



# Agent-Based Inference for Animal Movement and Selection

Mevin B. HOOTEN, Devin S. JOHNSON, Ephraim M. HANKS, and John H. LOWRY

Contemporary ecologists often find themselves with an overwhelming amount of data to analyze. For example, it is now possible to collect nearly continuous spatiotemporal data on animal locations via global positioning systems and other satellite telemetry technology. In addition, there is a wealth of readily available environmental data via geographic information systems and remote sensing. We present a modeling framework that utilizes these forms of data and builds on previous research pertaining to the quantitative analysis of animal movement. This approach provides additional insight into the environmental drivers of residence and movement as well as resource selection while accommodating path uncertainty. The methods are demonstrated in an application involving mule deer movement in the La Sal Range, Utah, USA. Supplemental materials for this article are available online.

**Key Words:** Agent-based model; Change of support; Continuous model; Hierarchical; Bayesian model; Individual-based model.

## 1. INTRODUCTION

In a recent synthesis paper, Holyoak et al. (2008) reported that nearly 9,000 entries arose in a search for literature concerning the quantitative analysis of animal movement. A large portion of these are no doubt focused on population level, or Eulerian, movement (e.g., Turchin 1991, 1998; Wikle 2003), but several recent examples exist concerning individual-based movement modeling (e.g., Morales et al. 2004; Jonsen, Flemming, and Myers 2005; Tracey, Zhu, and Crooks 2005; Christ, Ver Hoef, and Zimmerman 2008; Horne et al. 2007; Barraquand and Benhamou 2008; Eckert et al. 2008; Johnson et al. 2008). Of those modeling approaches that are based on the analysis of animal paths, Barraquand and Benhamou

---

Mevin B. Hooten (✉) (E-mail: [mevin.hooten@usu.edu](mailto:mevin.hooten@usu.edu)) is Assistant Professor and Ephraim M. Hanks (E-mail: [ephraim.hanks@aggiemail.usu.edu](mailto:ephraim.hanks@aggiemail.usu.edu)) is Ph.D. Candidate, Department of Mathematics and Statistics, Utah State University, Logan, UT 84322-3900, USA. Devin S. Johnson is Statistician, NMML, National Oceanic and Atmospheric Administration, Seattle, WA 98115, USA (E-mail: [devin.johnson@noaa.gov](mailto:devin.johnson@noaa.gov)). John H. Lowry is Associate Director, RS/GIS Laboratory, Utah State University, Logan, UT 84322, USA (E-mail: [john.lowry@usu.edu](mailto:john.lowry@usu.edu)).

© 2010 International Biometric Society  
*Journal of Agricultural, Biological, and Environmental Statistics*, Volume 15, Number 4, Pages 523–538  
DOI: [10.1007/s13253-010-0038-2](https://doi.org/10.1007/s13253-010-0038-2)

(2008) note that they fall into two categories: egocentric and exocentric; that is, pertaining to either the animal's system of reference (e.g., move distances and turn angles) or on some other arbitrary spatial system of reference (e.g., changes in latitude or longitude), respectively. Both Morales et al. (2004) and Tracey, Zhu, and Crooks (2005) present methods that are exemplar of the egocentric approach, whereas, according to Barraquand and Benhamou (2008), the models employed by Jonsen, Myers, and Flemming (2003), Jonsen, Flemming, and Myers (2005) and Johnson et al. (2008) are classified as exocentric.

In terms of habitat or resource selection (e.g., Ramsey and Usner 2003; Johnson et al. 2008), traditional inference based on the analysis of animal paths follows one of two basic approaches: model-based methods (e.g., Christ, Ver Hoef, and Zimmerman 2008) and post hoc analyses (e.g., Barraquand and Benhamou 2008). More recently, Tracey, Zhu, and Crooks (2005) presented a model-based egocentric approach that links turn angles and move distances to known landscape features, and they are in the process of generalizing these methods to accommodate covariates and multi-animal inference using artificial neural networks (Tracey, Zhu, and Crooks 2010). These methods, along with ongoing improvements in telemetry technology, are a boon for animal ecologists who are able to obtain spatial location data for known animals at relatively high temporal resolutions and regularity. Jonsen, Flemming, and Myers (2005) deal with temporal irregularity by explicitly building it into the likelihood portion of a Markovian state-space movement model. Johnson et al. (2008) build on this idea in a continuous-time setting where directional persistence and velocity are modeled by an Ornstein–Uhlenbeck process. These models are also capable of accommodating measurement uncertainty, which is common, but varying in severity, in all telemetry data. Resulting inference can then be made on velocity and movement parameters, and recent extensions of these models allow one to characterize the posterior predictive distribution of the animal's path itself. This latter form of inference is especially appealing because it not only provides information concerning the probable space use of an individual but also information about the individual's location at any given time as well as the innate uncertainty associated with it.

In developing the methods presented herein, we set out to provide a means for answering scientific questions pertaining to animal movement in heterogeneous landscapes that contains several critical features:

- **Scientific:** In order to address specific scientific questions about selection and response to various landscape features, models should be able to accommodate explicit scientific structure.
- **Intuitive:** If these models are to be used by researchers and managers in scientific fields, resulting statistical inference needs to be readily interpretable.
- **Retrospective:** Much important research concerning animal movement has been conducted and needs to be incorporated into future modeling efforts.
- **Robust:** The framework should be robust, both scientifically and algorithmically, so that it is applicable in many situations (e.g., aquatic and terrestrial) and is computationally stable with minimal user tuning required. It should also be able to accommodate common forms of both telemetry and spatial covariate data.

- **Extendible:** Any models constructed for studying individual animal movement should be able to be generalized to handle more complex situations or additional scales of inference.

Using the above list as a guideline, in what follows, we present a general modeling framework that relies on telemetry data, contains both ego- and exocentric components, and is harmonious with many of the previous research efforts described above. To this end, we formulate our model using notation similar to that of Johnson et al. (2008) and rely on a change of spatiotemporal support to match the scale of commonly available gridded covariate information. We then couch the dynamics of the system in an agent-based specification (e.g., Grimm et al. 2005) that yields a model-based approach for studying resource selection. To illustrate the methods, we focus the application on characterizing the seasonal migrations of mule deer (*Odocoileus hemionus*) in the La Sal Mountain Range, Utah, USA.

## 2. METHODS

### 2.1. CONTINUOUS STATISTICAL MODEL

To begin, we introduce notation for the observed spatial locations of individuals over time (obtained via telemetry techniques). Let the location of an individual at time  $t$  be defined as  $\mathbf{s}_t \equiv (s_{1,t}, s_{2,t})'$  and the collection of such measurements be denoted as  $\mathbf{S}$ , a  $T \times 2$  matrix, corresponding to a total of  $T$  measured times. In a general sense, the methods presented by Jonsen, Flemming, and Myers (2005), Christ, Ver Hoef, and Zimmerman (2008), and Johnson et al. (2008) allow one to make inference on model parameters ( $\theta$ ) as well as the complete path itself ( $\tilde{\mathbf{S}}$ ); where  $\tilde{\mathbf{S}}$  is an infinite dimensional set of spatial locations in theory, but a large finite set, in practice. In particular, Johnson et al. (2008) provide a means of finding the posterior distribution of the latent state variables and parameters given the observed location data:  $[\theta|\mathbf{S}]$ , where the standard bracket notation  $[\cdot]$  denotes a probability distribution. Additionally, recent developments have also allowed for the estimation of the posterior predictive path distribution:  $[\tilde{\mathbf{S}}|\mathbf{S}]$ . In practice, realizations from this latter distribution can be obtained for any finite set of times in the data interval. This provides a means for finding a quasi-continuous path realization at any desired level of temporal resolution (Figure 1). In essence, this amounts to a model-based data augmentation.

### 2.2. AGENT-BASED STATISTICAL MODEL

Given that many forms of landscape-level covariate data are available in a gridded format (e.g., remotely sensed imagery, land use and land cover grids, and digital elevation models), we will specify the dynamics of the underlying movement process directly on the same areal spatial support as the covariates. In doing so, we consider the problem from an occupancy modeling perspective (e.g., MacKenzie et al. 2002, 2003; Royle and Kery 2007; Royle and Dorazio 2008), where, if the animal of interest occupies cell  $i$  at time  $t$  (i.e.,  $\mathbf{s}_t \in \text{cell}_i$ ), we let  $y_{i,t} = 1$ . In general, we could write this as a transformation of the path  $\mathbf{S}$  to a finite set of occupancy states  $\mathbf{Y} = \mathbf{Y}(\mathbf{S})$ . The issue that arises when inducing change of support through this transformation is that there is some loss of information about the

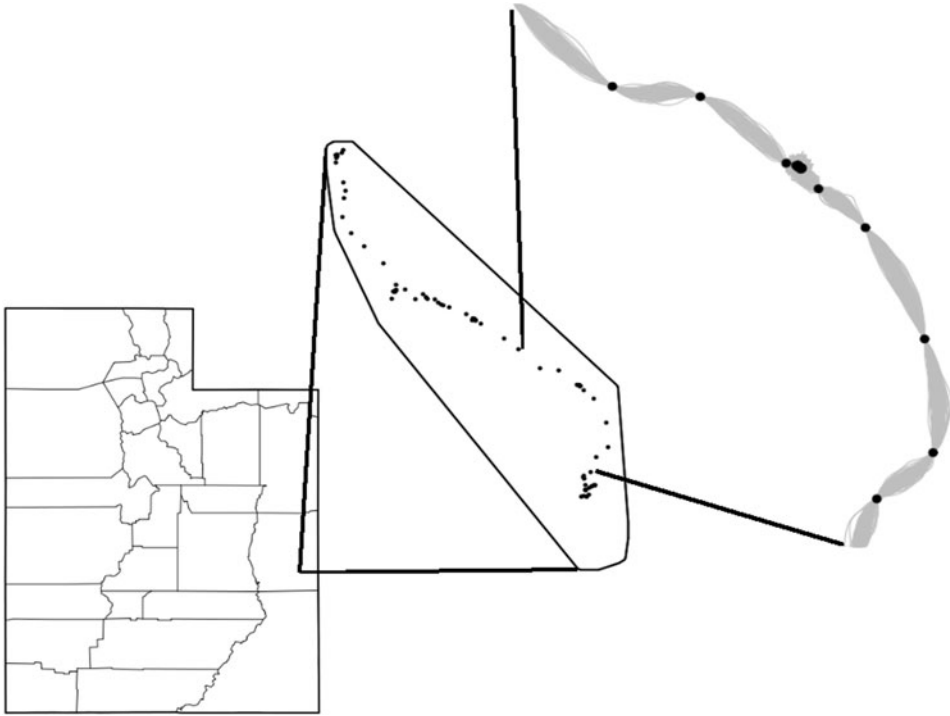


Figure 1. Telemetry data ( $\mathbf{S}$ , black points) for male mule deer in Castle Valley, Utah. Posterior predictive distribution ( $[\tilde{\mathbf{S}}|\mathbf{S}]$ , gray lines, inset) for section of path. Distribution represented as an finite ensemble of posterior predictive realizations.

animal's path when scaling up in this manner. In other words, the inverse transformation ( $\mathbf{Y}^{-1}$ ) does not exist since there is no way to recover  $\mathbf{S}$  from  $\mathbf{Y}$ . However, our desired inference is on the scale of the covariates, so this does not present a problem. Another issue that arises is that we do not have a completely observed path,  $\mathbf{S}$ . We do, however, have the posterior predictive realizations  $\tilde{\mathbf{S}}$ , which can provide a spatially contiguous discrete path on the support of the covariates (i.e.,  $\mathbf{Y} = \mathbf{Y}(\tilde{\mathbf{S}}) = \{y_{i,t}, \forall i, t\}$ , where  $y_{i,t} \in \{0, 1\}$ ). In fact, because the path cannot technically pass through the corner of a cell, the movement of the animal on this discrete spatial domain can be characterized by a first-order neighborhood of cells (i.e., the rook's neighborhood, Figure 2).

In an Eulerian occupancy model, with a widespread phenomenon being modeled across the spatial domain over time, one may want to use the entire set of occupancy states (e.g., Royle and Kery 2007; Hooten and Wikle 2010). However, with occupancy for a single animal arising in a small portion of the total spatial domain of interest, we seek a more parsimonious conditional representation of movement. That is, rather than consider the occupancy status for each cell in the spatial domain ( $y_{i,t}, \forall i$ ) at each time, we consider the status of the first-order neighborhood (a 5-dimensional vector), in a Markovian fashion, conditioned on the location of the previously occupied cell:  $\mathbf{y}_{\mathcal{N}_i,t} | \{y_{i,t-1} = 1\}$ . For notational convenience we define the first-order neighborhood on a regular grid oriented such

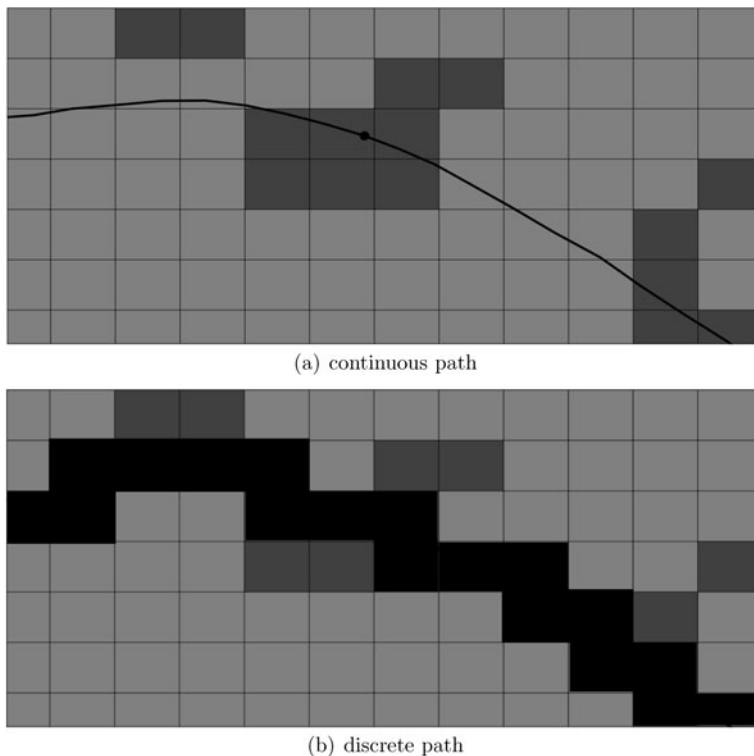


Figure 2. Change of Spatiotemporal Support. (a) A continuous posterior predictive path realization ( $\tilde{S}$ , black line) overlaid on the discrete spatial landscape with two habitat types. (b) The resulting discretized path ( $y_{i,t}$ , black cells) overlaid on the discrete spatial landscape with two habitat types.

that North is up:

$$\mathcal{N}_i = \begin{bmatrix} & N_2 & \\ N_1 & N_3 & N_5 \\ & N_4 & \end{bmatrix}_i = \begin{bmatrix} & N & \\ W & O & E \\ & S & \end{bmatrix}_i .$$

Using this specification, if the animal does not move out of cell  $i$  from time  $t - 1$  to  $t$ , then  $\mathbf{y}_{\mathcal{N}_i,t} = (0, 0, 1, 0, 0)'$ . Conversely, if the animal moves to the northern neighbor, for example, then  $\mathbf{y}_{\mathcal{N}_i,t} = (0, 1, 0, 0, 0)'$ . Thus, the  $\mathbf{y}_{\mathcal{N}_i,t}$  represent conditional multinomial response variables with transition probabilities  $\mathbf{p}_{\mathcal{N}_i,t}$ . These probabilities control the directional movement and can vary by location (i.e., non-stationarity). Now the dynamics of movement are Lagrangian (or agent-based, Grimm et al. 2005; Grimm and Railsback 2005) and can be readily interpreted as transition probabilities. Hooten and Wikle (2010) modeled the probabilities with a Dirichlet distribution that depended on an underlying movement potential surface, whereas, in a similar situation, Royle and Dorazio (2008, Chapter 9) suggest a general linear model (GLM) specification due to ease of implementation.

In what follows, we strike a compromise between these approaches and take advantage of conjugate full-conditional distributions in a robust Bayesian Gibbs sampling algorithm. Our methods primarily differ from those of Hooten and Wikle (2010) and Royle and Do-

razio (2008) in that we are focused on an individual rather than a population. However, we do utilize a latent Gaussian process to connect individual movement with a set of spatial covariates (not unlike the approaches discussed in Royle and Dorazio 2008) and we also use derivatives of environmental surfaces to separate different features of movement (similarly to the approach described in Hooten and Wikle 2010).

### 2.3. ENVIRONMENTAL DRIVERS OF MOVEMENT

Before specifying the dynamic process model that will allow for movement inference, we will take this opportunity to explain how scientific concepts can be directly incorporated into these models. In doing so, consider a decomposition of the dynamic process into two broad components: residence and movement. In turn, let each of these be decomposed further into two additional components: static and potential drivers:

1. Static Residence ( $\alpha_1$ ): Residence time depends on the environment in cell  $i$ .
2. Residence Potential ( $\alpha_2$ ): Residence time depends on convexity or concavity of the neighborhood of cell  $i$ .
3. Static Movement ( $\gamma_1$ ): Movement direction depends on the environment in cell  $i$ .
4. Movement Potential ( $\gamma_2$ ): Movement direction depends on the environmental gradient in the neighborhood of cell  $i$ .

Thus, consider each of the four drivers of residence and movement as spatial fields or surfaces that influence how long an animal occupies an area and then which way it goes when it leaves that area. The last field ( $\gamma_2$ ) in particular was considered the dominant driver of epidemic spread by Hooten and Wikle (2010), and can be conceptualized as a wavy surface that, if a ball were placed on, would tend to roll downhill and along corridors as well as around barriers.

Now, in order to provide a means for scientific inference, we need to link each of these drivers to a set of environmental covariates ( $\mathbf{X} = \{\mathbf{x}_i, \forall i\}$ , where  $\mathbf{x}_i$  are the covariates for cell  $i$ ) and then link the dynamic process of animal movement to the drivers. To achieve this, we introduce a set of latent auxiliary variables  $\mathbf{Z}$  such that

$$E(z_{i,j}) = \begin{cases} \alpha_{1,i} + \alpha_{2,i} & i = j \\ \gamma_{1,j} + \nabla\gamma_{2,i} & i \neq j, \end{cases}$$

where we drop the temporal notation involving  $t$  and  $t - 1$ , and without loss of generality, use the indices  $(i, j)$  to correspond to a move from cell  $i$  to its  $j$ th neighbor, for each sequentially occupied cell  $i$ . Recall that, when  $i = j$ , the animal resides in cell  $i$  during the time period  $(t - 1, t]$ . Then, in connecting the drivers to the covariates, let  $\alpha_{1,i} = \beta_{\alpha_0} + \mathbf{x}'_i \boldsymbol{\beta}_{\alpha_1}$ ,  $\alpha_{2,i} = \sum_{k \neq i} (\mathbf{x}'_k - \mathbf{x}'_i) \boldsymbol{\beta}_{\alpha_2}$ ,  $\gamma_{1,j} = \mathbf{x}'_j \boldsymbol{\beta}_{\gamma_1}$ , and  $\nabla\gamma_{2,i} = (\mathbf{x}'_j - \mathbf{x}'_i) \boldsymbol{\beta}_{\gamma_2}$ , where the movement potential field is  $\gamma_2 = \mathbf{X} \boldsymbol{\beta}_{\gamma_2}$ .

## 2.4. LATENT PROCESS MODEL

Using the above definitions and Gaussian assumptions on the auxiliary variables (for  $z_{i,j} \in \mathbb{R}$ ) we can now write

$$\begin{aligned} z_{i,j} &= 1_{i=j} \beta_{\alpha_0} + 1_{i=j} \mathbf{x}'_i \boldsymbol{\beta}_{\alpha_1} + 1_{i=j} \sum_{k \neq i} (\mathbf{x}'_k - \mathbf{x}'_i) \boldsymbol{\beta}_{\alpha_2} \\ &\quad + 1_{i \neq j} \mathbf{x}'_j \boldsymbol{\beta}_{\gamma_1} + 1_{i \neq j} (\mathbf{x}'_j - \mathbf{x}'_i) \boldsymbol{\beta}_{\gamma_2} + \varepsilon_{i,j} \\ &= \mathbf{q}'_{i,j} \boldsymbol{\beta} + \varepsilon_{i,j}, \end{aligned} \quad (2.1)$$

where  $\mathbf{q}'_{i,j} = (1_{i=j}, 1_{i=j} \mathbf{x}'_i, 1_{i=j} \sum_{k \neq i} (\mathbf{x}'_k - \mathbf{x}'_i), 1_{i \neq j} \mathbf{x}'_j, 1_{i \neq j} (\mathbf{x}'_j - \mathbf{x}'_i))$ , and the vector of regression coefficients is grouped in  $\boldsymbol{\beta} = (\beta'_{\alpha_0}, \boldsymbol{\beta}'_{\alpha_1}, \boldsymbol{\beta}'_{\alpha_2}, \boldsymbol{\beta}'_{\gamma_1}, \boldsymbol{\beta}'_{\gamma_2})'$ , and the terms  $1_{i=j}$  and  $1_{i \neq j}$  are indicator variables. The model specification in (2.1) can be condensed to a move-by-move form if we let  $\mathbf{z}_i = (z_{i,1}, z_{i,2}, z_{i,3}, z_{i,4}, z_{i,5})'$  be a collection of auxiliary variables for each cell in the neighborhood of cell  $i$ :

$$\mathbf{z}_i = \mathbf{Q}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad (2.2)$$

where  $\mathbf{Q}_i = (\mathbf{q}'_{i,1}, \mathbf{q}'_{i,2}, \mathbf{q}'_{i,3}, \mathbf{q}'_{i,4}, \mathbf{q}'_{i,5})'$  and  $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ . The covariance matrix  $\boldsymbol{\Sigma}$  then controls the correlation between movement directions and could be simplified as  $\boldsymbol{\Sigma} = \mathbf{I}$ .

Finally, letting  $\mathbf{z} = (\mathbf{z}'_1, \dots, \mathbf{z}'_\tau)'$ ,  $\mathbf{Q} = (\mathbf{Q}'_1, \dots, \mathbf{Q}'_\tau)'$ , and  $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \dots, \boldsymbol{\varepsilon}'_\tau)'$ , for a total of  $\tau$  augmented time points, we have

$$\mathbf{z} = \mathbf{Q} \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.3)$$

where  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I} \otimes \boldsymbol{\Sigma})$ . Now one can see that the latent process model involving the unobserved auxiliary variables  $\mathbf{z}$  takes on a standard linear form (2.3) and involves the four spatial drivers of residence and movement through the environmental covariates ( $\mathbf{X}$ ) and a set of regression coefficients  $\boldsymbol{\beta}$ .

## 2.5. HIERARCHICAL MODEL

### 2.5.1. Fully Observed Path

In completing the hierarchical Bayesian model formulation (e.g., Cressie et al. 2009), we would need to specify a model linking the multinomial cell probabilities to the auxiliary variables and then put priors on the model parameters. Assuming for now that we only have a single path realization  $\mathbf{Y}(\tilde{\mathbf{S}})$ , we can adopt the computationally robust algorithm presented by Albert and Chib (1993) to implement the model. Specifically, if  $\mathbf{Y}$  is a  $\tau \times 5$  matrix, we can formulate the model such that the multinomial cell probabilities, corresponding to the  $i$ th row of  $\mathbf{Y}$ , are  $p_{i,j} = P(z_{i,j} > z_{i,k}, \forall k \neq j)$ . Then, letting the  $z_{i,j}$  be defined precisely as in the previous section yields a very computationally tractable posterior distribution with conjugate full-conditionals (see Supplemental Materials):

$$[\boldsymbol{\beta}, \mathbf{Z} | \mathbf{Y}(\tilde{\mathbf{S}})] \propto [\mathbf{Y}(\tilde{\mathbf{S}}) | \mathbf{P}(\mathbf{Z})][\mathbf{Z} | \boldsymbol{\beta}][\boldsymbol{\beta}], \quad (2.4)$$

where  $\mathbf{P}$  is a  $\tau \times 5$  matrix of multinomial cell probabilities,  $\mathbf{Z}$  is a  $\tau \times 5$  matrix of the auxiliary variables, and  $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ .

This model certainly allows one to make inference on the importance of the covariates as well as the residence and movement fields ( $\alpha_1, \alpha_2, \gamma_1, \gamma_2$ ), but it does not accommodate the uncertainty in the path  $\tilde{\mathbf{S}}$ .

### 2.5.2. Incorporating the Path Distribution

To properly account for uncertainty when making inference on model parameters, we seek the posterior distribution of the coefficients ( $\beta$ ), and the auxiliary variables ( $\mathbf{Z}$ ) given the *observed* data  $\mathbf{S}$  rather than the *augmented* data  $\tilde{\mathbf{S}}$ . Given that the prior model output (Johnson et al. 2008) includes the posterior predictive path distribution  $[\tilde{\mathbf{S}}|\mathbf{S}]$ , we merely need to integrate over it to find the correct posterior:

$$\begin{aligned} [\beta, \mathbf{Z}|\mathbf{S}] &= \int [\beta, \mathbf{Z}|\mathbf{Y}(\tilde{\mathbf{S}})][\tilde{\mathbf{S}}|\mathbf{S}] d\tilde{\mathbf{S}} \\ &\propto \int [\mathbf{Y}(\tilde{\mathbf{S}})|\mathbf{P}(\mathbf{Z})][\mathbf{Z}|\beta][\beta][\tilde{\mathbf{S}}|\mathbf{S}] d\tilde{\mathbf{S}}. \end{aligned} \quad (2.5)$$

In practice, the implementation of (2.5) is a trivial extension of (2.4) because all parameters remain conjugate and we perform the necessary integration by drawing a realization of  $\tilde{\mathbf{S}}$  from its posterior predictive distribution  $[\tilde{\mathbf{S}}|\mathbf{S}]$  on each iteration of the Gibbs sampler; then we condition on it for the remainder of the full-conditional samples. Using the auxiliary variable approach (Albert and Chib 1993), we never actually need to sample the multinomial cell probabilities  $\mathbf{P}$ ; however, if inference is desired directly on the transition probabilities themselves, they could be approximated numerically using the posterior  $\mathbf{Z}$ . Performing inference directly on the  $\mathbf{Z}$  is possible, but not intuitive, and thus not commonly done, as these latent variables are often thought of as nuisance parameters.

The approach described above is quite similar to data augmentation (Tanner and Wong 1987) where one would sample an augmented data set from the current posterior predictive distribution in the algorithm. In our case, we cannot obtain  $\tilde{\mathbf{S}}$  directly from the current posterior, due to the inability to invert  $\mathbf{Y}$ . However, we can use the results of the previous model fit that depend directly on  $\mathbf{S}$ . Thus, we are conditioning on the originally observed path while incorporating previous model results and their inherent uncertainty. In a certain sense, this approach consists entirely of posterior predictive inference based on the methods of Johnson et al. (2008).

## 3. APPLICATION: MULE DEER MOVEMENT AND SELECTION

To illustrate the utility of the methods presented above as well as forms of possible inference, we fit models to global positioning system (GPS) data representing the spring and fall migrations of a male mule deer (*Odocoileus hemionus*) in the La Sal Mountain Range, Utah, USA (McFarlane 2007). Mule deer movement in this region is an excellent candidate for the methods presented herein because the data are collected with temporal irregularity and the study species exhibits complex movement, much of which may be due to its spatially heterogeneous environment (D'Eon and Serrouya 2005). Additionally, mule deer are one of the primary hosts for chronic wasting disease (CWD) in the western United States (Miller, Hobbs, and Tavener 2006). CWD is caused by a contagious prion-



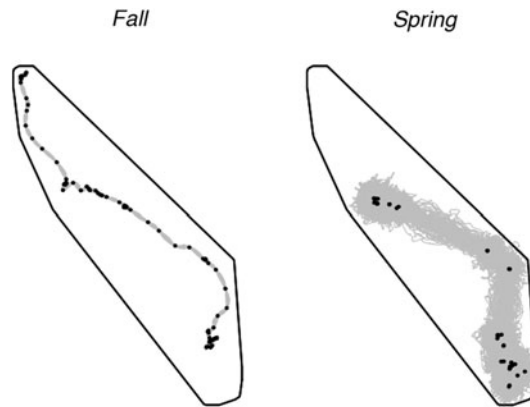


Figure 3. GPS data (black points) and posterior predictive realizations (gray lines), for the fall and spring male mule deer paths. Fall migration moves generally from southeast to northwest. Spring migration moves generally from northwest to southeast.

based pathogen that is capable of persistence in the environment and thus does not need animal contact to spread (Miller, Hobbs, and Tavener 2006). Characterizing the movement behavior of mule deer in various environments is a critical first step in understanding the epidemiology of CWD.

In the study conducted by McFarlane (2007), GPS collars were set to transmit location data more frequently during certain scientifically important periods of the year. For example, during the breeding season (1 November to 15 December), location data are available every 30 minutes, whereas from 1 January to 15 March, data are only available every 12 hours. To capture the behavior during the fall migration for a male mule deer in 2005, we focused on the period from 24 October to 27 October (in 30-minute intervals), and also 17 May to 28 May (in 6-hour intervals) for the spring migration the same year (Figure 3). For spatial covariates, we used land cover data from the National Land Cover Dataset 2001 (Homer et al. 2007) and obtained elevation data from the USGS National Elevation Dataset. In addition to the three dominant land cover types in the study area, slope (degree) and solar exposure (cosine aspect transformation) were derived from elevation (Figure 4). For this study, the grid cell size was increased from 30 meters squared to 300 meters squared and thus the land cover types were represented as a percentage of the given land cover within each cell. These covariates ( $\mathbf{X}$ ) were then centered and scaled so that the relative importance of the estimated coefficients could be assessed visually.

The model described above, with the four drivers of movement specified as in Section 2.3, was fit to both sets of data using the covariates ( $\mathbf{X}$ ) and Markov Chain Monte Carlo (MCMC). In fitting the model, we found that convergence occurred rapidly and trace plots indicated well-mixing Markov chains. Specifically, in this application, we ran the Gibbs sampler for 10,000 MCMC iterations and discarded the first 1,000 as the burn-in period. A relatively vague Gaussian prior was used for  $\beta$  with mean zero and variance  $100(\mathbf{Q}'\mathbf{Q})^{-1}$ . We have made the data, computer code, and necessary instructions to replicate the results available at the following URL: <http://www.math.usu.edu/~hooten/other/>. Additionally, we have conducted a simulation study to verify that the model is able to cor-

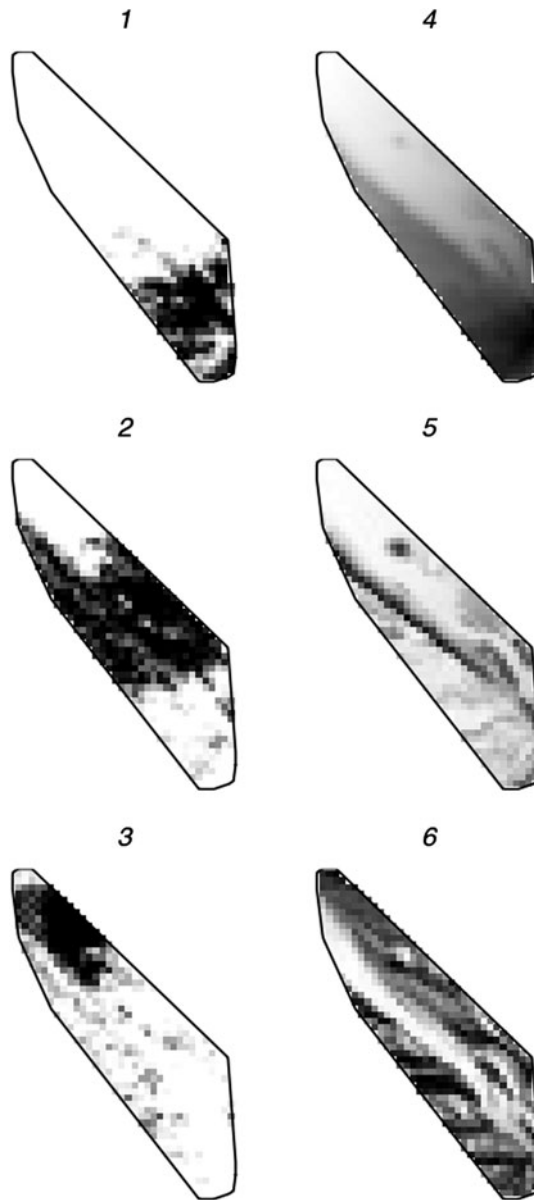


Figure 4. Spatial covariates for male mule deer fall and spring migrations; grayscale represents covariate intensity with dark shades corresponding to high values. 1—Deciduous forest; 2—Evergreen forest; 3—Scrub/Shrub; 4—Elevation; 5—Slope; 6—Solar exposure (aspect transformations such that southwest exposures are high values).

rectly recover the coefficients ( $\beta$ ) pertaining to residence and movement (see Supplemental Materials).

The resulting posterior distributions for the residence and movement parameters ( $\beta_{\alpha_1}, \beta_{\alpha_2}, \beta_{\gamma_1}, \beta_{\gamma_2}$ ) are presented in Figure 5, while the posterior mean fields for the residence and movement drivers ( $\alpha_1, \alpha_2, \gamma_1, \gamma_2$ ) are presented in Figure 6. Also, to assess the

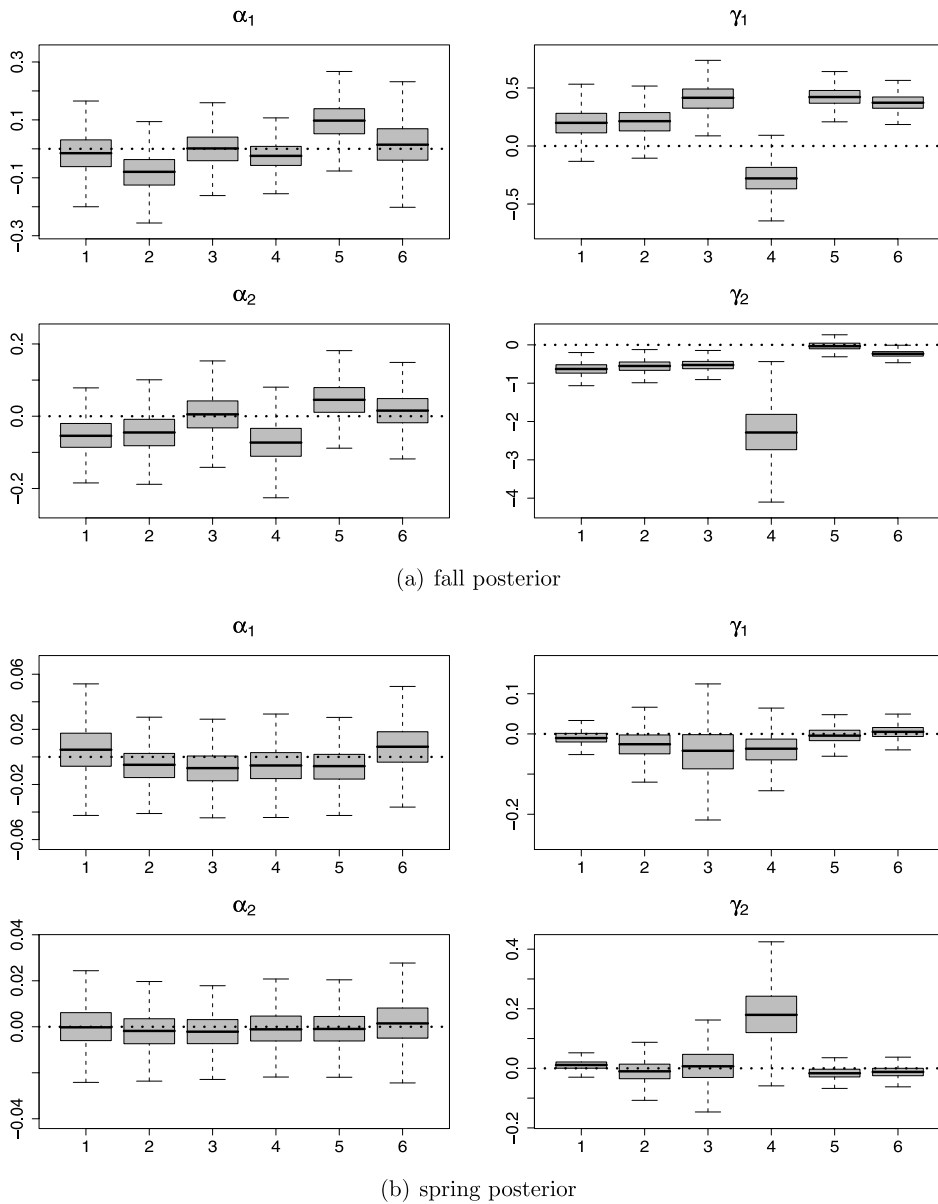


Figure 5. Boxplots summarizing the posterior distributions for the movement parameters in terms of each set of covariates. Recall that the corresponding covariates are described in Figure 4 and were standardized so that these coefficients can be compared directly on the same scale. (a) Fall migration; (b) Spring migration.

reliability of our posterior estimates (as suggested by Flegel, Haran, and Jones 2008 and Jones et al. 2006), we also computed the MCMC standard errors for each coefficient in  $\beta$ . The asymptotic 95% confidence intervals for these quantities suggest that our posterior mean estimates are accurate to at least 4 digits.

In terms of statistical inference, we can evaluate the effect of the heterogeneous environment on movement in two ways: using the coefficients ( $\beta_{\alpha_1}, \beta_{\alpha_2}, \beta_{\gamma_1}, \beta_{\gamma_2}$ ) and using

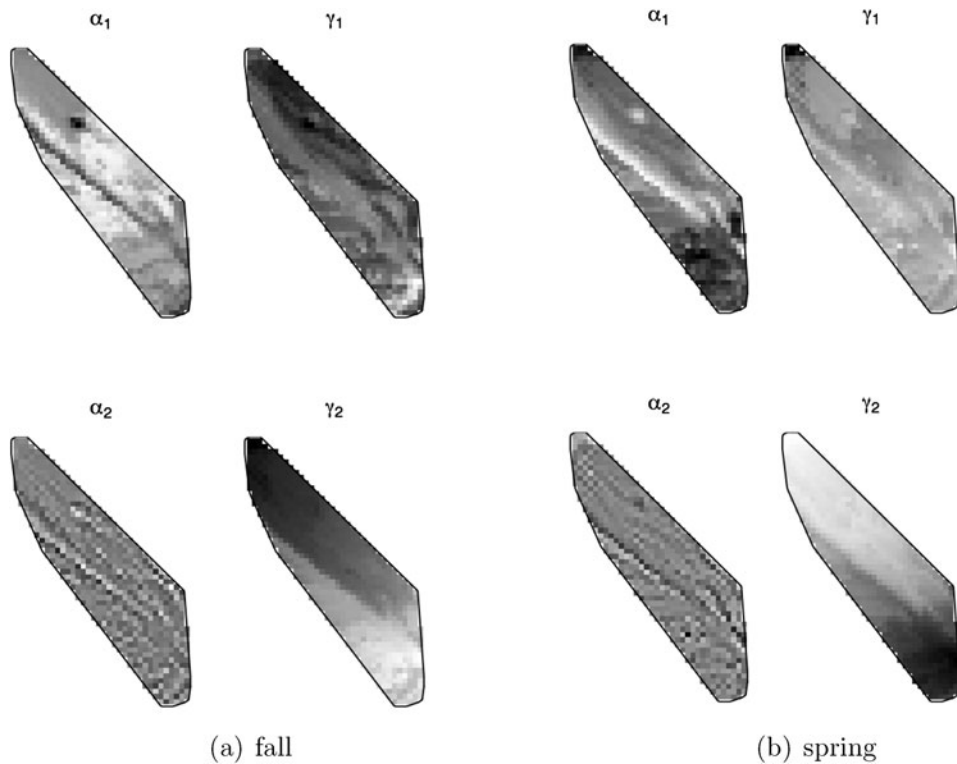


Figure 6. Posterior expectations for the static residence field ( $\alpha_1$ ), residence potential field ( $\alpha_2$ ), static movement field ( $\gamma_1$ ), and the movement potential field ( $\gamma_2$ ). Grayscale represents posterior expectation with dark shades corresponding to high values. (a) Fall migration. (b) Spring migration.

the residence and movement fields ( $\alpha_1$ ,  $\alpha_2$ ,  $\gamma_1$ ,  $\gamma_2$ ). First, with regard to the coefficients, Figure 5 indicates that, in general, the posterior distributions for the fall migration are more significant than those resulting from the spring migration data. This illustrates one of the primary advantages to these methods, and is likely a direct result of the discrepancy in uncertainty regarding our knowledge of the actual spring path (Figure 3); though, of course, could be due to a clear difference in selection between the fall and spring migrations. To further examine the latter, a more temporally fine set of data would need to be collected. Additionally, in each of the migrations, the posterior coefficients show differences in both significance (i.e., distance from zero) and precision; this indicates that there may indeed be several drivers influencing residence and movement for this mule deer. Notably, in the fall migration, while the residence parameters ( $\beta_{\alpha_1}$ ,  $\beta_{\alpha_2}$ ) appear to be telling a similar story, the movement parameters ( $\beta_{\gamma_1}$ ,  $\beta_{\gamma_2}$ ) corresponding to the land cover types (1: Deciduous forest, 2: Evergreen forest, and 3: Scrub/Shrub) appear to have opposing influences in the static movement field ( $\gamma_1$ ) versus the movement potential field ( $\gamma_2$ ). The implications for movement due to this difference in the sign of the posterior coefficients are that given the presence of the other covariates in the model, each of the land covers positively influences movement direction ( $\beta_{\gamma_1}$ ) when the animal is in that land cover, with deciduous forest being the strongest influence. Conversely, for movement potential ( $\gamma_2$ ), since  $\beta_{\gamma_2}$  interacts

with the difference in environment between cell  $j$  and cell  $i$  (i.e.,  $\mathbf{x}_j - \mathbf{x}_i$ ), the strong negative coefficient corresponding to covariate one indicates that during the fall migration the mule deer tends to move out of areas with deciduous forest. Note that, in this study area, the deciduous forest land cover is also at higher elevations and thus the mule deer are fleeing colder temperatures as winter approaches. Similar inference can be made for the spring migration, though the results are less conclusive given the uncertainty in the path due to the coarser temporal data.

The residence and movement fields themselves (Figure 6) can also offer insight into the environmental drivers of resource selection for mule deer and how it differs between migrations. Overall, the residence and movement fields shown in Figure 6a (fall migration) clearly indicate differing spatial patterns than those in Figure 6b (spring migration). Take, for example, the contrasting static residence fields ( $\alpha_1$ ) for the fall and spring migrations. In this case, notice first that middle region of the study area is lighter shaded in the fall than in the spring but the southern region is darker shaded in the spring than in the fall. These regions correspond mainly to two different land covers and thus the scientific interpretation is that the male mule deer moves the quickest through the evergreen land cover in the fall and moves the slowest through the deciduous land cover in the spring. Similarly contrasting fields can be seen in the fall and spring movement potentials ( $\gamma_2$ ) which correspond strongly to the elevation covariate. These fields can be viewed as surfaces where mule deer will generally move from lighter areas to darker areas. With this in mind, the fall movement potential shows a general northwesterly movement is likely, while the spring movement potential shows the opposite is likely.

## 4. CONCLUSION

Given that numerous previous research efforts have described methods for characterizing animal movement, our goal was to incorporate as much of the existing methodology as possible while providing a new set of tools for analyzing animal response to their environment. This is achieved by specifying and implementing a model that relies directly on the results from a prior analysis using telemetry data in addition to new spatial covariate information. The modeling framework we presented herein depends on a change of spatiotemporal support (from continuous to discrete) in order to reconcile the differences between the prior model output and the most common form of large-scale covariate data. We then employ an agent-based specification that allows the data to inform dynamic parameters that have direct scientific interpretations on the spatial scale of the covariates.

As described in Section 1, other objectives of this work involved the methods being intuitive, robust, and extendible. Our model specification is purposefully Lagrangian in that it is individual-based and can be easily understood. We also specifically designed the model such that it would have conjugate full-conditional distributions (see Supplemental Materials) to ensure a robust computational implementation. Additionally, the model framework we present can be extended in numerous ways for other specific scientific situations. Several of the following extensions are the focus of ongoing research, but are mentioned here so that they may be considered in other projects as well.

Perhaps the simplest extension or modification would be a focus on different environmental drivers of residence and movement. For example, rather than specify the residence potential field ( $\alpha_2$ ) as the total convexity or concavity of a local neighborhood, one might want to consider the general tilt of the environmental surface. Moreover, it may be of interest to the user to perform some model reduction. That is, in situations where not all covariates ( $\mathbf{X}$ ) are important in each driver of movement, one may seek to perform variable selection to reduce the parameter space for improved inference. Given that we have specified the model as a form of multinomial regression, standard selection methods could be employed by modifying (2.1) and comparing model fits via the selection criterion of choice (e.g., deviance information criterion). Alternatively, it may be possible to take a model-based selection approach and implement it using reversible-jump MCMC or a similar method (e.g., O'Hara and Sillanpaa 2009). In some situations, it may become important to explicitly assess identifiability among parameters, in which case, the Bayesian learning methods based on differences in precision and Kullback–Leibler divergence proposed by Xie and Carlin (2006) could be considered.

Another extension could be made regarding the modeling of an ensemble of individual animals. The methods presented here can be directly employed in the case where telemetry data are available for multiple animals in the same study region. In this case, the model would be fit independently for each animal using the proximity to other tracked animals as covariates in  $\mathbf{X}$  (e.g., perhaps as a kernel density surface based on the other animals' location data). A formal extension of the model could be made to accommodate direct correlation between animals by augmenting (2.1) to include the auxiliary variables from other animals as autocovariates (e.g.,  $z_{i,j,k} = \sum_{l \neq k} z_{i,j,l} \beta_l + \mathbf{q}'_{i,j} \boldsymbol{\beta} + \varepsilon_{i,j}$ , for animal  $k$ ). Other extensions such as time-varying covariates, temporal change-points, and path feedback are also being investigated.

Overall, the modeling framework presented here is harmonious with previous advances in the study of animal movement and is general enough to allow for many possible extensions and improvements while being sufficiently intuitive and tangible for animal ecologists and resource managers.

## SUPPLEMENTAL MATERIALS

**Full-Conditionals:** Full conditional distributions for an MCMC implementation of the model.

**Simulation:** Simulation experiment to assess the model's ability to recover coefficients.

## ACKNOWLEDGEMENTS

This research was funded by USGS 1434-06HQRU1555 and NOAA AB133F-09-SE-4640. The authors would also like to thank Mary Conner, Leslie McFarlane, and the Utah Division of Wildlife Resources for providing data and many helpful suggestions.

## REFERENCES

- Albert, J. H., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.
- Barraquand, F., and Benhamou, S. (2008), "Animal Movements in Heterogeneous Landscapes: Identifying Profitable Places and Homogeneous Movement Bouts," *Ecology*, 89, 3336–3348.
- Christ, A., Ver Hoef, J., and Zimmerman, D. L. (2008), "An Animal Movement Model Incorporating Home Range and Habitat Selection," *Environmental and Ecological Statistics*, 15, 27–38.
- Cressie, N. A. C., Calder, C. A., Clark, J. S., Ver Hoef, J. M., and Wikle, C. K. (2009), "Accounting for Uncertainty in Ecological Analysis: The Strengths and Limitations of Hierarchical Statistical Modeling," *Ecological Applications*, 19, 553–570.
- D'Eon, R. G., and Serrouya, R. (2005), "Mule Deer Seasonal Movements and Multiscale Resource Selection Using Global Positioning System Radiotelemetry," *Journal of Mammalogy*, 86, 736–744.
- Eckert, S. A., Moore, J. E., Dunn, D. C., Sagarminaga Van Buiten, R., Eckert, K. L., and Halpin, P. N. (2008), "Modeling Loggerhead Turtle Movement in the Mediterranean: Importance of Body Size and Oceanography," *Ecological Applications*, 18, 290–308.
- Flegal, J. M., Haran, M., and Jones, G. L. (2008), "Markov Chain Monte Carlo: Can We Trust the Third Significant Figure?" *Statistical Science*, 23, 250–260.
- Grimm, V., and Railsback, S. F. (2005), *Individual-Based Modeling and Ecology*, Princeton, New Jersey: Princeton University Press.
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., and DeAngelis, D. L. (2005), "Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology," *Science*, 310, 987–991.
- Holyoak, M., Casagrandi, R., Nathan, R., Revilla, E., and Spiegel, O. (2008), "Trends and Missing Parts in the Study of Movement Ecology," *Proceedings of the National Academy of Sciences*, 105, 19060–19065.
- Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J. N., and Wickham, J. (2007), "Completion of the 2001 National Land Cover Database for the Conterminous United States," *Photogrammetric Engineering and Remote Sensing*, 73, 337–341.
- Hooten, M. B., and Wikle, C. K. (2010), "Statistical Agent-Based Models for Discrete Spatiotemporal Systems," *Journal of the American Statistical Association*, 105, 236–248.
- Horne, J. S., Garton, E. O., Krone, S. M., and Lewis, J. S. (2007), "Analyzing Animal Movements Using Brownian Bridges," *Ecology*, 88, 2354–2363.
- Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W. (2008), "Continuous-Time Correlated Random Walk Model for Animal Telemetry Data," *Ecology*, 89, 1208–1215.
- Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006), "Fixed-Width Output Analysis for Markov Chain Monte Carlo," *Journal of the American Statistical Association*, 101, 1537–1547.
- Jonsen, I. D., Myers, R. A., and Flemming, J. M. (2003), "Meta-Analysis of Animal Movement Using State-Space Models," *Ecology*, 84, 3055–3063.
- Jonsen, I. D., Flemming, J. M., and Myers, R. A. (2005), "Robust State-Space Modeling of Animal Movement Data," *Ecology*, 86, 2874–2880.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and Langtimm, C. A. (2002), "Estimating Site Occupancy Rates when Detection Probabilities are Less than One," *Ecology*, 83, 2248–2255.
- MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., and Franklin, A. B. (2003), "Estimating Site Occupancy, Colonization, and Local Extinction when a Species is Detected Imperfectly," *Ecology*, 84, 2200–2207.
- McFarlane, L. R. (2007), "Breeding Behavior and Space Use of Male and Female Mule Deer: An Examination of Potential Risk Differences for Chronic Wasting Disease Infection," M.S. Thesis, Utah State University, Dept. of Wildland Resources.
- Miller, M. W., Hobbs, N. T., and Tavener, S. J. (2006), "Dynamics of Prion Disease Transmission in Mule Deer," *Ecological Applications*, 16, 2208–2214.

- Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004), "Extracting More Out of Relocation Data: Building Movement Models as Mixtures of Random Walks," *Ecology*, 85, 2436–2445.
- O'Hara, R. B., and Sillanpaa, M. J. (2009), "A Review of Bayesian Variable Selection Methods: What, How and Which," *Bayesian Analysis*, 4, 85–118.
- Ramsey, F. L., and Usner, D. (2003), "Persistence and Heterogeneity in Habitat Selection Studies Using Radio Telemetry," *Biometrics*, 59, 332–340.
- Royle, J. A., and Dorazio, R. M. (2008), *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*, San Diego: Academic Press.
- Royle, J. A., and Kery, M. (2007), "A Bayesian State-Space Formulation of Dynamic Occupancy Models," *Ecology*, 88, 1813–1823.
- Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.
- Tracey, J. A., Zhu, J., and Crooks, K. (2005), "A Set of Nonlinear Regression Models for Animal Movement in Response to a Single Landscape Feature," *Journal of Agricultural, Biological, and Environmental Statistics*, 10, 1–18.
- Tracey, J. A., Zhu, J., and Crooks, K. R. (2010, in press), "Modeling and Inference of Animal Movement Using Artificial Neural Networks," *Environmental and Ecological Statistics*. doi:[10.1007/s10651-010-0138-8](https://doi.org/10.1007/s10651-010-0138-8).
- Turchin, P. (1991), "Translating Foraging Movements in Heterogeneous Environments into the Spatial Distribution of Foragers," *Ecology*, 72, 1253–1266.
- (1998), *Quantitative Analysis of Movement*, Sunderland, Massachusetts: Sinauer Associates, Inc. Publishers.
- Wikle, C. K. (2003), "Hierarchical Bayesian Models for Predicting the Spread of Ecological Processes," *Ecology*, 84, 1382–1394.
- Xie, Y., and Carlin, B. (2006), "Measures of Bayesian Learning and Identifiability in Hierarchical Models," *Journal of Statistical Planning and Inference*, 136, 3458–3477.