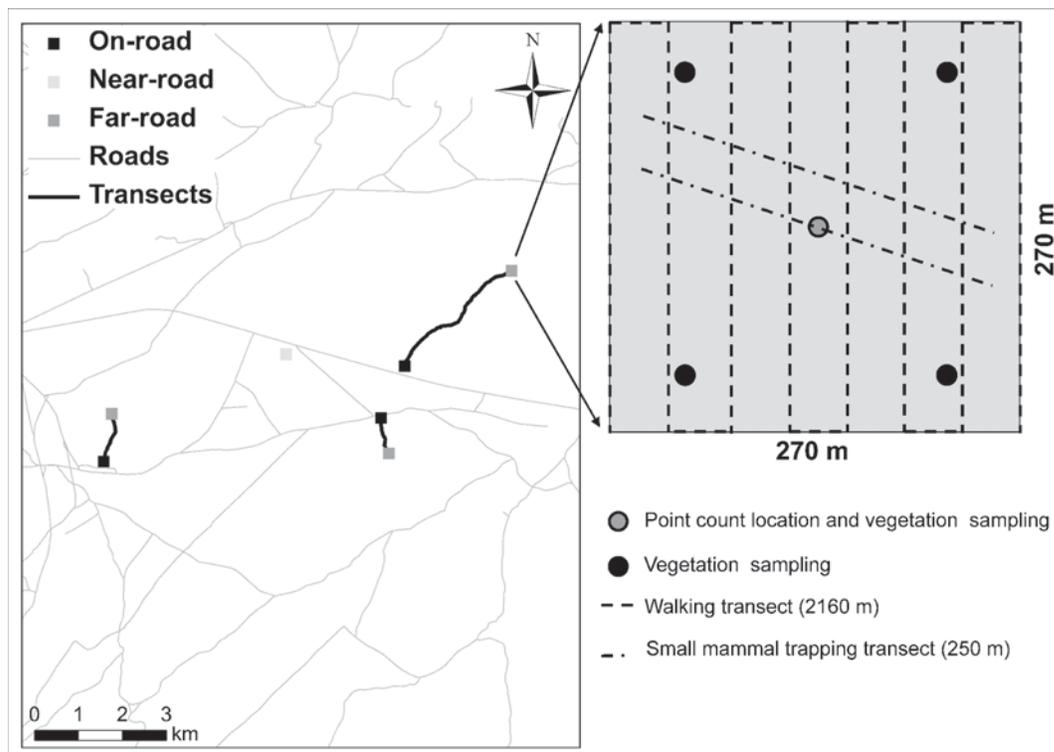


FIG. 4.3. Sampling layout within a survey block. Survey blocks were quadratic in shape with sides measuring 0.27 km. Points were used to survey vegetation ( $n = 5$ ), with the center point used as songbird point count location. We used walking transects (2.16 km) to survey medium-sized mammals (grounds squirrels, prairie dogs, and chipmunks, lagomorphs), reptiles, and greater sage-grouse pellets. We surveyed small mammal diversity along two 0.25-km long transects (50 traps total); direction of transects was chosen randomly and transects were spaced 15 m apart.

veyed plant communities (shrub cover and composition, selected exotic forb and grass cover and composition, native annual and perennial forb and grass cover, and shrub height) in five 20-m radius ( $1,257 \text{ m}^2$ ) plots systematically located in the survey block at the center and 127.3 m from the center at  $45^\circ$ ,  $135^\circ$ ,  $225^\circ$ , and  $315^\circ$  azimuths (Fig. 4.3). For exotic plants, we sampled a subset of plant species deemed noxious and invasive by land management agencies (Appendix 4.3). We trapped small mammals at a subset of survey blocks using two parallel 0.25-km long transects centered on the survey block, but oriented randomly (Fig. 4.3). Detailed descriptions of specific sampling protocols are provided in chapters that follow.

We combined surveys throughout the field season to maximize sampling efficiency and minimize cost. Three field crews (two observers per team) worked independently throughout the field season. During the first round of surveys from 28 April – 31 May, all crews sampled medium to large-sized mammals on transects en route to survey blocks. Within survey blocks, crews sampled songbirds, pygmy rabbit signs, ant mounds, and medium-sized mammals (grounds squirrels [*Spermophilus* spp.], prairie dogs [*Cynomys* spp.], and chipmunks [*Tamias* spp.]). During the second round of sampling from 1 June – 2 July, all crews again sampled medium to large-sized mammals on transects en route to survey blocks; on survey

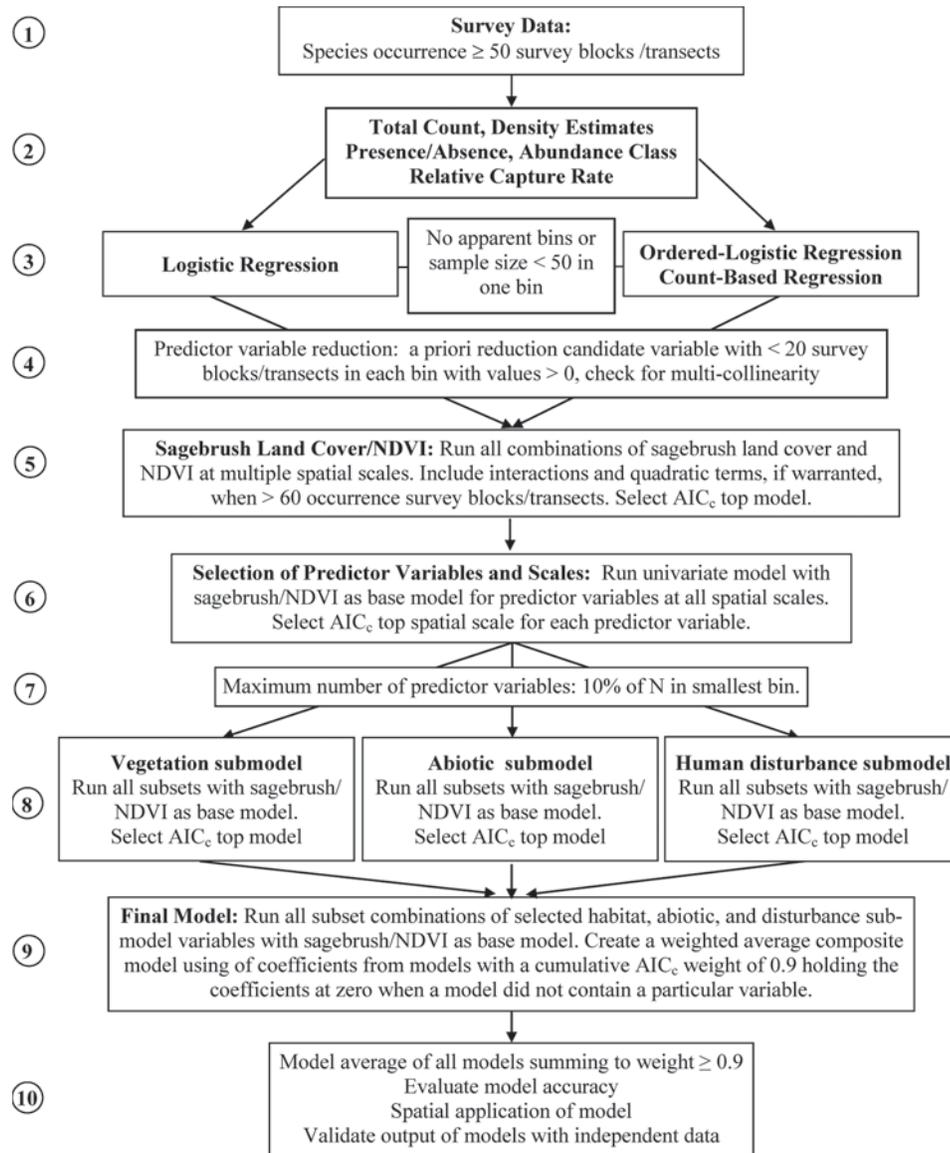


FIG. 4.4. Flow chart outlining hierarchical multi-stage modeling approach for floral and faunal presence/absence and abundance data.

blocks, song birds, and vegetation (species specific shrub and tree cover and height, exotic and native herbaceous cover, and ground cover) were sampled. During the last round of sampling between 6 July and 2 September, we only sampled on survey blocks. Crew one counted reptiles, mammals, and sage-grouse pellets; crew two measured vegetation and habitat char-

acteristics (shrub cover, total, sagebrush [live, woody, and total], exotic and native herbaceous cover, dominant species by cover type, rock out-crop cover, and ground cover); and crew three trapped small mammals on a subset of survey blocks. We assigned field crews to sample the various taxonomic groups based on individual expertise.

## ANALYSES

We developed species occurrence and abundance models based on habitat, abiotic and land use predictor variables (Franklin 2009). Our modeling procedure followed an Akaike Information Criterion (AIC) approach (Burnham and Anderson 2002); however, for most species we could not develop *a priori* candidate models because we lacked knowledge about species-specific responses to land use as well as appropriate spatial extents for assessing land cover conditions. As a result, our modeling effort was exploratory and followed a hierarchical analysis based on multiple steps to select the most plausible final models (Fig 4.4). We first selected the best extent and form of variables of interest and then chose top variables among competing variables within categories, of influence (Fig. 4.4). We used empirical information and/or our own knowledge to guide selection of predictor variables whenever possible but ultimately used AIC corrected for small sample sizes (AIC<sub>c</sub>; Burnham and Anderson 2002) to select among competing predictor variables. Once predictor variables were selected within categories, we used all possible variable combinations within and across categories to develop candidate models. We used AIC<sub>c</sub> to rank these models, produced a final model-averaged composite model based on a 90% confidence model set, and used independent data when possible to evaluate predictive capacity of final models. For all species, we modeled species presence/absence, abundance, or density, as summarized on survey blocks or transects, using a set of predictor variables consisting of a variety of environmental, habitat, and land-use covariates. Below, we outline detailed analytical approaches that apply to Chapters 5-9; methods used in Chapter 10 (exotic plants) deviate from this approach and are detailed in that chapter.

### Predictor Variables

We used a suite of common GIS predictor variables consisting of land cover mea-

sured at different radii, land cover patch metrics, vegetation productivity, soil characteristics, terrain-derived variables, distance from water, climate, and density of and distance from anthropogenic features. Little is known about how sagebrush-associated species perceive ecological patterns. Therefore, we explored landscape perception of these species by selecting a range of circular moving window sizes based on the radius of seven model home range sizes that best represented 38 sagebrush steppe-associated species (Appendix 4.2). We evaluated land cover, vegetation productivity, and terrain-derived variables at six radii (0.27, 0.54, 1, 3, 5, and 18 km) and landscape metrics (contagion, edge density, mean patch size) at three radii (1, 3, 5 km). The 18-km radius reflected the recommended scale for habitat management around lek locations of migratory greater sage-grouse populations (Connelly et al. 2000). All predictor variable data sets are available on the SAGEMAP website (<http://sagemap.wr.usgs.gov/wbea.aspx>).

We modeled distance variables using exponential distance decay functions (value =  $e^{(\text{Euclidean distance to feature (km)}/\text{distance parameter})}$ ) with the distance parameter set at 0.25, 0.5, and 1 km (Nielsen et al. 2009, Carpenter et al. 2010), allowing for nonlinear responses of species to distance from water sources or anthropogenic features. For anthropogenic features such as power lines that attract synanthropic predators (predators that benefit from human features [Johnston 2001]), the asymptote of the 1-km distance decay function (~4.5 km) approximates the maximum home range size ( $\approx 54 \text{ km}^2$ ) for golden eagles (*Aquila chrysaetos*) breeding in the Intermountain West (Kochert et al. 2002); the asymptote of the 0.5-km distance decay function (~2.4 km) approximates the mean home range size (22.8 km<sup>2</sup>) for golden eagles in southwest Idaho (Marzluff et al. 1997); and the asymptote of the 0.25-km distance decay function (~1.2 km) approximates the mean common raven (*Corvus corax*) feeding dis-

tance around nests in arid regions of California ( $0.57 \text{ km} \pm 0.71 \text{ SD}$  [Boarman and Heinrich 1999]).

We initially identified a total of 154 candidate predictor variables likely to influence species occurrence and abundance; inclusion or exclusion of specific predictor variables are discussed in each chapter separately. We screened candidate predictor variables for sufficient representation of non-zero data values (i.e., values  $>0$ ) across survey blocks and extents to avoid model fitting based on predictor variables dominated by zeros or having non-zero data values only at large extents. As a cut-off point, we only included predictor variables with non-zero data values on at least 20 survey blocks (6%,  $n = 326$ ) or transects (14%,  $n = 141$ ) at the smallest radius of 0.27-km. We omitted three land-cover variables (agriculture,  $n = 4$  survey blocks with values  $> 0$  [retained as a distance predictor variable]; juniper [*Juniperus* spp.],  $n = 2$ ; and mountain shrub,  $n = 19$ ) and one anthropogenic variable (oil and gas wells,  $n = 5$  [retained as a distance predictor variable]). For distance to anthropogenic feature, we selected predictor variables with at least 20 survey blocks or transects located  $\leq 1$  km from a feature. We omitted three predictor variables as a result, including human populated area ( $n = 4$  survey blocks within 1 km of feature), railroad ( $n = 2$ ), and tower ( $n = 2$ ). We were left with a total of 122 candidate predictor variables after this screening (Table 4.2).

Fifty-four of 122 candidate predictor variables consisted of nine land cover types (Table 4.2) evaluated at the six radii (0.27, 0.54, 1, 3, 5, and 18 km); these included four sagebrush land cover classes as well as coniferous forest (CFRST), grassland (GRASS), mixed shrubland (MIX), riparian (RIP), and salt-desert shrubland (SALT) land cover (Table 4.2). We used the LANDFIRE existing vegetation type (EVT) data layer (LANDFIRE 2007), reclassified per the cross-walk listed in Ap-

pendix 1.1, as our base land cover map and moving window analyses in ArcMap 9.2 (ESRI 2006) to calculate proportion of land cover class for each of six radii. For the all sagebrush species (ALLSAGE) land cover (all sagebrush species and subspecies combined), we calculated landscape metrics in FragStats (McGarigal et al. 2002) including patch size (PATCH), edge density (EDGE), and contagion (CONTAG), at three radii (1, 3, and 5 km).

We determined land cover productivity values for the plot center and computed mean values at six spatial extents (Table 4.2). Land cover productivity values were calculated for each pixel on the landscape using the maximum value of NDVI from all available data during the growing season (May through August in 2005 and 2006). NDVI values were derived from the 250-m resolution Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery (Carroll et al. 2006) re-sampled to 90-m resolution using cubic convolution for interpolation in ArcMap 9.2 (ESRI 2006).

We derived 18 abiotic variables (Table 4.2) based on terrain, soil, climate, and hydrography. Terrain variables were derived from 90-m DEM and consisted of compound topographic index (CTI) (Gessler et al. 1995), elevation (ELEV), slope (SLOPE), solar radiation index (SOLAR, developed using HILLSHADE analysis with parameters set to solar angle and direction at noon on the summer solstice, ESRI 2006), and topographic ruggedness index (TRI) (Riley et al. 1999). We derived TRI across the six radii in addition to the plot center. For soil variables, we used the conterminous United States multilayer soil characteristics dataset (Miller and White 1998) to develop spatial datasets for acidity (pH), available water capacity (AWC), bulk density (BULKd), clay content (CLAY), depth (SOIL<sub>cm</sub>), salinity (SALIN), sand content (SAND), and silt content (SILT). For climate variables, we used climate normals from Parame-

TABLE 4.2. Descriptive statistics (mean, standard error, and range) for 38 independent candidate variables and associated spatial extents ( $n = 122$  variables) measured on 330 survey blocks in the Wyoming Basins Ecoregional Assessment area. We used a subset of candidate predictive variables in the species models introduced in Chapters 5–10.

Category	Description	Radius/distance parameter (km)	Variable	Unit	$\bar{x}$	SE	Min	Max
<b>Vegetation</b> Land cover	All big sagebrush (Intermountain basins big sagebrush shrubland, Intermountain basins big sagebrush steppe, Intermountain basins montane sagebrush steppe, and <i>Artemisia tridentata</i> ssp. <i>vaseyana</i> shrubland alliance) <sup>a</sup>  All sagebrush species (All big sagebrush ecological systems, plus Colorado Plateau mixed low sagebrush shrubland, Columbia Plateau low sagebrush steppe, Wyoming basins dwarf sagebrush shrubland and steppe) <sup>a</sup>  Big sagebrush (Intermountain basins big sagebrush shrubland, and Intermountain basins big sagebrush steppe) <sup>a</sup>  Mountain sagebrush (Intermountain basins montane sagebrush steppe and <i>A. t.</i> spp. <i>vaseyana</i> shrubland alliance) <sup>a</sup>	0.27	ABIGSAGE <sub>270</sub>	Proportion	0.75	0.02	0.00	1.00
		0.54	ABIGSAGE <sub>540</sub>	Proportion	0.73	0.01	0.00	1.00
		1	ABIGSAGE <sub>1km</sub>	Proportion	0.72	0.01	0.01	1.00
		3	ABIGSAGE <sub>3km</sub>	Proportion	0.69	0.01	0.03	0.99
		5	ABIGSAGE <sub>5km</sub>	Proportion	0.67	0.01	0.03	0.98
		18	ABIGSAGE <sub>18km</sub>	Proportion	0.62	0.01	0.10	0.90
		0.27	ALLSAGE <sub>270</sub>	Proportion	0.77	0.02	0.00	1.00
		0.54	ALLSAGE <sub>540</sub>	Proportion	0.75	0.01	0.00	1.00
		1	ALLSAGE <sub>1km</sub>	Proportion	0.74	0.01	0.01	1.00
		3	ALLSAGE <sub>3km</sub>	Proportion	0.71	0.01	0.03	0.99
		5	ALLSAGE <sub>5km</sub>	Proportion	0.69	0.01	0.03	0.99
		18	ALLSAGE <sub>18km</sub>	Proportion	0.64	0.01	0.11	0.93
		0.27	BIGSAGE <sub>270</sub>	Proportion	0.59	0.02	0.00	1.00
		0.54	BIGSAGE <sub>540</sub>	Proportion	0.58	0.02	0.00	1.00
		1	BIGSAGE <sub>1km</sub>	Proportion	0.58	0.02	0.00	1.00
		3	BIGSAGE <sub>3km</sub>	Proportion	0.56	0.02	0.00	0.97
		5	BIGSAGE <sub>5km</sub>	Proportion	0.55	0.02	0.00	0.94
		18	BIGSAGE <sub>18km</sub>	Proportion	0.51	0.01	0.00	0.87
0.27	MTNSAGE <sub>270</sub>	Proportion	0.16	0.02	0.00	1.00		
0.54	MTNSAGE <sub>540</sub>	Proportion	0.15	0.02	0.00	1.00		
1	MTNSAGE <sub>1km</sub>	Proportion	0.15	0.01	0.00	0.95		
3	MTNSAGE <sub>3km</sub>	Proportion	0.13	0.01	0.00	0.91		
5	MTNSAGE <sub>5km</sub>	Proportion	0.13	0.01	0.00	0.87		
18	MTNSAGE <sub>18km</sub>	Proportion	0.11	0.01	0.00	0.48		

TABLE 4.2. Continued

Category	Description	Radius/distance parameter (km)	Variable	Unit	$\bar{x}$	SE	Min	Max
Coniferous forest <sup>a</sup>		0.27	CFRST <sub>270</sub>	Proportion	0.02	0.01	0.00	0.79
		0.54	CFRST <sub>540</sub>	Proportion	0.03	0.01	0.00	0.72
		1	CFRST <sub>1km</sub>	Proportion	0.04	0.01	0.00	0.74
		3	CFRST <sub>3km</sub>	Proportion	0.05	0.01	0.00	0.73
		5	CFRST <sub>5km</sub>	Proportion	0.05	0.01	0.00	0.71
		18	CFRST <sub>18km</sub>	Proportion	0.08	0.01	0.00	0.53
		0.27	GRASS <sub>270</sub>	Proportion	0.05	0.01	0.00	0.97
		0.54	GRASS <sub>540</sub>	Proportion	0.05	0.01	0.00	0.84
		1	GRASS <sub>1km</sub>	Proportion	0.05	0.01	0.00	0.74
		3	GRASS <sub>3km</sub>	Proportion	0.04	<0.01	0.00	0.61
Mixed shrubland <sup>a</sup>		5	GRASS <sub>5km</sub>	Proportion	0.04	<0.01	0.00	0.43
		18	GRASS <sub>18km</sub>	Proportion	0.04	<0.01	0.00	0.15
		0.27	MIX <sub>270</sub>	Proportion	0.00	<0.01	0.00	0.28
		0.54	MIX <sub>540</sub>	Proportion	0.01	<0.01	0.00	0.24
		1	MIX <sub>1km</sub>	Proportion	0.01	<0.01	0.00	0.12
		3	MIX <sub>3km</sub>	Proportion	0.01	<0.01	0.00	0.08
		5	MIX <sub>5km</sub>	Proportion	0.01	<0.01	0.00	0.06
		18	MIX <sub>18km</sub>	Proportion	0.01	<0.01	0.00	0.04
		0.27	RIP <sub>270</sub>	Proportion	0.03	<0.01	0.00	0.79
		0.54	RIP <sub>540</sub>	Proportion	0.03	<0.01	0.00	0.57
Riparian <sup>a</sup>		1	RIP <sub>1km</sub>	Proportion	0.03	<0.01	0.00	0.34
		3	RIP <sub>3km</sub>	Proportion	0.04	<0.01	0.00	0.26
		5	RIP <sub>5km</sub>	Proportion	0.04	<0.01	0.00	0.19
		18	RIP <sub>18km</sub>	Proportion	0.04	<0.01	0.00	0.12

TABLE 4.2. Continued

Category	Description	Radius/distance parameter (km)	Variable	Unit	$\bar{x}$	SE	Min	Max
	Salt-desert shrubland <sup>a</sup>	0.27	SALT <sub>270</sub>	Proportion	0.05	0.01	0.00	0.83
		0.54	SALT <sub>540</sub>	Proportion	0.05	0.01	0.00	0.76
		1	SALT <sub>1km</sub>	Proportion	0.05	0.01	0.00	0.69
		3	SALT <sub>3km</sub>	Proportion	0.05	0.01	0.00	0.58
		5	SALT <sub>5km</sub>	Proportion	0.06	0.01	0.00	0.58
		18	SALT <sub>18km</sub>	Proportion	0.06	0.01	0.00	0.49
	All sagebrush species contagion	1	CONTAG <sub>1km</sub>	%	39.51	1.45	0.55	97.64
		3	CONTAG <sub>3km</sub>	%	36.15	1.31	2.01	97.90
		5	CONTAG <sub>5km</sub>	%	29.61	1.00	3.27	91.38
		1	EDGE <sub>1km</sub>	m/ha	41.92	1.20	0.88	91.60
		3	EDGE <sub>3km</sub>	m/ha	42.85	1.20	0.00	84.20
	All sagebrush species edge density	5	EDGE <sub>5km</sub>	m/ha	45.98	0.89	3.84	80.10
		1	PATCH <sub>1km</sub>	m <sup>2</sup>	178.27	6.24	0.81	304.56
		3	PATCH <sub>3km</sub>	m <sup>2</sup>	718.45	47.66	1.55	2,745.09
	All sagebrush species mean patch size	5	PATCH <sub>5km</sub>	m <sup>2</sup>	1,011.6	119.1	1.8	9,866.6
		Plot center	NDVI	Cell value	0.32	0.01	0.13	0.76
		0.27	NDVI <sub>270</sub>	Mean value	0.32	0.01	0.14	0.76
	Normalized Difference Vegetation Index	0.54	NDVI <sub>540</sub>	Mean value	0.32	0.01	0.15	0.75
		1	NDVI <sub>1km</sub>	Mean value	0.32	0.01	0.17	0.76
		3	NDVI <sub>3km</sub>	Mean value	0.33	0.01	0.18	0.77
		5	NDVI <sub>5km</sub>	Mean value	0.34	0.01	0.19	0.76
		18	NDVI <sub>18km</sub>	Mean value	0.35	0.01	0.20	0.74
	Compound topographic index	Plot center	CTI	Value	8.87	0.12	4.96	19.64

TABLE 4.2. Continued

Category	Description	Radius/distance parameter (km)	Variable	Unit	$\bar{x}$	SE	Min	Max	
Terrain	Elevation	Plot center	ELEV	m	2,102	18.5	1,286	3,161	
	Slope	Plot center	SLOPE	Deg	4.27	0.27	0.00	32.15	
	Solar radiation index	Plot center	SOLAR	Value	148.52	0.90	76.00	226.00	
	Topographic ruggedness index		Plot center	TRI	Cell value	20.78	1.21	0.00	149.47
			0.27	TRI <sub>270</sub>	Mean value	21.40	1.08	0.00	114.50
			0.54	TRI <sub>540</sub>	Mean value	22.17	1.02	0.59	94.64
		1	TRI <sub>1km</sub>	Mean value	22.65	0.98	2.12	96.76	
<b>Abiotic</b>	Acidity	3	TRI <sub>3km</sub>	Mean value	23.87	0.94	2.18	82.26	
		5	TRI <sub>5km</sub>	Mean value	24.03	0.93	2.63	86.94	
		18	TRI <sub>18km</sub>	Mean value	25.91	0.92	5.98	95.79	
		Plot center	pH	Value	6.73	0.04	2.87	8.74	
		Plot center	AWC	inches/inch	5.18	0.09	1.46	9.16	
		Plot center	BULKd	g/cc	1.53	0.01	1.21	2.19	
Soil	Available water capacity	Plot center	CLAY	%	16.51	0.39	0.00	47.00	
		Plot center	SOIL <sub>cm</sub>	cm	100.90	1.58	38.00	152.00	
<b>Abiotic</b>	Salinity	Plot center	SALIN	mmhos/cm	2.28	0.09	0.00	9.53	
		Plot center	SAND	%	39.14	0.81	0.00	88.25	
		Plot center	SILT	%	26.70	0.54	0.00	58.38	
		Plot center	Tmax	Deg C	12.24	0.12	4.49	16.46	
Climate	Mean annual maximum temperature	Plot center	Tmin	Deg C	-2.95	0.11	-7.37	1.22	
		Plot center	PRECIP	cm	33.72	0.70	17.07	80.74	

TABLE 4.2. Continued

Category	Description	Radius/distance parameter (km)	Variable	Unit	$\bar{x}$	SE	Min	Max
<b>Abiotic</b> Water Sources	Decay distance from intermittent water	0.25	iH2Od <sub>250</sub>	Probability	0.22	0.016	0.00	1.00
		0.5	iH2Od <sub>500</sub>	Probability	0.36	0.017	0.00	1.00
	Decay distance from permanent water	1	iH2Od <sub>1km</sub>	Probability	0.53	0.016	0.02	1.00
		0.25	pH2Od <sub>250</sub>	Probability	0.05	0.009	0.00	1.00
		0.5	pH2Od <sub>500</sub>	Probability	0.11	0.011	0.00	1.00
		1	pH2Od <sub>1km</sub>	Probability	0.20	0.014	0.00	1.00
<b>Disturbance</b> Distance	Decay distance from agricultural land	0.25	AG <sub>250</sub>	Probability	0.02	<0.01	0.00	0.45
		0.5	AG <sub>500</sub>	Probability	0.06	0.01	0.00	0.67
	Decay distance from interstate highways, federal and state highways	1	AG <sub>1km</sub>	Probability	0.13	0.01	0.00	0.82
		0.25	MjRD <sub>250</sub>	Probability	0.04	0.01	0.00	1.00
		0.5	MjRD <sub>500</sub>	Probability	0.08	0.01	0.00	1.00
		1	MjRD <sub>1km</sub>	Probability	0.13	0.01	0.00	1.00
	Decay distance from pipeline	0.25	PIPE <sub>250</sub>	Probability	0.06	0.01	0.00	1.00
		0.5	PIPE <sub>500</sub>	Probability	0.10	0.01	0.00	1.00
	Decay distance from power line	1	PIPE <sub>1km</sub>	Probability	0.15	0.02	0.00	1.00
		0.25	POWER <sub>250</sub>	Probability	0.04	0.01	0.00	1.00
		0.5	POWER <sub>500</sub>	Probability	0.06	0.01	0.00	1.00
		1	POWER <sub>1km</sub>	Probability	0.12	0.01	0.00	1.00
Decay distance from secondary roads	0.25	2RD <sub>250</sub>	Probability	0.41	0.02	0.00	1.00	
	0.5	2RD <sub>500</sub>	Probability	0.54	0.02	0.01	1.00	
	1	2RD <sub>1km</sub>	Probability	0.69	0.02	0.08	1.00	
Decay distance from oil and gas wells	0.25	WELL <sub>250</sub>	Probability	0.01	<0.01	0.00	0.70	
	0.5	WELL <sub>500</sub>	Probability	0.03	0.01	0.00	0.84	
	1	WELL <sub>1km</sub>	Probability	0.07	0.01	0.00	0.91	

TABLE 4.2. Continued

Category	Description	Radius/distance parameter (km)	Variable	Unit	$\bar{x}$	SE	Min	Max
<b>Disturbance</b> Density	Density of all roads (interstate highways, federal and state highways, and secondary roads)	0.27	RDdens <sub>270</sub>	km/km <sup>2</sup>	1.78	0.10	0.00	9.54
		0.54	RDdens <sub>540</sub>	km/km <sup>2</sup>	1.44	0.07	0.00	7.59
		1	RDdens <sub>1km</sub>	km/km <sup>2</sup>	1.28	0.06	0.00	6.24
		3	RDdens <sub>3km</sub>	km/km <sup>2</sup>	1.43	0.04	0.07	5.03
		5	RDdens <sub>5km</sub>	km/km <sup>2</sup>	1.43	0.03	0.35	4.19
		18	RDdens <sub>18km</sub>	km/km <sup>2</sup>	1.44	0.02	0.32	2.31

<sup>a</sup> Ecological systems reclassified from the LANDFIRE existing vegetation type data set (LANDFIRE 2007); see Ch. 1 for details.

ter-Elevation Regressions on Independent Slopes Model (PRISM) to estimate mean annual precipitation (PRECIP; PRISM Group 2006a), maximum temperature (Tmax; PRISM Group 2006b), and minimum temperature (Tmin; PRISM Group 2006c). Last, we developed hydrographic variables based on distance to perennial (pH2Od) and intermittent (iH2Od) water sources; as with other distance-based variables, we used exponential distance decay functions fit to 0.25-km, 0.50-km, and 1-km distance parameters.

We included seven anthropogenic feature types in our analyses. Spatial data sets for anthropogenic features were clipped from input data used to create the human footprint of the western U.S. (Leu et al. 2008) and updated with recent spatial data sets (see metadata for detailed information on data acquisition). We derived 18 anthropogenic proximity variables (Table 4.2) based on six anthropogenic features (agriculture [AG], interstate and state/federal highways [MjRD], pipelines [PIPE], power lines [POWER], secondary roads [2RD], and oil-gas wells as of August 2005 [WELL]) and exponential distance decay functions fit with three distance parameters (0.25 km, 0.50 km, 1 km). We also developed a road density (RDdens) (interstate highways, federal and state highways, and secondary roads combined) spatial data set evaluated at the six radii.

## Modeling Approach

### Step 1 – Candidate species selection

Our goal at the onset of this study was to develop occurrence or abundance models for all species surveyed during the breeding seasons of 2005 and 2006. However, many species were rare or difficult to detect (Ch. 5–10). We restricted development of models to species with occurrences on at least 50 survey blocks or transects (Fig. 4.4) because sample sizes below this threshold result in regression models with poor predictive capabilities (Coudun and

Gégout 2006). Only 43.2% ( $n = 37$  species) of all species sampled in our study were detected on  $>50$  survey blocks and only 10.0% ( $n = 10$ ) on  $>50$  transects. We present a complete list of species sampled on the 330 survey blocks or the 145 transects in following chapters.

### *Step 2 – Survey data*

Our survey data consisted of four types: (1) counts on survey blocks for sage-grouse pellets, ant mounds, lagomorphs, medium-sized rodents, and reptiles; (2) counts with distance estimates for birds and large-bodied mammals (lagomorphs and ungulates); (3) relative capture rates for small mammals; and (4) plant composition and cover estimates (discussed separately in Ch. 10). We derived detection probabilities for species sampled when possible (Buckland et al. 2001) (Ch. 6-8).

### *Step 3 – Model structure*

We used three modeling approaches to develop species occurrence or abundance models: count-based regressions, generalized ordered-logistic regressions, and logistic regressions (Fig. 4.4). The decision on which analysis to employ was based on (1) the sample size of survey blocks or transects with presences, and (2) whether data collected were counts or presence/absence. For species with counts, we used count-based models, investigating appropriate distributional form of the data (e.g., Poisson versus negative binomial), and also whether data were inherently zero-inflated. The expected output from count-based models is based on count estimates. We used ordered-logistic regression where the distribution of the counts prevented us from implementing count-based models (e.g., few counts over a broad range) or counts were an indicator rather than a direct measure of species abundance (e.g., sage-grouse pellets). For ordered-logistic regression models, we required a minimum of 50 observations within each count/abundance class. Classes were determined

based on apparent break points in counts/density frequency distributions. For species with less than 50 observations in each count/abundance class, we simply reverted to a presence/absence model using logistic regression. The expected outcome from ordered-logistic regression and logistic-regression analyses is based on a probability of occurrence estimate. All analyses were conducted in STATA 10.1 (STATA Corporation, College Station, TX).

We followed a recently developed two-staged approach for count-based models that incorporates detectability into count-based regression models when distance was recorded for individual detections (see Buckland et al. 2009). We first modeled detectability using the Multiple Covariate Sampling Engine in Program DISTANCE (Thomas et al. 2006). We develop the detection-function model for all observations for a given species by identifying the best detection function and form using AIC. We did so only for species with a minimum of 60 detections, allowing for proper estimation of the species detection function (Buckland et al. 2001). Note that 60 distance estimates could be obtained even if occurrence was less than 50 survey blocks or transects. We used observer team, time of year, time of day, and a shrub volume index (based on field measured data) when possible to assess the influence of covariates on detectability and to adjust density estimates. We used the top detection function to predict density on each survey block or transect. We then developed a generalized linear model (GLM) for each species using observed counts as the response variable and an offset term that included detection probability (that varied among sites) and survey effort (constant across sites) (Buckland et al. 2009). We restricted raw counts based on the truncation distance as identified in Program DISTANCE (Buckland et al. 2001). We used the offset term in the GLM to model observed counts while incorporating detectability differences across sites (Buckland et al. 2009).

Count data are typically Poisson-distributed, but when data are over-dispersed, a negative binomial distribution (mixture distribution of Poisson and gamma) may be more appropriate. Although a negative binomial regression model may account for excess zeros, a zero-inflated model (type of mixture model) is typically required to properly account for excess zeros in the dataset (Hilbe 2007). We evaluated different model structures and assessed the fit of each structure using a Vuong test (Vuong 1989). We first conducted a Vuong test using an intercept only model to identify the most appropriate of four exponential model forms: Poisson, negative binomial, zero-inflated Poisson (ZIP), or zero-inflated negative binomial (ZINB). We used the identified model form to evaluate the sagebrush land cover/NDVI sub-model (Step 5 below). After the top sagebrush land cover/NDVI sub-model was identified, we re-ran the Vuong test to confirm the top model form with base covariates. When zero-inflated processes were warranted, we maintained candidate model variables in both count and inflated portions of the model. Otherwise, potential model combinations became too cumbersome to evaluate. When incorporating offsets, expected outcome from count-based models result in density estimates.

We used generalized ordered-logistic regression analyses (Willams 2006) when distribution of the counts made it difficult to estimate count-based models or if counts were an indicator of species abundance rather than density of individuals (Ch. 5 and 7). We binned data into high and low abundance classes (0 = absence, 1 = low-medium abundance, 2 = high abundance) according to natural breaks in frequency distributions. Ordered-logistic regression uses an ordered (from low to high) categorical dependent variable to simultaneously estimate multiple equations, resulting in separate intercepts for each level (number of abundance classes in the dependent vari-

able minus one) and a single set of coefficients for each predictor variable. Unlike ordered-logistic regression, which assumes parallel regression lines of each abundance class, generalized ordered-logistic regression analyses relax this assumption (Willams 2006). We used the “GOLOGIT2” command in STATA 10 (STATA Corporation, College Station, TX), with the “autofit” option, which automatically relaxes the parallel constraint for those predictor variables that do not meet the parallel-line assumption and fits a separate slope for each abundance class.

We used logistic regression analyses (Hosmer and Lemeshow 2000) for those species whose survey data was an indicator of occurrence, no natural breaks in frequency distributions could be identified, or when count/abundance classes contained <50 survey blocks or transects. Survey blocks and transects were coded as presence if one or more individuals were detected.

#### *Step 4 – Predictor variable reduction*

We avoided perfect fit of predictor variables, variables containing almost exclusively zero-values, by screening each variable for presence of non-zero data values (Fig. 4.4). We set the threshold where at least 20 presence survey blocks or transects contained non-zero data values. We removed predictor variables from the standard candidate set if this criterion was not met. After we selected all candidate predictor variables, we checked for collinearity (Spearman rank correlation  $r_s \geq 0.7$ ) among the predictor variables. In cases where predictor variables were correlated, we retained variables at uncorrelated spatial scales or used *a priori* knowledge and ease of biological interpretation to select a single variable from the pair. We document the predictor variables, including descriptive statistics, used in each species distribution model in chapters to follow.

#### *Step 5– Sagebrush land cover/NDVI sub-model*

Our sampling design was based on presence of sagebrush-grassland land cover and NDVI. Thus, we first evaluated which combination of sagebrush land-cover class (0.27, 0.54, 1, 3, 5, and 18 km) and/or NDVI (0.27, 0.54, 1, 3, 5, and 18 km) had the best model fit when predicting species occurrence/abundance. We used *a priori* biological knowledge to select sagebrush land-cover classes to be included in this analysis. For example, if a species did not primarily inhabit mountain big sagebrush (*A. tridentata* ssp. *vaseyana*) land cover, we excluded mountain sagebrush only land cover class (MTNSAGE) from the regression analyses. We included all radii of selected sagebrush types in the analyses because little is known about the scale of sagebrush land cover important to species. We used  $AIC_c$  for model selection and carried forward the  $AIC_c$ -selected top sagebrush, NDVI, or sagebrush-NDVI model (parameters ( $k$ ) = 2–4 [intercept, sagebrush variable, NDVI variable, two variables for quadratic term or interaction]). We did not test interactions or quadratic terms if the sample size was  $\leq 60$  due to sample size limitations. We visually inspected presence/absence bi-plots and abundance scatter plots to evaluate whether interactions of sagebrush-NDVI or quadratic terms for both sagebrush and NDVI were justified.

#### *Step 6 – Selection of predictor variable scales*

We used univariate regression models to determine the best scale for each predictor variable that explained species occurrence/abundance (Fig. 4.4). Each univariate model included the sagebrush-NDVI sub-model selected from Step 5, along with a predictor variable at the given radii. We carried forward the  $AIC_c$ -best scale for each predictor variable.

#### *Step 7 – Number of predictor variables included in sub-models and final models*

We limited the number of predictor variables to 10% (Hosmer and Lemshow

2000) of the smallest sample size in each abundance or presence/absence class to avoid model over-fitting in logistic, ordered logistic, negative binomial, zero-inflated negative binomial, Poisson and zero-inflated Poisson regression analyses (Fig. 4.4). For example, candidate models could only include a maximum of ten predictor variables if the presence sample size was 104 survey blocks, including the variables from the sagebrush-NDVI base model in submodels and final models.

#### *Step 8 – Sub-model development for vegetation, abiotic, and anthropogenic disturbance variables*

We developed three sub-models based on vegetation, abiotic, and anthropogenic disturbance variables (Fig 4.4). Our goal was to select the best combination of each predictor variable and extent within each sub-model group. Candidate models for each sub-model group consisted of the sagebrush-NDVI sub-model selected in Step 5 and all possible combinations of predictor variables in each category selected in Step 6, limited to the number of variables identified in Step 7. We carried forward the  $AIC_c$ -selected top sub-model to the next step.

#### *Step 9 – Final model*

We allowed all predictor variables within each of the  $AIC_c$ -best submodels for vegetation, abiotic, and anthropogenic disturbance categories (Step 8) to compete, both within and across submodels (Fig. 4.4). The sagebrush/NDVI submodel (Step 5) was again held constant in all models. All possible candidate models were competed; final models were ranked based on  $AIC_c$ , and model weights ( $w_i$ ) were calculated. We incorporated model uncertainty into the final composite predictive model by using model-averaged coefficients based on weights from all candidate models within a cumulative  $AIC_c$  weight just  $\geq 0.9$  (Burnham and Anderson 2001). We set

coefficients to zero when a model did not contain a particular variable.

*Step 10 – Spatial application, dose response curves, and model evaluation*

We develop maps of species occurrence or abundance at a 90-m cell size by spatially applying the final composite model using raster calculator in ArcMap 9.3.1 (ESRI 2006) (Fig. 4.4). We binned final model predictions for summary and display. Non-sagebrush habitats (areas with <3% sagebrush habitat in a 5-km radius) where we did not sample were masked, and no predictions were made to these areas.

We evaluated accuracy of generalized ordered logistic and logistic regression models using receiver operating characteristic (ROC) estimating the area under the curve (AUC, Metz 1978). AUC is a discrimination index based on likelihood for a presence to have a higher species occurrence probability when compared to a randomly selected absence point. We used this metric as one indicator of model performance, fully cognizant of potential problems if ROC is the only metric used to evaluate model performance (Lobo et al. 2008). We used the sensitivity-specificity equality approach (Liu et al. 2005) to determine the optimal cutoff threshold for predicting presence-absence of each species (habitat or non-habitat) and used this threshold to assess the predictive capacity for each model (Nielsen et al. 2004, Lobo et al. 2008).

We created dose response curves for each species by plotting predicted probability of occurrence or density relative to changes in sagebrush quantity. This permitted us to assess critical levels of sagebrush required for a species across the WBEA landscape, as well as characterize response to losses or fragmentation of sagebrush habitat. We used the Dose Response Calculator for ArcGIS (Hanser et al. 2011) to calculate the mean probability of occurrence or density from the spatial model output across one percent intervals of the

sagebrush predictor variable, 0.01 intervals of NDVI, or distance intervals from anthropogenic features, where appropriate. We used the optimal cutoff or minimum densities to identify the sagebrush or productivity threshold values above which a species was likely to occur.

We used independent survey data when available to evaluate predictive outputs of species models (Pearce and Ferrier 2000, Strauss and Biederman 2007). We used three data sets to validate models: (1) Wyoming Fish and Game (pronghorn, Bob Oaklef pers. comm.; sage-grouse, Tom Christiansen pers. comm.), (2) Wyoming Natural Diversity Database (reptile models; Wyoming Natural Diversity Database 2009), and (3) Breeding Bird Survey (USGS Breeding Bird Survey Data, <http://www.mbr-pwrc.usgs.gov/bbs/>) data sets (songbird models). To examine performance of models based on logistic regression analyses, we first binned each model into 10 equal probability classes, and then counted presence locations and calculated area in each bin. We used this information to determine expected observations per bin and regressed proportion of expected against observed observations (Johnson et al. 2006). A model well supported by validation data will have (1) a slope not differing from one, (2) an intercept near zero, and (3) a high  $R^2$  value (Nielsen et al. 2004). As a more general evaluation of songbird models (Ch. 6) we used BBS data from 2005–06 and compared mean counts across entire BBS routes with averaged model predictions (density or probability of occurrence) along each BBS survey route. Predictive models should have a significant and positive correlation with independent count data, even though BBS data do not account for differences in detectability.

## DISCUSSION

Conducting floral and faunal sampling across large scales is a costly endeavor and

logistically challenging (Franklin 2009). Given these hurdles, few studies to date have investigated how wildlife and plant communities respond to habitat-anthropogenic disturbance gradients across large scales (Franklin 2009). Moreover, most studies do not sample all possible habitat-anthropogenic disturbance combinations or gradients (e.g., low habitat suitability – high anthropogenic disturbance). Yet such field data are crucial when evaluating ecoregional assessment outcomes and predictions. To our knowledge, our study is one of a few that has sampled habitat-anthropogenic disturbance interactions across large spatial extents and covered the possible range of habitat-anthropogenic disturbance combinations.

An inherent problem of faunal surveys is to find trained field biologists capable of sampling a suite of species in different taxonomic groups (Noss et al. 1997). Although some taxonomic groups are easier to sample than others, we had difficulty training field technicians in identifying all possible bird species by sound. We recommend that a subset of bird species be sampled rather than a complete inventory of the avian community to minimize errors associated with identifying all breeding species that may possibly occur. This approach can be applied to any taxonomic group. Subsets of species should be selected according to habitat associations, life history traits, or sensitivity to perceived anthropogenic threats. Ultimately these species should be potential indicators of biodiversity (Mac Nally and Fleishman 2004). The cost of sampling and logistics associated with training field technicians can be reduced by having at least one well-trained technician per survey protocol in each team to assist in training inexperienced biological technicians.

Our hierarchical multi-stage modeling approach, although exploratory in nature, worked well in developing species occurrence and abundance models for sagebrush-associated species. Very little was known about how most species in our assessment

responded to land cover composition and configuration and human disturbance and at which spatial extents these responses might be strongest. Therefore, field data collection and an exploratory analytical approach, as we have outlined here, was the first step in conducting statistically rigorous studies that investigate thresholds at which species occurrence and abundance are influenced by human disturbance.

## LITERATURE CITED

- ALDRIDGE, C. L., AND M. S. BOYCE. 2007. Linking occurrence and fitness to persistence: habitat-based approach for endangered greater sage-grouse. *Ecological Applications* 17:508–526.
- AVILA-FLORES, R., M. S. BOYCE, AND S. BOUTIN. 2010. Habitat selection by prairie dogs in a disturbed landscape at the edge of their geographic range. *Journal of Wildlife Management* 74:945–953.
- BIBBY, C. J., N. D. BURGESS, AND D. A. HILL. 1992. *Bird census techniques*. Academic Press, London, UK.
- BOARMAN, W. I., AND B. HEINRICH. 1999. Common raven (*Corvus corax*). in A. Poole and F. Gill (editors). *The birds of North America* No. 476. The Academy of Natural Sciences, Philadelphia, PA and The American Ornithologists Union, Washington, DC.
- BUCKLAND, S. T., D. R. ANDERSON, K. P. BURNHAM, J. L. LAAKE, D. L. BORCHERS, AND L. P. THOMAS. 2001. *Introduction to distance sampling: estimating abundance of biological populations*. Oxford University Press, Oxford, UK.
- BUCKLAND, S. T., D. R. ANDERSON, K. P. BURNHAM, J. L. LAAKE, D. L. BORCHERS, AND L. P. THOMAS. 2004. *Advanced distance sampling*. Oxford University Press, Oxford, UK.
- BUCKLAND, S. T., R. E. RUSSELL, B. G. DICKSON, V. A. SAAB, D. N. GORMAN, AND W. M. BLOCK. 2009. Analyzing designed experiments in distance sampling. *Journal of Agricultural, Biological, and Environmental Statistics* 14:432–442.
- BURNHAM, K. P., AND D. R. ANDERSON. 2002. Model selection and multimodel inference: a

- practical information-theoretic approach. Second edition. Springer Verlag, New York, NY.
- CARPENTER, J. E., C. L. ALDRIDGE, AND M. S. BOYCE. 2010. Sage-grouse habitat selection during winter in Alberta. *Journal of Wildlife Management* 74:1806–1814.
- CARROLL, M. L., C. M. DiMICELI, R. A. SOHLBERG, AND J. R. G. TOWNSHEND. 2006. 250 m MODIS Normalized Difference Vegetation Index. University of Maryland, College Park, MD.
- COMER, P., J. KAGAN, M. HEINER, AND C. TOBALSKE. 2002. Current distribution of sagebrush and associated vegetation in the western United States (excluding NM and AZ). Digital Map 1:200,000 scale. USGS Forest and Rangeland Ecosystems Science Center, Boise, ID, and The Nature Conservancy, Boulder, CO. <<http://sagemap.wr.usgs.gov/ftp/regional/TNC/sagestitch.zip>> (20 September 2011).
- CONNELLY, J. W., M. A. SCHROEDER, A. R. SANDS, AND C. E. BRAUN. 2000. Guidelines to manage sage grouse populations and their habitats. *Wildlife Society Bulletin* 28:967–985.
- COUDUN, C., AND J.-C. GÉGOUT. 2006. The derivation of species response curves with Gaussian logistic regression is sensitive to sampling intensity and curve characteristics. *Ecological Modelling* 199:164–175.
- DOHERTY, K. E., D. E. NAUGLE, B. L. WALKER, AND J. M. GRAHAM. 2008. Greater sage-grouse winter habitat selection and energy development. *Journal of Wildlife Management* 72:187–195.
- ESRI. 2006. ARC/INFO version 9.2. Environmental Systems Research Institute, Inc., Redlands, CA.
- FRANKLIN, J. 2009. Mapping species distributions: spatial inference and predictions. Cambridge University Press, New York, NY.
- FRELICH, J., B. BUDD, T. KOHLEY, AND B. HAYDEN. 2001. Wyoming Basins ecoregional plan. The Nature Conservancy, Boulder, CO.
- GESSLER, P. E., I. D. MOORE, N. J. MCKENZIE, AND P. J. RYAN. 1995. Soil–landscape modeling and spatial prediction of soil attributes. *International Journal of Geographic Information Systems* 4:421–432.
- GROVES, C., L. VALUTIS, D. VOSICK, B. NEELY, K. WHEATON, J. TOUVAL, AND B. RUNNELS. 2000. Designing a geography of hope: a practitioner's handbook for ecoregional conservation planning. Second edition, Volume 1. The Nature Conservancy, Arlington, VA. <<http://conserveonline.org/docs/2000/11/GOH2-v1.pdf>> (20 September 2011).
- HANSER, S. E., C. L. ALDRIDGE, M. LEU, AND S.E. NIELSEN. 2011. Dose-response calculator for ArcGIS. United States Geological Survey Data Series 631. <<http://pubs.usgs.gov/ds/631/>> (20 September 2011).
- HILBE, J. M. 2007. Negative binomial regression. Cambridge University Press, New York, NY.
- HOSMER, D. W., AND S. LEMESHOW. 2000. Applied logistic regression, second edition. John Wiley and Sons, New York, NY.
- JOHNSON, C. J., S. E. NIELSEN, E. H. MERRILL, T. L. McDONALD, AND M. S. BOYCE. 2006. Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. *Journal of Wildlife Management* 70:347–357.
- JOHNSTON, R. F. 2001. Synanthropic birds of North America. Pp. 49–67 in J. M. Marzluff, R. Bowman, and R. Donnelly (editors). *Avian ecology and conservation in an urbanizing world*. Kluwer Academic Publishers, Boston, MA.
- KOCHERT, M. N., K. STEENHOF, C. L. MCINTYRE AND E. H. CRAIG. 2002. Golden eagle (*Aquila chrysaetos*). In A. Poole and F. Gill (editors). *The birds of North America* No. 684. The Academy of Natural Sciences, Philadelphia, PA and The American Ornithologists' Union, Washington, DC.
- LANDFIRE. 2007. LANDFIRE 1.0.0 existing vegetation type layer. U.S. Department of Interior, Geological Survey. <<http://landfire.cr.usgs.gov/viewer/>> (20 September 2011).
- LEU, M., AND S. E. HANSER. 2011. Influences of the human footprint on sagebrush landscape patterns: implications for sage-grouse conservation. Pages 253–271 in S. T. Knick and J. W. Connelly (editors). *Greater sage-grouse: ecology and conservation of a landscape species and its habitats*. Studies in Avian Biology 38. University of California Press, Berkeley, CA.

- LEU, M., S. E. HANSER, AND S. T. KNICK. 2008. The human footprint in the west: a large-scale analysis of anthropogenic impacts. *Ecological Applications* 18:1119–1139.
- LIU, C., P. M. BERRY, T. P. DAWSON, AND R. G. PEARSON. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385–393.
- LOBO, J. M., A. JIMÉNEZ-VALVERDE, AND R. REAL. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17:145–151.
- MAC NALLY, R., AND E. FLEISHMAN. 2004. A successful predictive model of species richness based on indicator species. *Conservation Biology* 18:646–654.
- MARZLUFF, J. M., S. T. KNICK, M. S. VEKASY, L. S. SCHUECK, AND T. J. ZARRIELLO. 1997. Spatial use and habitat selection of golden eagles in southwestern Idaho. *Auk* 114:673–687.
- MCGARIGAL, K., S. A. CUSHMAN, M. C. NEEL, AND E. ENE. 2002. FRAGSTATS: spatial pattern analysis program for categorical maps. <<http://www.umass.edu/landeco/research/fragstats/fragstats.html>> (20 September 2011).
- MCCMAHON, G., S. M. GREGONIS, S. W. WALTMAN, J. M. OMERNIK, T. D. THORSON, J. A. FREEOUF, A. H. RORICK, AND J. E. KEYS. 2001. Developing a spatial framework of common ecological regions for the conterminous United States. *Environmental Management* 28:293–316.
- METZ, C. E. 1978. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* 8:283–298.
- MILLER, D. A., AND R. A. WHITE. 1998. A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth Interactions* 2:1–26. <[http://www.soilinfo.psu.edu/index.cgi?soil\\_data&index.html](http://www.soilinfo.psu.edu/index.cgi?soil_data&index.html)> (20 September 2011).
- NEELY, B., P. COMER, C. MORITZ, M. LAMMERT, R. RONDEAU, C. PAGUE, G. BELL, H. COPELAND, J. HUMKE, S. SPACKMAN, T. SCHULZ, D. THEOBALD, AND L. VALUTIS. 2001. Southern Rocky Mountains: an ecoregional assessment and conservation blueprint. The Nature Conservancy, U.S. Forest Service, Rocky Mountain Region, Colorado Division of Wildlife, and U.S. Bureau of Land Management. <<http://conserveonline.org/coldocs/2002/02/SRMreport.pdf>> (20 September 2011).
- NIELSEN, S. E., J. CRANSTON, AND G. B. STENHOUSE. 2009. Identification of priority areas for grizzly bear conservation and recovery in Alberta, Canada. *Journal of Conservation Planning* 5:38–60.
- NIELSEN, S. E., R. H. M. MUNRO, E. L. BAINBRIDGE, G. B. STENHOUSE, AND M. S. BOYCE. 2004. Grizzly bears and forestry II. Distribution of grizzly bear foods in clearcuts of west-central Alberta, Canada. *Forest Ecology and Management* 199:67–82.
- NOSS, R. F., M. A. O'CONNELL, AND D. D. MURPHY. 1997. *The science of conservation planning: habitat conservation under the Endangered Species Act*. Island Press, Washington, DC.
- NOSS, R., G. WUERHNER, K. VANCE-BORLAND, AND C. CARROLL. 2001. A biological conservation assessment for the Utah-Wyoming-Rocky Mountains ecoregion: a report to The Nature Conservancy. Conservation Science, Inc., Corvallis, OR. <[http://conserveonline.org/coldocs/2003/10/uwrm\\_plan\\_ver2001.pdf](http://conserveonline.org/coldocs/2003/10/uwrm_plan_ver2001.pdf)> (20 September 2011).
- PEARCE, J., AND S. FERRIER. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133:225–245.
- PRISM GROUP. 2006a. United States average annual precipitation, 1971 – 2000. Oregon State University. <<http://www.prismclimate.org>> (20 September 2011).
- PRISM GROUP. 2006b. United States maximum annual temperature, 1971 – 2000. Oregon State University. <<http://www.prismclimate.org>> (20 September 2011).
- PRISM GROUP. 2006c. United States minimum annual temperature, 1971 – 2000. Oregon State University. <<http://www.prismclimate.org>> (20 September 2011).
- RABE, M. J., M. S. SIDERS, C. R. MILLER, AND T. K. SNOW. 1998. Long foraging distance for a spotted bat (*Euderma maculatum*) in northern Arizona. *Southwestern Naturalist* 43:266–269.
- RILEY, S. J., S. D. DEGLORIA AND R. ELLIOT. 1999. A terrain ruggedness index that quan-

- tifies topographic heterogeneity. *Intermountain Journal of Sciences* 5:1–4.
- SAWYER, H., F. LINDZEY, AND D. MCWHIRTER. 2005. Mule deer and pronghorn migration in western Wyoming. *Wildlife Society Bulletin* 33:1266–1273.
- STRAUSS, B., AND R. BIEDERMANN. 2007. Evaluating temporal and spatial generality: how valid are species-habitat relationship models? *Ecological Modelling* 204:104–114.
- THE NATURE CONSERVANCY. 2000. Middle Rockies-Blue Mountains ecoregional conservation plan. Middle Rockies-Blue Mountains Planning Team. <[http://conserveonline.org/coldocs/2002/05/ERP\\_with\\_appendices.pdf](http://conserveonline.org/coldocs/2002/05/ERP_with_appendices.pdf)> (20 September 2011).
- THOMAS, L., J. L. LAAKE, S. STRINDBERG, F. F. C. MARQUES, S. T. BUCKLAND, D. L. BORCHERS, D. R. ANDERSON, K. P. BURNHAM, S. L. HEDLEY, J. H. POLLARD, J. R. B. BISHOP, AND T. A. MARQUES. 2006. Distance 5.0, Release 2. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. <<http://www.ruwpa.st-and.ac.uk/distance/>> (20 September 2011).
- VUONG, Q. H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57:307–334.
- WALKER, B. L., D. E. NAUGLE, AND K. E. DOHERTY. 2007. Greater sage-grouse population response to energy development and habitat loss. *Journal of Wildlife Management* 71:2644–2654.
- WELLER, C., J. THOMPSON, P. MORTON, AND G. APLET. 2002. *Fragmenting our lands: the ecological footprint from oil and gas development*. The Wilderness Society, Washington, DC.
- WISDOM, M. J., M. M. ROWLAND, AND L. H. SURRING. 2005. *Habitat threats in the sagebrush ecosystem: methods of regional assessments and applications in the Great Basin*. Alliance Communication Group, Lawrence, KS.
- WYOMING NATURAL DIVERSITY DATABASE. 2009. Unpublished report. Wyoming Natural Diversity Database, University of Wyoming, Laramie, WY.

#### APPENDIX 4.1

Summary of literature review on effect area of various anthropogenic disturbances (if not reported in the metric system, units were converted from originally reported values). This appendix is archived electronically and can be downloaded at the following URL: <http://sagemap.wr.usgs.gov/wbea.aspx>.

#### APPENDIX 4.2

Home range estimates for 40 species of concern in the Wyoming Basins Ecoregional Assessment area. Scientific names for species are provided in Chapter 2. The minimum and maximum home range or territory size are provided along with the corresponding citations, where available. Values other than the minimum and maximum are also provided. Where applicable we provide the citation if different from those used to obtain the minimum and maximum values. Values are in hectares unless specified otherwise. This appendix is archived electronically and can be downloaded at the following URL: <http://sagemap.wr.usgs.gov/wbea.aspx>.

#### APPENDIX 4.3

Exotic plants species sampled during 2005 and 2006 for the Wyoming Basins Ecoregional Assessment. This appendix is archived electronically and can be downloaded at the following URL: <http://sagemap.wr.usgs.gov/wbea.aspx>.