Environ Ecol Stat (2011) 18:393–410
DOI 10.1007/s10651-010-0138-8

Modeling and inference of animal movement using artificial neural networks

Jeff A. Tracey · Jun Zhu · Kevin R. Crooks

Received: 3 September 2009 / Revised: 2 February 2010 / Published online: 6 April 2010
© Springer Science+Business Media, LLC 2010

Abstract Movement of animals in relation to objects in their environment is important in many areas of ecology and wildlife conservation. Tools for analysis of movement data, however, still remain rather limited. In previous work, we developed nonlinear regression models for movement in relation to a single landscape feature. Here we greatly expand these previous models by using artificial neural networks. The new models add substantial flexibility and capabilities, including the ability to incorporate multiple factors and covariates. We devise a likelihood-based model fitting procedure that utilizes genetic algorithms and demonstrate the approach with movement data for red diamond rattlesnakes. The proposed methodology can be useful for global positioning system tracking data that are becoming more common in studies of animal movement behavior.

Keywords Circular Statistics · Genetic algorithm · Movement path · Semi-parametric model · Telemetry data · von Mises distribution

J. A. Tracey · K. R. Crooks
Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO, USA

J. A. Tracey
SigmaLogistic Consulting, Inc., San Diego, CA, USA

J. Zhu (✉)
Department of Statistics, Colorado State University, Fort Collins, CO, USA
e-mail: jzhu@stat.colostate.edu

J. Zhu
Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA

1 Introduction

Understanding the movement of individual animals in space is of great importance in many areas of ecology and wildlife conservation (Van Vuren 1998; Tracey 2006). Of particular interest here is *object orientation*, which refers to movement in relation to objects in a landscape (or landscape features; Jander 1975; Lima and Zollner 1996). Examples of objects are prey or den sites that may be represented by points, rivers or roads that may be represented by lines, and regions of urban development or resource patches that may be represented by polygons. Movement data for individual animals are often collected using field techniques such as radio-telemetry or global positioning system (GPS) tracking, where animal positions are acquired at regular or irregular time intervals. Data on landscape features are often collected via remote sensing or ground measurement using GPS or other techniques. For example, radio-telemetry data were collected to study the effects of roads and housing development on the movement of red diamond rattlesnakes (*Crotalus ruber*) in San Diego, California (Tracey 2000). Figure 1 shows the boundaries between a habitat patch and regions of urban development (or, *patch boundaries*). Within the patch are movement paths of four individual rattlesnakes. It is well understood, at least qualitatively, that an animal may respond to objects through movement via repulsion or attraction (Jander 1975; Lima and Zollner 1996). However, it remains a challenge to describe object orientation quantitatively based on empirical data.

Most models of individual movement found in the ecological literature, including simple random walk, correlated random walk, and directional bias models, do not describe movement in relation to landscape features (Batschelet 1981; Kareiva and Shigesada 1983; Turchin 1998). In recent developments, more sophisticated animal

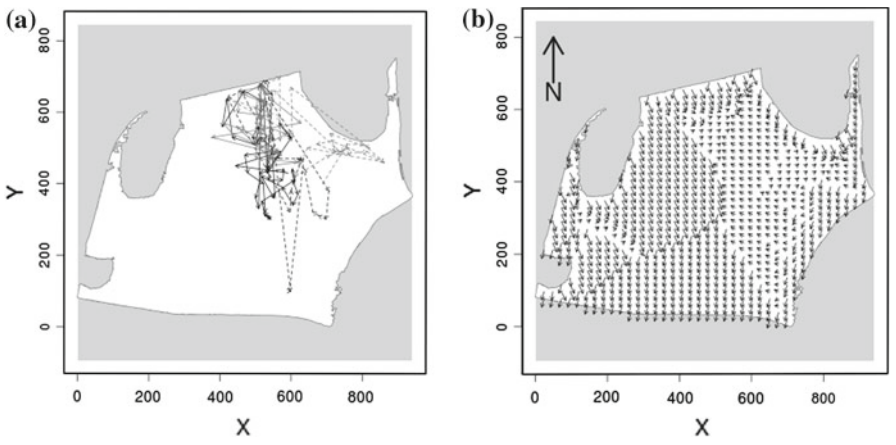

Fig. 1 **a** Observed movement paths for four adult male red diamond rattlesnakes in a habitat patch isolated by roads and urban development. The *solid black line* corresponds to rattlesnake $i = 1$, the *dashed black line* to $i = 2$, the *solid gray line* to $i = 3$, and the *dashed gray line* to $i = 4$. **b** Responses from the best model (Model #17, Table 1) on a lattice of locations within the habitat patch. The angle of an *arrow* indicates the fitted mean angle (of movement) and the length of an *arrow* is proportional to the fitted concentration parameter (strength of the tendency to move in the mean angle) of the von Mises distribution

Table 1 AIC_c values of neural network models responsive to the patch boundaries

Model	Hidden nodes	Hidden thresholds	Output thresholds	AIC_c	ΔAIC_c	Akaike weight
Individual-adjusted						
1	2	Yes	Yes	517.24	9.78	6.27E-003
2	3	Yes	Yes	523.01	15.56	3.49E-004
3	4	Yes	Yes	547.79	40.33	1.45E-009
4	5	Yes	Yes	566.38	58.92	1.34E-013
5	2	Yes	No	520.08	12.63	1.51E-003
6	3	Yes	No	520.70	13.24	1.11E-003
7	4	Yes	No	545.82	38.37	3.89E-009
8	5	Yes	No	570.56	63.10	1.66E-014
9	2	No	Yes	527.72	20.27	3.31E-005
10	3	No	Yes	528.30	20.84	2.49E-005
11	4	No	Yes	528.94	21.48	1.80E-005
12	5	No	Yes	558.56	51.10	6.67E-012
13	2	No	No	523.30	15.84	3.03E-004
14	3	No	No	522.27	14.82	5.05E-004
15	4	No	No	533.87	26.42	1.53E-006
16	5	No	No	561.35	53.90	1.65E-012
Population-level						
17	2	Yes	Yes	507.46	0.00	8.34E-001
18	3	Yes	Yes	516.72	9.26	8.11E-003
19	4	Yes	Yes	516.41	8.95	9.49E-003
20	5	Yes	Yes	536.55	29.09	4.02E-007
21	2	Yes	No	511.83	4.37	9.38E-002
22	3	Yes	No	522.19	14.74	5.26E-004
23	4	Yes	No	522.59	15.14	4.31E-004
24	5	Yes	No	529.91	22.46	1.11E-005
25	2	No	Yes	514.91	7.46	2.00E-002
26	3	No	Yes	521.15	13.70	8.85E-004
27	4	No	Yes	520.28	12.83	1.37E-003
28	5	No	Yes	527.97	20.52	2.92E-005
29	2	No	No	515.02	7.57	1.90E-002
30	3	No	No	521.69	14.24	6.75E-004
31	4	No	No	519.95	12.50	1.61E-003
32	5	No	No	529.84	22.38	1.15E-005

movement models are emerging that combine movement, resource selection, and home range of an animal (see, e.g., Brillinger et al. 2004; Christ et al. 2008; Forester et al. 2009; Hooten et al. 2010; Johnson et al. 2008). Some of the movement models involved are formulated using kernel densities. However, it is not always straightforward to assess the effect of landscape features directly and computation for statistical inference

tends to be intensive. See also Patterson et al. (2008) for an excellent description of a general modeling framework, namely state-space models, for animal movement.

Tracey et al. (2005) developed a set of nonlinear regression models for the analysis of individual animal movement data in relation to a single type of landscape feature. The main idea was to let the parameters of the von Mises distribution change based on the animal's location relative to the location of the object. The von Mises distribution for move angles has a concentration parameter that changes as a function of the animal-to-object distance and a mean angle that changes as a function of the animal-to-object angle. Although this approach allows for nonlinear models, the forms of nonlinearity are restricted to exponential and logistic functions, which may not always be suitable for complex behavioral responses. In addition, the models are limited to movement in relation to only one landscape feature at a time and can not be extended to accommodate multiple objects on the landscape simultaneously. Furthermore, this methodology is individual-based, quantifying the movement of one individual animal at a time. Although animal movement is inherently an individual-level process, often data consist of many observations for an individual animal and a small number of individuals, which provide opportunities to draw inference not only for the individual movement, but also at a population level.

In this paper, we consider artificial neural networks (ANNs) as a far more flexible and powerful alternative to the models featured in Tracey et al. (2005). Neural networks are composed of interconnected computational units called *nodes* that are collectively more powerful than the individual nodes (Reed and Marks 1999). As a result, we broaden the scope of Tracey et al. (2005) substantially in several important directions. First, ANN models are nonparametric in spirit and are much more general than the previous exponential or logistic functions. The structure of the model can be easily modified by changing the architecture of the network. Second, within this general framework of ANNs, we can easily accommodate multiple inputs to neural network models by increasing the number of input nodes. Consequently, animal response to multiple objects on the landscape can be readily incorporated into the model simultaneously. This is important because animals must often weigh the importance of multiple sensory inputs when making decisions, such as the presence of food, risk from predators, and environmental conditions (Krebs and Davies 1993). Finally, it is also possible to incorporate other factors and covariates. In particular, we may draw population-level inference by pooling information across animals while accounting for individual effects.

Neural networks have been used elsewhere in modeling animal behavior and movement. For example, a strong case has been made for the usefulness of neural networks for modeling animal behavior (Enquist and Ghirlanda 2005). A recent example can be found in Dalziel et al. (2008), which places a neural network directly on the dispersal kernel. The approach is fully nonparametric without distributional assumptions and thus is very flexible. But it does not assess covariates directly, requires the spatial domain to be discretized (i.e., a grid), and focuses on individual-based inference. In contrast, our approach is semi-parametric, with neural networks placed on the model parameters of probability distributions. Semi-parametric modeling provides a nice balance between model flexibility and statistical efficiency and thus is a viable alternative. Our approach incorporates landscape features as covariates and utilizes the

idea of regression which eases interpretation. Moreover, our model allows continuous move angle and distance without having to discretize space and as we will demonstrate, can be inferred at both individual and population level.

For statistical inference, we adopt maximum likelihood. However, computation to obtain the maximum likelihood estimates (MLE) can be very challenging. Traditional methods such as Newton-Raphson type or back-propagation type algorithms are prone to local extrema (Hassoun 1995). To overcome such difficulty, we consider genetic algorithms (GA) as a powerful alternative computational method. GAs belong to the domain of evolutionary computing, which is characterized by evolving populations of solutions (Eiben and Smith 2003). GAs have been used in many applications to train weights in neural networks (Holland 1992; Mitchell 1996; Reed and Marks 1999; Strand et al. 2002; Morales et al. 2005). For example, Strand et al. (2002) proposed the concept of individual-based neural network genetic algorithm as a general approach to predicting behavioral and life-history traits that are fittest in a particular environment given the physiology of a species.

Due to the flexibility of ANN models and the feasibility of computation, we believe that our proposed methods will equip ecologists and biologists with far more adequate tools for studying animal movement across the landscape and hence, can impact future research in areas such as behavioral ecology and conservation biology. We propose responsive move models in Sect. 2 and ANN models in Sect. 3. In Sect. 4, we develop statistical inference via maximum likelihood and GAs. In Sect. 5, we demonstrate applications of the method using the rattlesnake data and provide a summary and further discussion in Sect. 6.

2 Model

2.1 Move model

Let n denote the number of animals. Let $\mathbf{s}_{ij}$ denote the jth observed location of the ith animal, where $j = 1, \ldots, n_i + 1$, $i = 1, \ldots, n$. We define a move as the difference between two adjacent observed locations and for the ith animal, let n_i denote the number of moves. Further, let A_{ij} denote the jth observed move angle and D_{ij} denote the observed move distance.

We model each move by specifying its move angle and distance:

$$\mathbf{s}_{i(j+1)} = \mathbf{s}_{ij} + D_{ij} \begin{bmatrix} \cos(A_{ij}) \\ \sin(A_{ij}) \end{bmatrix}. \tag{1}$$

Alternatively, we could model the move directly using the methodology to be developed here with some modification. However, modeling move angle and distance as in (1) enjoys ease in interpretability and visualization which we will demonstrate in the data example; thus it is the approach taken here.

2.2 Responsive move angle model

For a basic move angle model, we model A_{ij} by a von Mises distribution:

$$[A_{ij}|\mu,\kappa] \sim \text{von Mises}(\mu,\kappa), \tag{2}$$

where $A_{ij} \in (-\pi,\pi]$. The two parameters are a mean angle $\mu \in (-\pi,\pi]$ and a concentration parameter $\kappa \geq 0$ (Mardia and Jupp 2000). The probability density is given by

$$p(A_{ij}|\mu,\kappa) = \left\{2\pi I_0(\kappa)\right\}^{-1} \exp\left\{\kappa\cos(A_{ij}-\mu)\right\}, \tag{3}$$

where $\{2\pi I_0(\kappa)\}$ is a normalizing constant and$I_0(\kappa)$ is a modified Bessel function of the first kind and zero order. The mean angle μ is the angle of maximum probability density and the distribution is symmetric about the mean angle. The concentration parameter κ controls the dispersion of the distribution about the mean angle. When the concentration parameter $\kappa = 0$, the distribution is uniform on $(-\pi,\pi]$. As κ increases, the distribution becomes more concentrated about the mean angle. The von Mises distribution can be thought of as a circular analogue to a Gaussian distribution.

To model move angle in response to the landscape, we let A_{ij} follow a von Mises distribution with a move-dependent mean angle μ_{ij} and a move-dependent concentration parameter κ_{ij}:

$$[A_{ij}|\mu_{ij},\kappa_{ij}] \sim \text{von Mises}(\mu_{ij},\kappa_{ij}). \tag{4}$$

In particular, we assume that the mean angle and the concentration parameter of A_{ij} depend on a vector of p covariates $\boldsymbol{x}_{ij} \in \mathbb{R}^p$ with

$$\mu_{ij} = f_A(\boldsymbol{x}_{ij};\boldsymbol{\theta}_A), \quad \kappa_{ij} = g_A(\boldsymbol{x}_{ij};\boldsymbol{\theta}_A), \tag{5}$$

where $f_A(\cdot;\boldsymbol{\theta}_A)$ and $g_A(\cdot;\boldsymbol{\theta}_A)$ are functions parameterized by $\boldsymbol{\theta}_A$. Thus, the resulting probability density function is given by

$$p(A_{ij}|\boldsymbol{x}_{ij};\boldsymbol{\theta}_A) = \left[2\pi I_0\{g_A(\boldsymbol{x}_{ij};\boldsymbol{\theta}_A)\}\right]^{-1} \exp\left[g_A(\boldsymbol{x}_{ij};\boldsymbol{\theta}_A)\cos\{A_{ij}-f_A(\boldsymbol{x}_{ij};\boldsymbol{\theta}_A)\}\right]. \tag{6}$$

Tracey et al. (2005) assumed a constant mean response angle, where the response angle is the angle of movement relative to the animal-to-object angle. They also assumed two parametric forms for g in the concentration parameter, an exponential or a logistic function constrained to be decreasing with the animal-to-object distance. That is, if an animal responds to an object, then the strength of the animal's response is expected to increase as it gets closer to the object so that the animal has a greater tendency to move in the mean response angle. Conversely, as the animal-to-object

distance increases, the von Mises distribution becomes more uniform, and the animal moves in angles that depend less on the mean response angle. While reasonable for certain situations, these modeling assumptions can be rather restrictive in many other situations. Here we relax these assumptions and cast the problem under a richer and more powerful modeling framework using artificial neural network (ANN).

2.3 Responsive move distance model

For a basic move distance model, we model D_{ij} by a gamma distribution:

$$[D_{ij}|\alpha, \beta] \sim \text{gamma}(\alpha, \beta), \tag{7}$$

where $D_{ij} > 0$ and the shape and scale parameters are $\alpha, \beta > 0$. To model move distance in response to the landscape, we let D_{ij} follow a gamma distribution with a move-dependent shape parameter α_{ij} and a move-dependent scale parameter β_{ij}:

$$[D_{ij}|\alpha_{ij}, \beta_{ij}] \sim \text{gamma}(\alpha_{ij}, \beta_{ij}), \tag{8}$$

where $\alpha_{ij} = f_D(\boldsymbol{x}_{ij}; \boldsymbol{\theta}_D)$ and $\beta_{ij} = g_D(\boldsymbol{x}_{ij}; \boldsymbol{\theta}_D)$ parameterized by $\boldsymbol{\theta}_D$. In Tracey et al. (2005), the variance of the gamma distribution was assumed to be constant and the mean of the gamma distribution varied according to either an exponential or a logistic function in relation to the animal-to-object distance. Again, these restrictive assumptions are relaxed under the ANN modeling framework.

3 Artificial neural network

A neural network is a computational system based on a simplified model of neural systems in animals (Reed and Marks 1999). It consists of individual computational units called *nodes*, which are analogous to neurons. The nodes are linked by directed *connections*, which are analogous to the axon-dendrite connections among neurons. A connection from node k to node l has a weight θ_{kl} associated with it. The magnitude of this weight determines the strength of the connection, while its sign determines whether it promotes or inhibits the activation of the node to which the connection leads. We follow the common practice of arranging nodes into layers. One layer holds input nodes that receive input from outside the network and one layer holds output nodes that pass results out of the network. Between the input and output layers, there may possibly be one or more layers of hidden nodes, although not all neural networks have such hidden layers. Here, we focus on *feedforward neural networks* which only allow connections to be formed from nodes in a preceding layer to nodes in a following layer in a network. The networks we consider have one hidden layer of nodes (Fig. 2).

We denote the variables on the input layer by x_k with $k = 1, \ldots, K$, where K is the number of input nodes. Similarly, we denote the variables on the hidden layer and the output layer by y_l with $l = 1, \ldots, L$ and z_m with $m = 1, \ldots, M$, respectively, where L is the number of hidden nodes and M is the number of output nodes. The model is specified by the relation of x_k and y_l:

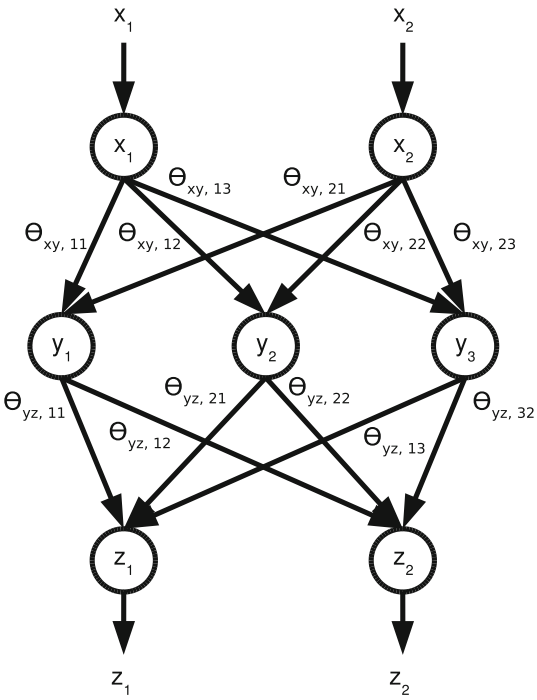

Fig. 2 Diagram of a feedforward neural network. The input nodes (*top row*) receive inputs x_1 and x_2 from outside the network. The input nodes are connected to the nodes in the hidden layer (*middle row*) which sum the weighted values from the inputs nodes and pass them through a logistic function to produce y_1, y_2, and y_3. The values y_1, y_2, and y_3 from the hidden nodes are weighted and summed in the output nodes (*bottom row*) to produce the outputs z_1 and z_2 which are passed out of the network. The weight on the connection from the kth input node to the lth hidden node is indicated by $\theta_{xy,kl}$ and from the lth hidden node to the mth output node is indicated by $\theta_{yz,lm}$

$$y_l = h\left(\sum_{k=1}^{K} \theta_{kl}^{xy} x_k - \theta_{0l}^{xy}\right), \quad l = 1, \ldots, L, \tag{9}$$

as well as the relation of y_l and z_m:

$$z_m = h\left(\sum_{l=1}^{L} \theta_{lm}^{yz} y_l - \theta_{0m}^{yz}\right), \quad m = 1, \ldots, M. \tag{10}$$

Here θ_{0l}^{xy} is a *threshold* of the lth hidden node and θ_{kl}^{xy} is the weight for the connection from input node k to hidden node l. Similarly, θ_{0m}^{yz} is the threshold of the mth output node and θ_{lm}^{yz} is the weight for the connection from hidden node l to output node m. Note that the individual parameters θ do not carry any direct biological interpretation and generally are not of interest in statistical inference, but as a whole allow flexible forms for the functions of interest such as f_A and g_A in (5).

If desired, we can fix specific connection weights or node thresholds at 0. The purpose in doing so is to reduce the number of parameters of the model. In the models we consider, we do not fix any of the connection weights, but allow all the hidden layer node and/or all the output layer node thresholds to be fixed at 0. This results in four cases: (1) do not fix any node thresholds, (2) fix only the hidden node thresholds at 0, (3) fix only the output node thresholds at 0, or (4) fix all hidden and output layer node thresholds at 0. We refer to the vector of all the weights and thresholds for a network as

$$\boldsymbol{\theta} = \left\{\theta_{kl}^{xy} : k = 0, \ldots, K, l = 1, \ldots, L\right\} \cup \left\{\theta_{lm}^{yz} : l = 0, \ldots, L, m = 1, \ldots, M\right\}. \tag{11}$$

A variety of options may be used as h, and we consider the logistic function $h(x) = \{1 + \exp(-x)\}^{-1}$ for the hidden nodes and the identity function $h(x) = x$ for the output nodes.

The set of x_k values are passed into the input nodes from outside of the network and are sometimes referred to as a *pattern*. Each node on the hidden layer and the output layer computes its output in the order that they are numbered. The set of z_m values from the output nodes are returned from the network after the computation is completed. In practice, observed data may be prepared for input to the network in a *pre-processing step* and network output may be prepared for use in a *post-processing step* (Bishop 2005). In our application, the pre-processing step involves computing the two components of a vector from the animal location to the nearest point on an urban edge, which may be passed to the network as inputs. These two inputs provide information on both the distance and angle of the habitat patch boundary from the animal's position.

When modeling move angle, our network has two outputs (z_1, z_2) and when modeling move distance, also two outputs (z_3, z_4). In the post-processing step, the mean angle of the von Mises distribution for move angles is computed as $\text{atan2}(z_2, z_1)$, which is the two-argument form of the arctangent function commonly implemented in most computer programming languages, and the concentration parameter as $(z_1^2 + z_2^2)^{1/2}$. For move distance, we model the shape and the scale parameters of the gamma distribution as $\exp(z_3)$ and $\exp(z_4)$, respectively. These transformations of the output are applied to ensure that the values of output are in the appropriate range of values.

Neural networks are, given enough nodes, *universal approximators*. In particular, Kolmogorov's theorem, applied to neural networks, demonstrates that one hidden layer is sufficient for this capability (Reed and Marks 1999). It has been shown that neural networks with two hidden layers can divide an input (i.e., covariate) space into classification regions of any shape and networks with one hidden layer can represent disjoint, nonconvex regions (Reed and Marks 1999). Basically, our networks divide the covariate space into regions and then assign a mean angle and concentration parameter to each region. We can increase the capacity of the network to do so by adding more hidden layer nodes. Furthermore, because we are using a logistic function in the hidden nodes (rather than a binary switch), the transitions of the network outputs between regions can be more gradual thereby taking the shape of a continuous function. We will take full advantage of these features when assessing covariates or factors in the system.

4 Inference

4.1 Population-level and individual-adjusted models

For any given individual animal, let $\boldsymbol{A}_i = (A_{i1}, \ldots, A_{in_i})'$ denote the vector of observed move angles, let $\boldsymbol{D}_i = (D_{i1}, \ldots, D_{in_i})'$ denote the vector of observed move

distances, and let $\boldsymbol{X}_i = [\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i}]$ denote the corresponding design matrix of covariates. For all the animals, let $\boldsymbol{A} = (\boldsymbol{A}_1', \ldots, \boldsymbol{A}_n')'$ denote the vector of all the observed move angles, let $\boldsymbol{D} = (\boldsymbol{D}_1', \ldots, \boldsymbol{D}_n')'$ denote the vector of all the observed move distances, and let $\boldsymbol{X} = [\boldsymbol{X}_1', \ldots, \boldsymbol{X}_n']'$ denote the design matrix of all the covariates. We define a *population-level model* by assuming that all the individuals share a same set of parameters $\boldsymbol{\theta}_A$ (or $\boldsymbol{\theta}_D$). Then, from (4) and (5), the log-likelihood function for $\boldsymbol{\theta}_A$ is

$$\ell(\boldsymbol{\theta}_A; \boldsymbol{A}, \boldsymbol{X}) = -\sum_{i=1}^{n} n_i \log(2\pi) - \sum_{i=1}^{n}\sum_{j=1}^{n_i} \log I_0\{g_A(\boldsymbol{x}_{ij}; \boldsymbol{\theta}_A)\} + \sum_{i=1}^{n}\sum_{j=1}^{n_i} g_A(\boldsymbol{x}_{ij}; \boldsymbol{\theta}_A) \cos\{A_{ij} - f_A(\boldsymbol{x}_{ij}; \boldsymbol{\theta}_A)\}. \tag{12}$$

The parameter value $\boldsymbol{\theta}_A$ in (12) reflects the effect of the population. Estimation of $\boldsymbol{\theta}_A$ would pool the data information across the animals, although it does not account for potential differences among the individuals. Similarly, it is straightforward to obtain the log-likelihood function $\ell(\boldsymbol{\theta}_D; \boldsymbol{D}, \boldsymbol{X})$ for $\boldsymbol{\theta}_D$ based on (8).

To account for individual effects, we create additional covariates and feed them into the ANN at additional input nodes. These covariates are essentially dummy variables that identify the individual animals. For example, for four animals, we use three binary variables, one for each animal $i = 2, 3, 4$. This results in an additional term unique to each individual $i > 1$ in (9). The additional weights will increase the dimension of $\boldsymbol{\theta}_A$ (or $\boldsymbol{\theta}_D$), but estimation and inference can be carried out in the same way as for (12). We note that individual effects are not necessarily linear or additive as in many traditional models, but rather can be highly nonlinear after being passed through the neural network. We will refer to this as an *individual-adjusted model*.

The aforementioned population-level models and individual-adjusted models both assume that the moves are unrelated to time for a given individual animal. To account for the effect of time, we develop a *time-adjusted model* by creating additional covariates and input nodes. First, we consider the time period (elapsed time) over which the observed movement occurs, which can adjust the model outputs based on the time duration of the observed movement. Second, we consider the elapsed linear time from an arbitrary point in time (in our case, 0:00, January 1st, 1999), which can adjust model outputs according to a temporal trend. Third, we consider circular time which represents seasonal effects. In this case we input the sine and cosine of the linear time, scaled by 2π, to the network.

4.2 Model assessment

For model assessment and comparison, we use Akaike's information criterion (AIC) with a finite-sample adjustment (Burnham and Anderson 2002). This information criterion penalizes the negative log-likelihood based on the number of model parameters. Since the number of parameters can be large relative to the number of observations, we use the small sample form which corrects for bias:

$$\mathrm{AIC}_c = -2\ell + 2dN/(N - d - 1) \tag{13}$$

where $\ell = \ell(\boldsymbol{\theta}_A; \boldsymbol{A}, \boldsymbol{X})$ (or $\ell(\boldsymbol{\theta}_D; \boldsymbol{D}, \boldsymbol{X})$) for the move angle (or distance) model, d is the number of model parameters and $N = \sum_{i=1}^{n} n_i$ is the sample size. AIC is interpreted as the relative Kullback-Leibler information loss among a set of models (Burnham and Anderson 2002). Hence, when performing model selection we seek the model with the lowest AIC_c value.

4.3 Genetic algorithm

Maximization of the log-likelihood function (12) can be performed using a Newton-Raphson type of optimization algorithm. Given the highly nonlinear structure of the neural network model, however, local extrema impose computational difficulty. Alternative algorithms, such as back-propagation, are available to train neural network weights (Reed and Marks 1999). These methods have many advantages such as computational efficiency; however, they are also prone to being trapped by local extrema (Hassoun 1995; Reed and Marks 1999; Enquist and Ghirlanda 2005). Thus we consider genetic algorithms (GA) as an alternative (Mitchell 1996).

In a GA, a solution is encoded as an array of a certain data type, depending on the nature of the problem. Each array representing a solution is referred to as a *chromosome*. We refer to each position on the chromosome as a *locus* and a specific value at a locus as an *allele*. We describe the specific formulation of the GA developed to fit our neural network models, but other GA operators and algorithm parameters are possible. In our model, a chromosome is designated by $\boldsymbol{\theta}^{(c,g)}$, where $c = 1, \ldots, C$ indexes the chromosome in a population and $g = 0, \ldots, G$ indexes the generation. Thus, each chromosome is an array of parameters of the neural network model for move angles and/or distances. At generation g, the entire set of solutions is denoted as $\boldsymbol{\theta}^{(g)} = (\boldsymbol{\theta}^{(1,g)'}, \ldots, \boldsymbol{\theta}^{(C,g)'})'$.

At $g = 0$, the parameter values are assigned by taking random draws from a uniform distribution. At each following generation, fitness is calculated for each chromosome, and the population of chromosomes for the next generation is created from the current generation. In our model, we let $\boldsymbol{\ell}^{(g)} = (\ell^{(1,g)}, \ldots, \ell^{(C,g)})'$ denote a vector of log-likelihood values for all the chromosomes at generation g, where $\ell^{(c,g)}$ is as in (12) evaluated at $\boldsymbol{\theta}^{(c,g)}$. We compute a vector of fitnesses for all the chromosomes $\boldsymbol{F}^{(g)} = (F^{(1,g)}, \ldots, F^{(C,g)})'$, where

$$\boldsymbol{F}^{(g)} = (\boldsymbol{\ell}^{(g)} - \ell^{(\min,g)})/(\ell^{(\max,g)} - \ell^{(\min,g)}) + \delta, \tag{14}$$

with $\ell^{(\max,g)} = \max\{\ell^{(1,g)}, \ldots, \ell^{(C,g)}\}$, $\ell^{(\min,g)} = \min\{\ell^{(1,g)}, \ldots, \ell^{(C,g)}\}$, and $\delta > 0$ is small value (say, 0.0001) that ensures all fitnesses are greater than zero.

Once the fitnesses have been calculated, a new generation of chromosomes is created. We use a survivor selection procedure called *elitism*, which takes the best C_{elite} chromosomes in the current generation and copies them to the next generation without modification. Elitism guarantees that the highest log-likelihood in the population will not decrease from one generation to the next. The remaining $C - C_{\text{elite}}$ members of the

next generation are created by an iterative process of parent selection, recombination (a.k.a. crossover), and mutation. Parent selection selects chromosomes for reproduction based on their fitness. Recombination mixes parent chromosomes to yield new solutions. Mutation introduces variability into the population.

For parent selection, we use a two-step tournament procedure. In the first step, two chromosomes are selected from the current generation with uniform probability. In the second step, the higher-fitness chromosome of the two is selected as a parent with probability p_{tour}; otherwise, the lower-fitness chromosome is selected to be a parent. This two-step procedure is repeated twice to select two parent chromosomes. After the parents are selected, a recombination operator is applied to mix the parent chromosomes with probability p_{cross} to produce two offspring chromosomes; otherwise, the two offspring are copied directly from the two parents. When mixing the parents, we use parameterized uniform recombination. If the recombination operator is applied, alleles at a given locus in the parent chromosomes are exchanged with probability p_{unif}; otherwise, they are not exchanged. Finally, a mutation operator is applied to each locus in the two offspring chromosomes. With probability p_{mut}, a uniform random variate on a specified range is drawn and added to the current allele at the locus. This procedure is applied to each locus in the two offspring chromosomes. After mutation, the two offspring chromosomes are added to the population for the next generation. If only one chromosome is needed, then the first offspring is used. The selection-recombination-mutation process is repeated until all $C - C_{\text{elite}}$ members of the next generation are created. This entire process of fitness calculation and offspring creation is repeated for G generations.

5 Rattlesnake movement data

5.1 Data description

We analyze movement data for four adult male red diamond rattlesnakes (*Crotalus ruber*) collected via radio-telemetry (Tracey 2000). These rattlesnakes inhabited a patch of coastal sage scrub habitat isolated by paved roads and urban development in coastal San Diego County, California, USA (Fig. 1a). Red diamond rattlesnakes, a species of pitviper, are endemic to coastal southern California and the Baja California, Mexico. In the United States, red diamond rattlesnake habitat is rapidly being destroyed, fragmented, and degraded by urban and agricultural development. An objective of the radio-telemetry study was to assess the effects of urban development on the movement of the rattlesnakes.

Data in this example include locations of each of the rattlesnakes and line data for a polygon that delineates the boundary of the habitat patch (Fig. 1a). Location data and other associated information were collected by radio-tracking each rattlesnake. Rattlesnake locations were obtained by surgically implanting a small transmitter in the snake's body, and then locating the transmitter by a receiver attached to a directional antenna. Spatial coordinates for each snake location were obtained using global positioning system (GPS) receivers and differential correction techniques were applied to improve their accuracy. These animals were typically located at 2–4 day

intervals between May and October (the time of the year when rattlesnakes in the study area were most active) of 1999 and 2000. Data for relocation intervals during which the animal did not move were excluded from this analysis. A total of $n_1 = 49, n_2 = 26, n_3 = 38,$ and $n_4 = 29$ moves were observed for the four animals. During the study, the animals were not observed moving beyond the boundary of the habitat patch. Line segments along the boundary of the habitat patch were digitized on one meter resolution US Geological Survey DOQQ imagery using ArcView 3.2 and the Spatial Analyst extension (Environmental Systems Reseach Institute Inc. 1999). The set of closed line segments, where the from-vertex of the first segment was identical to the to-vertex of the last segment, form a polygon representing the habitat patch (Fig. 1a). The resulting ArcView shapefile was exported as a tab delimited text file for input to R (R Development Core Team 2008) environment in which we perform the remaining analysis.

5.2 Model fitting and comparison

In previous work, we did not detect an effect of distance from an animal location to the urban-habitat boundary on move distance (Tracey et al. 2005). Therefore, we will focus only on modeling move angle in this example. The pre-processing step consists of computing the inputs to the neural networks and the observed responses from the data. Observed responses are the move angles of the rattlesnakes, which we compute as $A_{ij} = \text{ang}(\boldsymbol{s}_{ij}, \boldsymbol{s}_{i(j+1)}) = \text{atan2}(s_{i(j+1)2} - s_{ij2}, s_{i(j+1)1} - s_{ij1})$, where $\boldsymbol{s}_{ij} = (s_{ij1}, s_{ij2})'$ is the jth location of the ith animal, $j = 1, \ldots, n_i + 1$, and $i = 1, \ldots, n$.

Inputs to the neural network consist of some combination of vectors from the animal location to the nearest point on the habitat patch boundary and indicator variables that identify the rattlesnake to which each observation corresponds. For models that include a boundary response, we compute the spatial coordinates of the nearest point on the habitat patch boundary ($\boldsymbol{t}_{ij}$) from the rattlesnake location ($\boldsymbol{s}_{ij}$) and pass ($\boldsymbol{t}_{ij} - \boldsymbol{s}_{ij}$) as two inputs to the neural network. These inputs provide information on the relation between the rattlesnake location and the edge of the habitat patch. For individual-adjusted models that include individual effects, we pass three indicator variables, such that we input (0, 0, 0), (1, 0, 0), (0, 1, 0), and (0, 0, 1) for observations from animal $i = 1, 2, 3$ and 4, respectively. These inputs allow the neural network to modify the response for each individual rattlesnake. The post-processing step for the neural network models consists of computing the log-likelihood (12) as described in Sect. 3.

We use a genetic algorithm described in Sect. 4.3 with $C = 40$, $G = 5,000$, $C_{\text{elite}} = 5$, $p_{\text{tour}} = 0.6$, $p_{\text{cross}} = 0.7$, $p_{\text{unif}} = 0.7$, $p_{\text{mut}} = 0.1$, and uniform random variates for mutation drawn on the interval $(-0.5, 0.5)$. These algorithm parameters are chosen based on trial-and-error. Strategies of selecting these parameters can be found in Eiben and Smith (2003). We fit a total of 32 different models responsive to the habitat boundary (Table 1). The neural network models have 2–5 hidden nodes and are with or without thresholds in the hidden and/or output nodes. This results in 16 combinations of possible models. In addition, we consider either population-level models or individual-adjusted models. The only inputs in population-level models are the vectors

from the rattlesnake locations to the nearest points on the habitat patch boundary and the weights in these neural networks are the same for all animals. In contrast, individual-adjusted models have indicator variables for animal identities as additional inputs, as well as the additional weights in the neural network that account for the individual effects.

For model assessment and comparison, we use Akaike information criterion corrected for finite-sample size (AIC_c; Burnham and Anderson 2002; Sect. 4.2). Table 1 gives the AIC_c values of the 32 boundary responsive models. Model #17 has the smallest AIC_c value and is selected as the best model. This model is a population-level model that has 2 hidden nodes and has thresholds on both the hidden layer and the output layer nodes. Based on the best model, we further assess the effect of urban edge on the movement of these rattle snakes. In the region where rattlesnake movements were observed (Fig. 1a), movement responses are generally avoidance of the areas outside the habitat patch (Fig. 1b). That is, for the northern portion of the habitat patch, the model produces move angles directed toward the interior of the patch, which is consistent with observed movements of real rattlesnakes. The concentration parameters produced by the model at locations along the northern and eastern boundaries of the habitat patch are also greater than those produced in the north-central region of the patch. This indicates an increased directionality of move angles near the boundaries, but less directionality of movement in the interior of the patch. Movement paths simulated from this model would have increased residency in the north-central part of the habitat patch due to the lower concentration of the move angles in this area. Again, this is consistent with the observed rattlesnake locations concentrated in the north-central part of the habitat patch (Fig. 1a).

As mentioned in Sect. 2.1, modeling move angle and distance is equivalent to modeling move between two adjacent animal locations. However, with our approach we can display an entire map of mean move angles and concentrations as in Fig. 1b, for any number of pre-specified locations. This is more flexible than showing a dispersal kernel which generally depends on only one starting location. Thus, our approach enables an intuitive visualization of movement on the landscape, which may not be straightforward to obtain with dispersal kernel outputs. The ability to display expected movement responses to landscape features would greatly facilitate communications in applications of animal movement models to conservation or management problems.

Note that in the southern part of the habitat patch in Fig. 1b, the model produced mean angles directed approximately south, toward the southern boundary of the habitat patch. This would produce movements that would cause a rattlesnake to leave the habitat patch and enter areas of urban development. However, no rattlesnake movements were actually observed in this area, so the model is extrapolating beyond the observed range of the neural network inputs; hence, the results are not as reliable in this region. Thus, caution is advised when applying the model outside the range of observed input values. We also note that, because the best model has two input nodes and two hidden nodes on just one hidden layer in the network, the output space is effectively partitioned into disjoint, nonconvex regions, as shown in Fig. 1b.

Comparing the AIC_c values of population-level models and individual-adjusted models, we note that there appear to be no detectable individual effects. The best six out of 16 of the responsive models were not individual-adjusted. The second, third,

and fourth best models (models #21, 25, 29; Table 1) are similar in structure to the best model except that they have no thresholds in the hidden layer, output layer, or both the hidden and output layers. The fifth and sixth best models (models #18 and 19; Table 1) are also similar in structure to the best model, except that they have additional hidden layer nodes.

5.3 Further analysis

Since the primary interest of the study is to examine the relation between animal movement and the landscape, we compare the models above with baseline neural network models where the boundary of the habitat patch is not taken into account. These baseline models have 2–3 hidden nodes and are with or without thresholds in the hidden and/or output nodes. There are a total of 8 individual-adjusted models with indicator variables for animal identities as the only input. However, there is only 1 population-based model, as such a model does not have any input (Table 2). In this case, we simply fit a von Mises distribution to the move angle data. The models that take into account the urban edge clearly out-perform the baseline models, as evidenced by the much smaller AIC_c values in Table 1. Nineteen of the 32 responsive models out-performed the best baseline model (Model #41). This suggests that the urban boundary plays an important role in the movement of the rattlesnakes.

In addition, we consider time inputs described in Sect. 4. In this case, we consider the time duration of the move, linear time, and circular time models with 2–5 hidden nodes. We allow more hidden nodes for the time-adjusted models, compared to the baseline models, because the additional processing capability might be required for the additional time inputs. In the time-adjusted models, all other properties of the models are the same as our best model, which has thresholds for both the hidden and output nodes, boundary inputs, and no individual adjustment. These models all have larger AIC_c values than the best model (Model #17, Table 1), indicating that the additional

Table 2 AIC_c values of baseline neural network models

Model	Hidden nodes	Hidden thresholds	Output thresholds	AIC_c	ΔAIC_c	Akaike weight
Individual-adjusted						
33	2	Yes	Yes	549.02	41.57	7.86E-010
34	3	yes	Yes	564.66	57.20	3.16E-013
35	2	No	Yes	544.13	36.68	9.05E-009
36	3	No	Yes	556.65	49.19	1.73E-011
37	2	Yes	No	544.19	36.74	8.79E-009
38	3	Yes	No	559.27	51.82	4.66E-012
39	2	No	No	539.44	31.98	9.47E-008
40	3	No	No	551.52	44.07	2.25E-010
Population-level						
41	–	–	–	524.90	17.44	1.36E-004

Table 3 AIC_c values of population-level boundary responsive neural network models with additional time inputs

Model	Hidden nodes	Time input	AIC_c	ΔAIC_c	Akaike weight
42	2	Move duration	537.98	30.52	1.97E-007
43	3	Move duration	548.00	40.54	1.31E-009
44	4	Move duration	559.56	52.10	4.04E-012
45	5	Move duration	570.57	63.12	1.64E-014
46	2	Linear elapsed	541.45	33.99	3.47E-008
47	3	Linear elapsed	551.06	43.60	2.84E-010
48	4	Linear elapsed	561.34	53.88	1.66E-012
49	5	Linear elapsed	572.30	64.84	6.93E-015
50	2	Circular elapsed	541.20	33.74	3.92E-008
51	3	Circular elapsed	550.86	43.40	3.14E-010
52	4	Circular elapsed	561.10	53.65	1.87E-012
53	5	Circular elapsed	572.04	64.59	7.88E-015

All models have thresholds for the hidden layer and the output layer nodes

time covariates do not improve the models (Table 3). However, they may be more important for data sets in other situations.

Analyses based on our models yield useful results even though the number of observations in this example is moderate. From the analysis, we find that individual adjustment and time inputs do not improve model performance. We also find that models that include inputs for the relation between the rattlesnakes and the patch boundary perform much better than those that do not include these inputs. Thus, the rattlesnakes responded to the habitat patch boundary and the response was one of avoidance. In the areas where rattlesnakes were most active, the best model produced smaller concentration parameters for the von Mises distribution, which corresponds to greater residency in those areas. The conclusions are consistent with our previous work in Tracey et al. (2005), but more compelling, as the inference here is drawn at the population level after evaluating both individual and time effects, rather than from merely one animal before.

6 Conclusion and discussion

In this paper, we have built upon previous work by using artificial neural networks in a semi-parametric modeling framework to model data on animal movement angles and distances. The semi-parametric modeling framework offers far greater flexibility for modeling movement than our previous work that assumed rigid parametric forms such as exponential or logistic functions in Tracey et al. (2005). It also offers advantages in interpreting the model outputs compared to some of the existing methods.

We have proposed model fitting using a genetic algorithm fitting procedure in which fitness is calculated from the log-likelihood. We further describe model selection based

on AICc. Even though AICc is commonly used for model selection, other criteria are possible such as Bayesian information criterion or cross validation. It would be instrumental to evaluate different model selection criteria for ANN models.

We have demonstrated the application of the methodology with rattlesnake movement data in relation to habitat patch boundaries by focusing on move angles. In the current framework, we have separated modeling move angles from move distances, which effectively assumed that the two quantities are independent. It is possible to extend to a more general framework such that move angles and move distances are modeled jointly via bivariate distributions and thus correlation between move angles and move distances is allowed.

We have also observed difficulties in predicting the mean angles in the portion of the study area where there was little data. For conservation management purpose, it is often of interest to make such predictions, however. An alternative approach that we are currently experimenting is to use computer simulation, which produces maps of predicted moves in different landscapes, based on parameter estimates from data as well as other hypothetical but plausible parameter values.

Another consideration for our approach is to account for variable time intervals between animal locations. This may be addressed by modeling move rates rather than distances. For move angles, our current approach is to input time between locations to the network as a covariate, so that the network can adjust move angle model parameters accordingly. Alternative approaches may be worth pursuing. We leave this and other considerations for future investigation.

Acknowledgments This research was supported in part by the National Science Foundation Ecology of Infectious Disease program grant NSF EF-0723676. We thank Sue VandeWoude of Colorado State University for her support. We gratefully acknowledge Ted J. Case for his support of the red diamond rattlesnake study conducted by J. A. Tracey at the University of California, San Diego. Early stages of this work were supported in part by the US Geological Survey, The Nature Conservancy, and the California Department of Fish and Game. We specifically thank Robert Fisher (USGS) and Scott Morrison (TNC) from these organizations. We also thank two referees, an associate editor, and the editor for their constructive comments that helped improve this paper.

References

Batschelet E (1981) Circular statistics in biology. Academic, New York

Bishop CM (2005) Neural networks for pattern recognition. Oxford University Press, New York

Brillinger DR, Preisler HK, Ager AA, Kie JG (2004) An exploratory data analysis (EDA) of the paths of moving animals. J Stat Plan Inference 122:43–63

Burnham KP, Anderson D.R. (2002) Model selection and multimodel inference, 2nd edn. Springer, New York

Christ A, Ver Hoef JM, Zimmerman DL (2008) An animal movement model incorporating home range and habitat selection. Environ Ecol Stat 15:27–38

Dalziel BD, Morales JM, Fryxell JM (2008) Fitting probability distributions to animal movement trajectories: using artificial neural networks to link distance, resources, and memory. Am Nat 172:248–258

Eiben AE, Smith JE (2003) Introduction to evolutionary computing. Springer, New York

Enquist M, Ghirlanda S (2005) Neural networks and animal behavior. Princeton Univesity Press, Princeton

Environmental Systems Reseach Institute Inc (1999) ArcView 3.2 with spatial analyst extension. Redlands, CA

Forester JD, Im HK, Rathouz PJ (2009) Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. Ecology 90:3554–3565

Hassoun MH (1995) Fundamentals of artificial neural network. MIT Press, Cambridge

Holland JH (1992) Adaptation in natural and artificial systems. MIT Press, Cambridge
Hooten MB, Johnson DS, Lowry JH (2010) Agent-based inference for animal movement and selection. J Agric Biol Environ Stat (to appear)
Jander R (1975) Ecological aspects of spatial orientation. Annu Rev Ecol Syst 6:171–188
Johnson DS, Thomas DL, Ver Hoef JM, Christ A (2008) A general framework for the analysis of animal resource selection from telemetry data. Biometrics 64:968–976
Kareiva PM, Shigesada N (1983) Analyzing insect movement as a correlated random walk. Oecologia 56:234–238
Krebs JR, Davies NB (1993) An introduction to behavioral ecology, 3rd edn. Blackwell Science Ltd, Oxford
Lima SL, Zollner PA (1996) Towards a behavioral ecology of ecological landscapes. Trends Ecol Evol 11:131–135
Mardia KV, Jupp PE (2000) Directional statistics. Wiley, New York
Mitchell M (1996) An introduction to genetic algorithms. MIT Press, Cambridge
Morales JM, Fortin D, Frair JL, Merrill EH (2005) Adaptive models for large herbivore movements in heterogeneous landscapes. Landsc Ecol 20:301–316
Patterson TA, Thomas L, Wilcox C, Ovaskainen O, Matthiopoulos J (2008) State-based models of individual animal movement. Trends Ecol Evol 23:87–94
R Development Core Team (2008) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07-0
Reed RD, Marks RJI (1999) Neural smithing. MIT Press, Cambridge
Strand E, Huse G, Giske J (2002) Artificial evolution of life history and behavior. Am Nat 159: 624–644
Tracey JA (2000) Movement of red diamond rattlesnakes (Crotalus ruber) in heterogeneous landscapes in coastal southern california. Master's thesis, University of California, San Diego
Tracey JA (2006) Individual-based modeling as a tool for conserving connectivity. In: Crooks KR, Sanjayan M (eds) Connectivity conservation. Cambridge University Press, Cambridge
Tracey JA, Zhu J, Crooks K (2005) A set of nonlinear regression models for animal movement in response to a single landscape feature. J Agric Biol Environ Stat 10:1–18
Turchin P. (1998) Quantitative analysis of movement. Sinauer Associates, Sunderland
Van Vuren D (1998) Mammilian dispersal and reserve design. In: Caro T (ed) Behavioral ecology and conservation biology. New York, Oxford

Author Biographies

Jeff A. Tracey received an M.S. in Biology from the University of California, San Diego and an M.S. in Biometry from the University of Wisconsin–Madison. He received a Ph.D. in Ecology from Colorado State University in 2006, and continued as a post doctorate research associate in the Department of Fish, Wildlife, and Conservation Biology at Colorado State University until 2009. Since then he has founded, and is President of, SigmaLogistic Consulting Inc. His primary research interests are applications of statistical computing, artificial intelligence, and agent-based modeling to problems in conservation biology and epidemiology.

Jun Zhu received her B.A. in Mathematics and Computer Science from Knox College in 1994, M.S.E. in Mathematical Sciences from Johns Hopkins University in 1995, and Ph.D. in Statistics from Iowa State University in 2000. She has been on the faculty of University of Wisconsin–Madison as an Assistant and Associate Professor since 2000 and is currently an Associate Professor in the Statistics Department at Colorado State University. Her primary research interests are environmental statistics, spatial statistics, and spatial-temporal statistics with applications in biology, ecology, and environmental sciences.

Kevin R. Crooks is currently an Associate Professor in the Department of Fish, Wildlife, and Conservation Biology and the Graduate Degree Program in Ecology at Colorado State University. He received his B.S. in Zoology at Colorado State University in 1989, his M.S. in Ecology at the University of California Davis in 1994, his Ph.D. in Biology at the University of California Santa Cruz in 1999, and most recently was an Assistant Professor in the Department of Wildlife Ecology at the University of Wisconsin Madison from 2001 to 2003. A primary research avenue is investigation of the effects of habitat fragmentation and urbanization on the ecology and conservation of wildlife in developing landscapes.