Can habitat selection predict abundance?

Mark S. Boyce¹, Chris J. Johnson², Evelyn H. Merrill¹, Scott E. Nielsen³, Erling J. Solberg⁴ and Bram van Moorter⁴

¹Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada; ²Ecosystem Science and Management Program, University of Northern British Columbia, 3333 University Way, Prince George, BC V2N 4Z9, Canada; ³Department of Renewable Resources, University of Alberta, Edmonton, AB T6G 2H1, Canada; and ⁴Norwegian Institute for Nature Research (NINA), Trondheim 7485, Norway

Summary

1. Habitats have substantial influence on the distribution and abundance of animals. Animals’ selective movement yields their habitat use. Animals generally are more abundant in habitats that are selected most strongly.

2. Models of habitat selection can be used to distribute animals on the landscape or their distribution can be modelled based on data of habitat use, occupancy, intensity of use or counts of animals. When the population is at carrying capacity or in an ideal-free distribution, habitat selection and related metrics of habitat use can be used to estimate abundance.

3. If the population is not at equilibrium, models have the flexibility to incorporate density into models of habitat selection; but abundance might be influenced by factors influencing fitness that are not directly related to habitat thereby compromising the use of habitat-based models for predicting population size.

4. Scale and domain of the sampling frame, both in time and space, are crucial considerations limiting application of these models. Ultimately, identifying reliable models for predicting abundance from habitat data requires an understanding of the mechanisms underlying population regulation and limitation.

Key-words: animal movement, occupancy, population estimation, population size, presence-only data, resource selection functions

Introduction

Habitat is a primary determinant of the distribution and abundance of organisms and is the target for most conservation efforts. Boyce & McDonald (1999) recognized that resource selection functions (RSF) could be used to map the probability of use of habitats, and by summing these probabilities, population size could be estimated. Since then, developments in computer-mapping technology (GIS) as well as global-positioning satellite (GPS) radiotelemetry have generated unprecedented opportunity (Cagnacci et al. 2010) and these developments have motivated a variety of statistical methods for defining patterns of habitat selection and use (Johnson, Hooten & Kuhn 2013; Hooten et al. 2014; Table 1). How these patterns relate to the abundance of animals is the focus of this review.

Habitat selection is the probability that when a resource unit (e.g. a pixel) is encountered that it will be used by the animal (Lele et al. 2013). Selection reflects animal behaviour that is mechanistically linked to animal movement (Moorcroft & Barnett 2008), usually estimated from location/available (= use/available) data (Johnson et al. 2006). For example, modern GPS biotelemetry yields relocation data of high precision and these observations can be contrasted with random landscape locations (available locations) assuming each of these locations has been encountered. These presence/available data are sometimes termed ‘presence-only’ data because we usually do not sample areas to find they were unused. Instead, usually random...
pixels or resource units are drawn using geographical information systems (GIS). Several attempts have been made to treat presence/available data as though they were presence/absence data (= occupancy) assuming that the random resource units were unused, sometimes referred to as pseudo-absences (Ward et al. 2009; Royle et al. 2012). Generally, this is a bad idea because results are biased estimates of occupancy (Lancaster & Imbens 1996; Keating & Cherry 2004; Hastie & Fithian 2013).

Instead of trying to force a ‘presence-only’ design into the occupancy framework, one can more appropriately use these data to estimate the selection of habitats by individual animals (Boyce 2010), which can then be used to estimate probability of use, occupancy, and as we argue, abundance. Selection can be estimated based on the distribution of attributes associated with used resource units contrasted with the distribution of available resource units using the logistic discriminant function (Seber 1984) to obtain a resource selection function (RSF) from which we can calculate the relative probability of selection (Johnson et al. 2006). Using weighted distribution theory, the resource selection probability function (RSPF) can be estimated from which we can calculate the actual probability of selection (Lele 2009; Lele et al. 2013).

Our review on the linkage between habitat selection and abundance consists of three parts. First, we start from the relationship between selection, movement and habitat use, because movement is the behavioural mechanism leading to selective habitat use. Secondly, we investigate the link between habitat selection, use and abundance directly. Finally, we look into factors limiting this relationship, in particular changing availabilities, seasonal migration, fitness and herding behaviour.

Relating movement to selection and resource use

Frequent relocations collected by GPS biotelemetry and other techniques now allow one to document animal locations and to infer the movement track and behaviour of individual animals (Bergman, Schaefer & Luttich 2000; Bailey & Thompson 2006). We consider selective movement to be the mechanism that produces spatial patterns of individual habitat use. As such, the attributes of movement paths can be quantified and modelled as step-selection functions that are a function of predictor covariates reflecting the mechanisms influencing selection (Thurfjell, Ciuti & Boyce 2014). GPS telemetry data have permitted investigations of the behavioural mechanisms dictating predator–prey interactions (Hebblewhite, Merrill & McDonald 2005), responses to patchy resource distribution at spatial scales finer than the annual or seasonal range, and life-history responses including dispersal and migration
defined by net displacement (Blinking resource use to fitness. For example, a zone of secondary statistical design for estimating selection or for characterizing habitat selection, and ultimately population size from multiple model structures and assumptions lends credence to the generality of the RSF as a model for characterizing movement and selection and the behavioural process of habitat selection and including types of movement or the specific interactions between habitat use and behaviours in those habitats (Frait et al. 2005; DeMars et al. 2013).

Integrating vital rates resulting from movements and habitat use is a more challenging problem requiring simulation. Of the tools and methods available, spatial population models permit direct exploration of the relationships among movement, habitat selection and population dynamics. As an example, the program HexSim (formerly known as PATCH, e.g. Schumaker et al. 2004) allows one to simulate movement of individuals among territories that vary in resource quality. The quality of each territory is used to scale survival and reproduction for individuals occupying that territory. At each time step, a range of movement rules can be implemented to dictate patterns of distribution (e.g. dispersal) to include potential interactions with territory quality. Model predictions then can include temporal and spatial (by territory) measures of population abundance and change. Marcot et al. (2013) used HexSim to evaluate the influence of the size and distribution of habitat patches on persistence of northern spotted owls (*Strix occidentalis caurina* Xantus De Vesey). In particular, they noted the importance of colonization of low-density patches by dispersing owls, recognizing the interaction of resource use and movement. Again, there is a direct link among habitat selection, movement, density and fitness (Mulder & Ruess 2001). State-space models and agent-based models offer alternative approaches for exploring the spatial and population dynamics of mobile animals responding to heterogeneous environments (Patterson et al. 2008; Semeniuk et al. 2012; Watkins et al. 2014).

### Relationships between habitat selection, use and abundance

Manly, McDonald & Thomas (1993) defined a RSF to be any function that is proportional to the probability of use of a resource unit. But their definition for the RSF actually applies to the probability of selection instead of use, creating some confusion in the literature (Lele et al. 2013). Boyce & McDonald (1999) recognized that the probability of use could be estimated from the RSF and availability of habitat types, and then summing the probability of use (see Table 1) over a landscape should yield total population size. They proposed that this provided a link between habitat use, $U(.)$, and abundance as:

$$
\text{abundance} = \int U(.) \, dA
$$
\[ U(x_i) = \frac{w(x_i)A(x_i)}{\sum_j w(x_j)A(x_j)}. \]

where RSF, \( w() \), times area, \( A() \), is summed over the habitat types, \( x_i \). The expected abundance in the \( i \)-th habitat type is \( N_i = N \cdot U(x_i) \) where \( N \) is population size and the density in the \( i \)-th habitat type can be estimated by multiplying population size by relative use adjusted by the area:

\[ D(x_i) = N \cdot \frac{U(x_i)}{A(x_i)}. \]

Naturally this requires explicit definition of scope and scale determined by the sampling scheme – the estimate of abundance would apply over the domain of the area sampled and the time period during which the sampling was conducted. When applied to a new area, it is assumed that the availabilities of habitats and the population size remain the same as under the conditions when the RSF was derived. The availability assumption can be evaluated explicitly, for example, in the prediction of grizzly bear (\( Ursus arctos \) L.) abundance for the Bitterroots of Idaho, Boyce & Waller (2003) compared the frequency distribution of RSFs for available resource units in both the reference areas and the landscapes where grizzly bear abundance was being predicted. If availability varies, we might be able to model the RSF as a function of availability (Knopff et al. 2014). This RSF-based method also has been used for estimating abundance of Amur leopards (\( Panthera pardus orientalis \) Schlegel; Hebblewhite et al. 2011) and wolverines (\( Gulo gulo \) L., Inman et al. 2013).

Alternative approaches to habitat-based estimates of abundance exist where habitat use is connected directly to density. For example, in British Columbia, a number of studies have been conducted to estimate grizzly bear density in selected habitat types, and from these density estimates, Fuhr & Demarchi (1990) simply multiplied population density estimates times the total area of each local habitat type then summed over all habitat types within the study area to estimate population size. In this Fuhr & Demarchi (1990) method, habitat selection is implicit in determining the density of animals occurring in each habitat type. This is identical to the estimator of Boyce & McDonald (1999) that was applied by Nielsen (2011) to a population of grizzly bears in British Columbia where \( D(x_i) \) is the density in the \( j \)-th habitat type characterized by a vector of \( x \) covariates in an area, \( A \):

\[ \hat{N} = \sum_j D(x_j)A(x_j). \]

Other resource (pixel)-based models that use either intensity of use by individual animals (e.g. telemetry points) or direct counts of animals require the definition of resource units. Dimensions of these units are often arbitrary, but the movement dynamics of the population of interest might guide such sampling decisions. For example, Sawyer et al. (2006) modelled the log frequency of the number of telemetry points of mule deer (\( Odocoileus hemionus \) Rafinesque) in 100-m-radius circular cells as a function of elevation, slope and the distance to oil wells and roads to directly estimate \( U(x_i) \) in the equation above rather than \( w(x_i) \). Another analysis of the intensity of use is where the census of wolves (\( Canis lupus \) L.) in Poland was used to estimate a RSF (Jedrzejewski et al. 2008). The log frequency of wolf observations was modelled as a function of habitat covariates, validated using \( k \)-fold cross validation and then extrapolated across all of Poland.

Because habitat selection shapes use, we can build models directly from counts of animals, albeit requiring different link functions. For count-based models, Nielsen et al. (2005) suggested that zero-inflated binomial (ZIB) models might be an appropriate structure, and provided examples using zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models to describe distribution and abundance. These ZIB regression models yield two equations: one that links habitat selection to abundance assuming either a Poisson or negative binomial distribution and another regression equation to model the excess zeros beyond those predicted by the binomial model. Likewise, zero-altered binomial (ZAB) or hurdle models can be a useful approach for dealing with the excess zeros that are typical of ecological data (Potts & Elith 2006), still using two equations but where the binomial model is zero-truncated and a separate equation estimates the probability that a resource unit is occupied vs. absent (i.e. occupancy). Again the binomial model can be Poisson (ZAP) or negative binomial (ZANB).

Much less informative than count data, occupancy is typically measured in context of a spatial unit where an organism is either present or absent (Austin 2007). Logistic regression is ideally suited for analysis of these data because resource units can be classified into discrete categories where the organism is present (1) vs. those where it is absent (0). For animals, a major challenge with this plot-based approach is determining absence in an area because animals might be difficult to detect or because sites might be used at various times (especially for large animals and studies with small spatial scales or resource units). Methods for adjusting occupancy estimates for detection probability have been developed (MacKenzie et al. 2006), although this often requires substantial increases in sampling effort (Solymos, Lele & Bayne 2012).

Occupancy–abundance relationships are among the most fundamental patterns in ecology (Andrewartha & Birch 1954) having been described for plants, fish, birds, amphibians and mammals (Winters & Wheeler 1985; Gibbons, Reid & Chapman 1993; Boecken & Shachak 1998; Mossman et al. 1998; Tosh, Reyers & van Jaarsveld 2004). Therefore, it is no surprise that occupancy surveys have been used to predict the abundance of organisms (Nachman 1981; He & Gaston 2003; Royle & Nichols 2003). Empirical patterns of abundance–occupancy
relations (AORs) have been studied in several taxa including birds, plants and butterflies, both within and among species (Gaston, Blackburn & Lawton 1997). AORs are generally strong positive functions for animals, but patterns are more complex among plants and for among-species patterns (Buckley & Freckleton 2010). The link between occupancy and abundance also might be expected given that selection is common to both occupancy and the RSF (Table 1).

When sampling fractions are known, the probability of occupancy can be estimated from samples where resource units are not selected randomly but are based on the occurrence of an animal (Manly, McDonald & Thomas 1993: p. 108), thus allowing occupancy to predict abundance (Boyce, Meyer & Irwin 1994). This requires estimates of the proportion of available resource units that were sampled (\( P_u \)), and the proportion of occupied resource units that were in the sample of used units (\( P_a \)). This was done for a population of northern spotted owls where a complete inventory of owl nests allowed calculation of sampling fractions where there was a single nest in each occupied territory (Boyce, Meyer & Irwin 1994). However, we know few examples where these sampling fractions are known, and thus, there are few applications of this method.

Caveats

If a population is at carrying capacity, that is equilibrium abundance, we might expect that habitat selection could be used to estimate abundance (e.g. Boyce & McDonald 1999). However, if the population is governed by an ideal-free distribution (Hache, Villard & Bayne 2013), fitness ought to be approximately equal in all habitats with animals adjusting their abundance by habitat according to population density (Fretwell & Lucas 1969), so it might not be a critical assumption that the population is at carrying capacity (McLoughlin et al. 2010). However, birds and mammals seldom adhere to an ideal-free distribution with social interactions such as dominance and territoriality playing a major role in habitat use leading to an ideal-despotic distribution (Bock & Jones 2004; McLoughlin et al. 2006, 2007). Whether the distribution is ideal free or despotic, the relation between habitat selection and abundance is not necessarily violated. In fact, the empirical pattern is that fitness is usually correlated with population density (Bock & Jones 2004; McLoughlin et al. 2006, 2007).

However, exceptions can be found where density does not reflect the patterns of habitat selection by individuals (Van Horne 1983; Stephens et al. 2015). These exceptions include source and sink habitats where abundance in unproductive (sink) habitats is maintained by dispersers from a source population (Pulliam & Danielson 1991), and ecological traps or attractive sinks where animals might be attracted to an area, say because of available foods, but these are risky habitats where mortality is high (Dwernychuk & Boag 1972; Delibes, Ferreras & Gaona 2001; Nielsen, Boyce & Stenhouse 2006).

Transient populations above or below carrying capacity cannot be predicted without information on the mechanisms that govern abundance, such as limiting factors or population regulating mechanisms, that is density dependence (Fortin, Morris & McLoughlin 2008). For example, data on lynx (Lynx canadensis Kerr) fur returns dating to 1763 have demonstrated a regular 10-year cycle of abundance that is fundamentally tied to a predator–prey interaction with snowshoe hares (Lepus americanus Erxleben) (Krebs, Boutin & Boonstra 2001). We might estimate an occupancy model for lynx during the nadir of abundance and use this to extrapolate abundance of the animals across a landscape. Obviously, however, this occupancy model would seriously underestimate abundance that would occur when the lynx population was at its peak. If habitat selection were density dependent (Fortin, Morris & McLoughlin 2008; McLoughlin et al. 2010), such an occupancy model would perform poorly except at comparable phases of the population cycle. Still, if snowshoe hare abundance was the driver for lynx abundance, one might construct a model to predict lynx populations by including hares as a covariate in the occupancy model. In practice, obtaining such data on prey abundance would usually limit our ability to apply such a model.

Although many examples exist where habitat selection has been used to model abundance, there are other instances when it is not expected to work. Extrapolating abundance, as in the Boyce & McDonald (1999) method, ideally assumes that the RSF was estimated in a reference population where abundance was estimated independently for a population near carrying capacity with similar selection and availabilities as in the area where abundance is being projected. And applications must involve the same scale for resource units (Boyce 2006). Violations to these assumptions will require creative solutions or the development of new methods even when there is a strong association between habitats and the distribution and abundance of animals.

Likewise, we cannot expect this method necessarily to work for populations that have substantial shifts in seasonal distribution or grouping behaviours such as herding. Animals often select different habitats in each season depending upon the seasonal distribution and availability of resources. Populations of migratory birds or mammals (e.g. caribou & elk) migrate large distances selecting very different habitats during winter than during the breeding and summer seasons (Johnson & Seip 2008). Depending on weather conditions, some seasonal ranges might not be used in any particular year with the animals shifting to alternative sites.

Population limitation or regulation might be most strongly tied to a particular season, and comparison among seasonal models can help to identify the season for which projected population estimates are lowest.
Snow goose (Chen caerulescens L.) populations, for example, appear to be limited by foods on wintering areas in southern North America. With expansion of rice farming and other agriculture, food for snow geese on wintering areas became less limiting resulting in large increases in abundance (Ankney 1996). Components of fitness can be decomposed by season so that the importance of seasonal habitats can be modelled explicitly (Aldridge & Boyce 2007).

MOOSE AND GRIZZLY BEAR EXAMPLES OF HABITAT SELECTION—ABUNDANCE

Selection of habitats by moose (Alces alces L.) during summer in central Norway (Bjørneraas et al. 2012) provides a case study for examining some of the potential limitations to habitat-determined distribution and abundance. Here, the population was maintained below carrying capacity by hunting, yet abundance during summer was positively correlated with availability of preferred habitat types, indicating a functional response (Mysterud & Ims 1998). This is possible if animals are moving into habitats that yield the greatest fitness advantages, and densities are not required to be at carrying capacity for this to function (Fortin, Morris & McLoughlin 2008). Also, moose in more productive habitats maintained smaller home ranges and higher density. The issue of scale is important here, because the scale of habitat selection and management often are different (Mowat, Heard & Schwarz 2013). Moreover, habitat selection often occurs at different spatio-temporal scales. For moose in Norway, the scale of the management unit (MU) is larger than the scale of small-scale habitat selection (not migration), while the scale of migration tends to be larger than the scale of MU. The MUs (municipality) are at the level of 100–3000 km². Within these MUs, there is fairly consistent management (hunting permits km⁻² forestland); however, variation exists among MUs. Therefore, we cannot rely on density as a measure of variation in habitat quality among MUs.

We can use performance relative to density as a measure of habitat quality, assuming no extensive lagged effects of previous browsing. Bjørneraas et al. (2012) included only density in their habitat-selection model and found a weak relationship with habitat selection. However, Solberg et al. (2012a,b) found a strong positive relationship between moose reproductive performance (twinning rate) and habitat quality (tree density and species) for a given density of moose among municipalities.

Within a MU, small-scale habitat selection occurs at the level of the forest stands (c. 1 km²). Because the number of hunting permits is scaled proportional to the area of forested land, harvest will reduce populations below carrying capacity more in low- than high-quality habitats, hence increasing the differences in density among habitats. This disrupts the relationship between habitat selection and abundance. However, movements can redistribute animals again according to habitat quality (sensu IFD). Large-scale habitat selection occurs when animals migrate across the boundaries of MUs, often with seasonal changes. The importance of these movements may depend on the relative size of the exclusive summer areas vs. the wintering areas or areas used year-round (Solberg et al. 2009). Several studies indicate that more moose aggregate in lower altitude valleys during winter in regions where the surrounding hills are extensive, receive much snow and therefore are used exclusively as summer ranges (e.g. Hjeljord 2001; Rolandsen et al. 2010). Due to seasonal migration, high-quality habitats in wintering areas will have high densities, possibly above the year-round carrying capacity. In the presence of lagged effects on forage availability (either

![Fig. 1. Yellowhead grizzly bear (Ursus arctos) habitat selection from a multi-seasonal resource selection function (RSF) summarized for each 49 km² (7 x 7 km) sampling cell (n = 164) based on the mean RSF at DNA hair-snag sites and illustrated as white (low RSF) to dark grey (high RSF) cells. Local abundance of grizzly bears detected per cell from the DNA hair-snag sites is shown as point locations increasing in size based on the number of detections from 0 (small) to ≥5 (large).](image)
The high abundance on the wintering range can carry over into the summer season. This will demonstrate a wedge-shaped (triangular) relationship with the resource selection functions (RSF) where maximum local abundance is related to habitat selection. Relationship between local abundance of bears detected per cell and mean RSF values at DNA hair-snag sites is illustrated for the 95th, 75th, and 50th centiles based on quantile regression of local abundance and mean RSF value.

For most vertebrate species tested in the Australian wet tropics (59 of 69 species), the relationship between local abundance and environmental suitability was a triangular wedge-shaped distribution where the upper limit of abundance was restricted in areas of low suitability and highest in sites of high suitability (VanDerWal et al. 2009). Thus, local abundance was consistently low in areas of low suitability and although abundance was highest in areas of highest suitability, selected habitats did not always have high density because a variety of factors can limit populations below their potential (Vanderwel, Malcolm & Caspersen 2012).

We illustrate such a wedge-shaped distribution for habitat selection and local abundance of grizzly bears measured from hair-snag DNA captures within 164 sampling cells (7 x 7 km) each containing seven-two-week sampling sessions during spring and early summer of 2004 (Fig. 2; sampling methods: Alberta Grizzly Bear Inventory Team 2005; Boulanger et al. 2006). RSFs were estimated using GPS radiotelemetry data from a similar time period (Nielsen, Boyce & Stenhouse 2006; Nielsen, Cranston & Stenhouse 2009). Quantile regression illustrated strong relationships between the upper limit (95th quantile) of local abundance and grizzly bear habitat selection ($\beta = 0.55$, SE = 0.09, $P < 0.001$; $R^2 = 0.32$) as compared to the 75th quantile ($\beta = 0.24$, SE = 0.07, $P = 0.001$; $R^2 = 0.12$) and 50th quantile ($\beta = 0.00$, SE = 0.04, $P = 1.00$; $R^2 = 0.00$) (Fig. 2). Factors affecting selection of habitats therefore relate to local abundance, but only the upper limits of potential abundance.

To recap, habitat is a primary determinant of distribution and abundance for most organisms. However, other processes can prevail such that densities in a particular habitat might not be reliable predictors of use in all circumstances. Also, reference populations at or near carrying capacity, that is where $dN/dt = 0$, should be the baseline for such habitat-based extrapolations of potential abundance. This reinforces the value of national parks and other natural areas to provide such baselines where as much as possible natural ecological processes are allowed to function with minimal human intervention (Boyce 1991).

**Conclusion**

Resource selection functions and related models of use and occupancy create a direct link between habitat and the distribution and abundance of animals. Such models must be used cautiously, however, because the assumption is that factors determining abundance are included as predictor covariates in the model, and that the model structure is appropriate for the data. This might be possible assuming that the ecology of the animal is understood sufficiently. Scale and domain of the sampling frame, both in time and space, are crucial considerations limiting application of these models. Ultimately, identifying reliable models for predicting abundance from habitat data require an understanding of the mechanisms underlying population limitation and regulation.

**Acknowledgements**

We thank Gordon Stenhouse from the Foothills Research Institute for access to data on grizzly bears. Thanks to D. Morris and G. Aarts for constructive comments on the manuscript. We acknowledge support from the Natural Sciences and Engineering Research Council of Canada and the Alberta Conservation Association. This work was presented in summer 2012 at a workshop hosted by Hedmark University College, campus Evenstad, and was co-funded by the Norwegian Research Council project 208434-F40 and by the COST Action IC0903 MOVE, 2009-2013.

**Data accessibility**

Data deposited in University of Alberta Dataverse Network doi: http://dx.doi.org/10.7939/DVN/10252 (Boyce et al. 2015).

**References**


