Submitted to the Annals of Applied Statistics
arXiv: arXiv:0000.0000

A MULTI-STATE CONDITIONAL LOGISTIC REGRESSION MODEL FOR THE ANALYSIS OF ANIMAL MOVEMENT

By Aurélien Nicosia, Thierry Duchesne, Louis-Paul Rivest and Daniel Fortin

Université Laval, Québec, Canada

A multi-state version of an animal movement analysis method based on conditional logistic regression, called Step Selection Function (SSF), is proposed. In ecology SSF is developed from a comparison between the observed location of an animal and randomly sampled locations at each time step. Interpretation of the parameters in the multi-state model and the impact of different sampling schemes for the random locations are discussed. We prove the relationship between the new model, called HMM-SSF, and a random walk model on the plane. This relationship allows one to use both movement characteristics and local discrete choice behaviors when identifying the model's hidden states. The new HMM-SSF is used to model the movement behavior of GPS-collared bison in Prince Albert National Park, Canada, where it successfully teases apart areas used to forage and to travel. The analysis thus provides valuable insights into how bison adjust their movement to habitat features, thereby revealing spatial determinants of functional connectivity in heterogeneous landscapes.

1. Introduction. In animal ecology, being able to understand and model the movement of animals is fundamental (Nathan et al., 2008). For example, animal behaviourists want to see to what extent animals have preferred movement directions or are attracted towards several environmental targets, such as food-rich patches and previously visited locations (spatial memory effect) (Latombe et al., 2014). The development of Global Positioning System (GPS) technology permits the collection of a large amount of data on animal movement. This can be combined to data available from geographic information systems (GIS) to investigate how the environment influences animal displacement. To achieve this goal, robust statistical techniques and flexible animal movement models are required.

Discrete time models for animal movement are actively being developed and investigated (Holyoak et al., 2008). Because displacement in discrete time

Keywords and phrases: Animal movement, Biased correlated random walk, Conditional logistic regression, GPS, Hidden Markov model, Step Selection Function

can be characterized by the distance and the direction between two consecutive localizations, circular-linear processes can be used to model movement in 2D. A basic model is the biased correlated random walk (BCRW, Turchin, 1998); it predicts the next motion angle as a compromise between the current one (often called directional persistence) and the direction towards a specific target (also called directional bias). This type of model handles environmental targets through their directions. Unfortunately, it cannot account for the impact of local pixel characteristics on the selection process. This can be done using *Step Selection Function* (SFF), introduced by Fortin et al. (2005). The SSF is a discrete choice model that compares the local characteristics of pixels selected by the animal at each time step (cases) to control pixels that could have been visited given the animal's previous position. Duchesne, Fortin and Rivest (2015) formally prove that the parameters of a BCRW can be estimated within a SSF. This is used in Avgar et al. (2016) to propose an *integrated Step Selection Analysis* (i-SSA) to estimate the parameters of a BCRW and of the local selection probabilities in a single analysis.

Often, animal movement involves multiple states or behaviors (Fryxell et al., 2008). For instance, Langrock et al. (2012) identified two states, "exploratory" and "encamped", in their analysis of bison movement. The former state is characterized by long traveled distances and turning angles between two consecutive locations that tend to be concentrated around zero, while the latter is characterized by short distances and nearly uniformly distributed turning angles. Multiple movement behaviors can be accounted for through hidden states. Baum and Petrie (1966) give a general presentation of these models and Morales et al. (2004), Jonsen, Flemming and Myers (2005), Holzmann et al. (2006), Langrock et al. (2012) and more recently Nicosia et al. (2017) use hidden state models to analyze angular-distance data on animal movement

The main contribution of this paper is, in Section 2, to propose a multi-state SSF, called HMM-SSF for hidden Markov model step selection function, handling both animal movement and local habitat selection. Section 2 also discusses parameter interpretation in a multi-state context, and the sampling of control locations. In Section 3, we prove that the proposed HMM-SSF model can be used to fit the multi-state random walk model of Nicosia et al. (2017); this theoretical result is validated using a simulation study and the analysis of a real data set. Section 4 identifies two states in the analysis of the movement trajectory of bison in Prince Albert Na-

tional Park, Canada using an HMM-SSF; both movement and local habitat selection parameters vary between states.

2. Multi-state Step Selection Function.

2.1. *Single-state Step Selection Function.* Let us suppose that we follow an animal equipped with a GPS collar which provides the animal's location at regular time intervals $t = 1, \ldots, T$, for example every 1 hour. The data are combined with information on the animal's habitat in a geographic information system (GIS). Step Selection Functions (SSF, Fortin et al., 2005), specify how the animal uses its habitat by modeling the discrete choices that it makes at each time step. At a given time step, an SSF compares the characteristics of the location (and of the trajectory leading to this location) visited by the animal with J control steps to other locations that the animal could have visited at that time given the previous step. This comparison uses GIS data, an $r \times 1$ vector $\mathbf{x}_{0t}$, for the observed location and the corresponding vectors $\mathbf{x}_{jt}, j = 1, \ldots, J$ for the J control locations.

Steps can be characterized by features encountered on the way (e.g., road, proportion of forest cover), at the step's end (e.g., land covers, elevation), bearing directions of habitat features at relatively far distances (e.g., road, canopy gap), and speed (e.g. Fortin et al. (2005), Vanak et al. (2013), Basille et al. (2015)). The wish of an animal to go to a specific location, e. g. a target meadow, can be entered in the model as an explanatory variable equal to the cosine of the difference between the direction to the next location and the direction to the target (Duchesne, Fortin and Rivest, 2015). A directional persistence, that is the wish of an animal to move forward, can be included in the analysis through the cosine of the difference between the motion angles of current and previous steps.

The data for an SSF analysis is $\{[\mathbf{x}_{0t}, \mathbf{x}_{1t}, \ldots, \mathbf{x}_{Jt}] : t = 1, \ldots T\}$. It is analyzed using a conditional logistic regression model for a matched case-control design (Hosmer and Lemeshow (2000), chapter 7). It is also equivalent to the multinomial logit discrete choice model (Train, 2003, chapter 3). Thus, at time step t, the probability that the animal chooses the location with step characteristics $\mathbf{x}_{0t}$, rather than one of the J control locations with respective step characteristics $\mathbf{x}_{jt}, j = 1, \ldots, J$ is

$$p_t = \frac{\exp\left(\mathbf{x}_{0t}^\top \beta\right)}{\sum_{j=0}^{J} \exp\left(\mathbf{x}_{jt}^\top \beta\right)}, \tag{2.1}$$

where β is a $r \times 1$ vector of unknown selection parameters. Following Hosmer and Lemeshow (2000), β is easily estimated by maximizing the conditional

logistic regression likelihood given by

$$L(\beta) = \prod_{t=1}^{T} \frac{\exp\left(\mathbf{x}_{0t}^{\top}\beta\right)}{\sum_{j=0}^{J}\exp\left(\mathbf{x}_{jt}^{\top}\beta\right)}. \tag{2.2}$$

To discuss the interpretation of β we first consider a simple model with a single dichotomous explanatory variable x identifying a particular type of habitat representing $100\times H\%$ of the study area. In a null model, with $\beta = 0$, the probability of selecting the habitat at a time step is H. When $\beta \neq 0$ this probability becomes $e^{\beta}H/(1 + e^{\beta}H)$ which is larger than H if $\beta > 0$, see Appendix of Duchesne, Fortin and Rivest (2015) for more details. With a continuous explanatory variable, the same interpretation holds. Suppose that a variable x, available at each location of the map, is distributed as a stationary random field with marginal density $f(x)$. If $\beta = 0$, then the density of x for the selected locations is $f(x)$. When $\beta \neq 0$, this density is proportional to $e^{\beta x}f(x)$; this gives a weighted distribution (see, Patil (2005)). If $\beta > 0$ the animal tends to select locations with higher values of x more often than would be expected with a purely random selection. As a matter of fact, the value of β is actually the log of the odds that the animal will choose a location with a value of the explanatory variable equal to $x + 1$ divided by the odds of choosing a location with a value of the explanatory variable equal to x.

When the x variable is the cosine of the difference between the angles of the direction of a potential target and the current motion angle, then a positive value of β means that the target is attractive (steps in its direction are selected more often). More details on this latter interpretation may be found in Duchesne, Fortin and Rivest (2015) who actually show that an SSF with such cosine explanatory variables and uniform sampling of the control locations can be used to estimate the parameters of a Biased Correlated Random Walk model (BCRW, Turchin (1998)).

2.2. *Extension to multi-state SSF.* Often, the animals exhibit more than a single step selection behavior (Fryxell et al., 2008). Such a change in behavior can be explained by a hidden-state model (Frühwirth-Schnatter, 2013) with a different SSF in each state. We propose to model this phenomenon with a HMM-SSF that uses a hidden Markov process $\{S_t, t = 1, \ldots, T\}$, where S_t represents the state (behavior) of the animal at time step t, to account for the changes between states. Let $\beta^{(k)}, k = 1, \ldots, K$ denote the selection coefficients of the SSF when the animal is in state k. Following the reasoning of Nicosia et al. (2017), the likelihood function of this multi-state SSF model is

$$L(\beta) = \prod_{t=1}^{T} \sum_{k=1}^{K} \left(p_t^{(k)} \cdot \mathbb{P}(S_t = k | \mathcal{F}_{t-1}^c) \right), \tag{2.3}$$

where

$$p_t^{(k)} = \frac{\exp\left(\mathbf{x}_{0t}^\top \beta^{(k)}\right)}{\sum_{j=0}^{J} \exp\left(\mathbf{x}_{jt}^\top \beta^{(k)}\right)}, \tag{2.4}$$

and $\mathcal{F}_t^c$ denotes the complete data history up to time t, which consists of the observed data and of the unobserved state; thus $\mathcal{F}_t^c$ contains S_ℓ and $[\mathbf{x}_{0\ell}, \mathbf{x}_{1\ell}, \ldots, \mathbf{x}_{J\ell}]$, for $\ell = 1, \ldots, t$. Figure 1 presents the dependence structure of the proposed model.

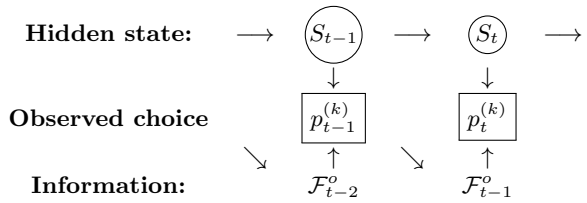


Fig 1: Dependence structure of the proposed model

The probability $\mathbb{P}(S_t = k|\mathcal{F}_{t-1}^c)$ in (2.3) is called the "predictive" probability. It can be efficiently computed using a filtering-smoothing algorithm (see Appendix B for more details) when $\{S_t\}$ is modeled as a Markov chain.

Inference about the parameters $\beta^{(k)}, k = 1, \ldots, K$ is based on the maximized log-likelihood $\ell(\beta) = \ln L(\beta)$. When $\{S_t\}$ is modeled as a Markov chain, we can use the EM algorithm and the filtering-smoothing algorithm to implement the inference and standard errors are estimated by computing the hessian of the observed log-likelihood function. Details of the procedure are given in Appendix A. To reach the global maximum of the observed likelihood function we use the short-run long-run EM algorithm strategy, see Appendix C.

2.2.1. *Interpretation of the parameters.* The interpretation of a parameter $\beta^{(k)}$ of an HMM-SSF is almost the same as that of β in a single-state SSF, except that it is conditional on the state in which the animal is. A non null β for a variable x means that the distribution of x constructed with the locations chosen by the animal differs from the stationary distribution of x, $f(x)$ over the study area. For instance if we have two states ($k = 1, 2$) and we

have a coefficient $\beta^{(1)} > 0$ for x in state 1 and a coefficient $\beta^{(2)} < 0$ for x in state 2, then this means that when the animal is in state 1, it tends to select locations, with distribution proportional to $e^{\beta^{(1)}x} f(x)$, with high values of x more often while in state 2 it tends to favor locations, with distribution proportional to $e^{\beta^{(2)}x} f(x)$, with small values of x. The value of β has the same log odds ratio interpretation within each state as in the single-state SSF.

2.2.2. *Sampling the control locations.* This section discusses the sampling of the control locations. At time t the animal is at a location $P_t \in \mathbb{R}^2$ and at time $t+1$ it will be at location P_{t+1}. The control locations for P_{t+1} are defined as a direction and a distance from P_t. Following Duchesne, Fortin and Rivest (2015) we select the control directions uniformly on $[0, 2\pi[$. Forester, Im and Rathouz (2009) argue that the method used to select the control distances influences the parameter estimates in a standard, single state, SSF. They emphasized that the range of the control distances needs to cover all the distances that the animal may possibly travel.

Let $\{(\phi_{jt}, h_{jt}) : \ j = 1, \ldots, J\}$, where $\phi_{jt} \in [0, 2\pi[$ is an angle and $h_{jt} > 0$ is a distance, be the polar coordinates (with P_t as origin) of the J control locations matched with P_{t+1}, and (ϕ_{0t}, h_{0t}) be the polar coordinates of P_{t+1}. As discussed above, ϕ_{jt} is sampled uniformly over $[0, 2\pi[$. The distances can be sampled uniformly over $[0, M]$, where M is large enough for $[0, M]$ to cover all possible observed distances. Let $D_k, k = 1, \ldots, K$, denote the support of the traveled distances in state k. Forester, Im and Rathouz (2009) have shown that, in a single state model, if the support of the control distances (i.e., $[0, M]$) does not include the support D_1 of the traveled distances, then we induce a bias in the estimation of the parameters $\beta^{(1)}$. This statement also applies to a multi-state model and therefore $[0, M]$ needs to cover $\cup_k D_k$.

Another way to sample the control distances is through a parametric distribution, such as a gamma distribution. This latter sampling procedure is discussed by Forester, Im and Rathouz (2009) for a single state SSF. We implement it in a HMM-SSF setting in the next section.

3. Multi-State SSF Model with Distances and Angles.

3.1. *Relationship with a Random Walk Model.* In this section we investigate whether the multi-state random walk model introduced by Nicosia et al. (2017) can be fitted using a multi-state SSF. Our goal is to generalize the findings of Duchesne, Fortin and Rivest (2015) to a complex multi-state SSF involving state dependent distributions for distances. This highlights

that the parameters of a specific multi-state SSF with covariates as (3.3) can be interpreted as those of a multi-state random walk model.

To see this we let both models depend on the directions from P_{t-1} to P_t, $\phi_{0,t-1}$, and from P_t to p potential targets in the landscape (e.g. the closest meadow, a canopy gap or the closest forest), denoted by the angles $\theta_{it}, i = 1, \ldots, p$. Figure 2 gives the notations with $p = 2$ targets.

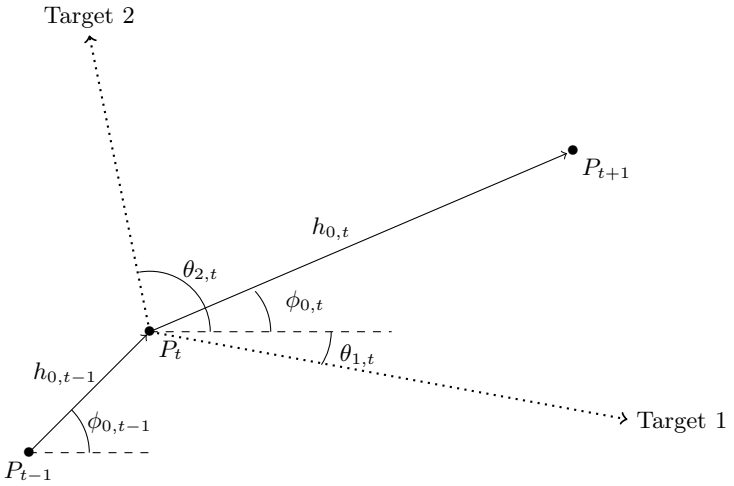


Fig 2: Notation of random walk model with $p = 2$ targets.

Knowing that the animal is in state k, The distribution of the direction $\phi_{0,t}$ at time t, observed when traveling from P_t to P_{t+1}, depends on the vector

$$\mathbf{V}_t^{(k)} = \kappa_0^{(k)} \begin{pmatrix} \cos(\phi_{0,t-1}) \\ \sin(\phi_{0,t-1}) \end{pmatrix} + \sum_{i=1}^{p} \kappa_i^{(k)} \begin{pmatrix} \cos(\theta_{it}) \\ \sin(\theta_{it}) \end{pmatrix}, t = 1, \ldots, T, \tag{3.1}$$

where $(\kappa_0^{(k)}, \ldots, \kappa_p^{(k)})$, $k = 1, \ldots, K$ are unknown parameters depending on the state k. The direction $\phi_{0,t}$ is assumed to have a von Mises distribution (see, Mardia and Jupp, 1999) that depends on state k. The mean direction is the direction of $\mathbf{V}_t^{(k)}$ and the concentration parameter is the length of $\mathbf{V}_t^{(k)}$, which correspond to the consensus model proposed by Rivest et al. (2016).

The traveled distances for the random walk model in state k are assumed to follow a distribution with density from the following exponential family (see Lehmann and Casella (2003), section 1.5):

$$g_k(d; \eta^{(k)}) = b(d) \exp\{\eta^{(k)\top} T(d) - A(\eta^{(k)})\}, d > 0, k = 1, \ldots, K. \tag{3.2}$$

In (3.2) $\eta^{(k)} \in \mathbb{R}^m$ is the vector of natural parameters, T is a $\mathbb{R}^m$ valued vector of sufficient statistics, b is a positive function, and A is a $\mathbb{R}$ valued function called the log-partition function.

An SSF that is equivalent to the random walk model specified by (3.1) and (3.2) has covariates that depend on the sufficient statistic $\{T(h_{jt})\}$ in (3.2) and the cosines of the differences between ϕ_{jt} and the directions to potential targets, $(\cos(\phi_{jt} - \phi_{0,t-1}), \cos(\phi_{jt} - \theta_{1t}), \ldots, \cos(\phi_{jt} - \theta_{pt}))^\top$. Thus the vector of explanatory variables for the SSF is

$$\mathbf{x}_{it} = (T(h_{it}), \cos(\phi_{it} - \phi_{0,t-1}), \cos(\phi_{it} - \theta_{1t}), \ldots, \cos(\phi_{it} - \theta_{pt}))^\top . \tag{3.3}$$

With this specification of $\mathbf{x}_{it}$, one has $\beta^{(k)} = (\eta^{(k)}, \kappa_0^{(k)}, \kappa_1^{(k)}, \ldots, \kappa_p^{(k)})$ and the function b in (3.2) appears in the SSF model as the offset $\log(b)$. The numerator of $p_t^{(k)}$ in (2.4) becomes

$$e^{\mathbf{x}_{0t}^\top \beta^{(k)} + \ln b(h_{0t})} = b(h_{0t}) e^{\eta^{(k)} T(h_{0t})} \times e^{\kappa_0^{(k)} \cos(\phi_{0t} - \phi_{0,t-1}) + \sum_{i=1}^p \kappa_i^{(k)} \cos(\phi_{0t} - \theta_{it})}, \tag{3.4}$$

which is the product of two terms, one for the distances and one for the directions. Note that (3.4) is the numerator of the time t contribution to the conditional likelihood function, for a given state, of the multi-state random walk model of Nicosia et al. (2017).

The denominator of $p_t^{(k)}$ is proportional to $J^{-1} \sum_{j=0}^J \exp(\mathbf{x}_{jt}^\top \beta^{(k)} + \log b(h_{jt}))$. The limit of this denominator as J goes to infinity depends on the way in which the controls have been selected. As mentioned in Section 2.2.2, the control angles $\{\phi_{jt}\}$ are drawn using a uniform distribution over $[0, 2\pi[$. One can sample the control distances uniformly in $[0, M]$, with a large M value as recommended in Section 2.2.2. With these methods for selecting the controls, the denominator is approximatively equal to the denominator of the time t contribution to conditional likelihood, for a given state, of the multi-state random walk model of Nicosia et al. (2017). Details are given in Supplementary Material B

Let us now suppose that the control distances are sampled from (3.2), with parameters $\tilde{\eta}$. Note that because the states are unobserved, the distribution from which the control locations are sampled cannot depend on the state, and hence $\tilde{\eta}$ is constant over k. In this case the offset does not appear anymore in $p_t^{(k)}$ as it is included in the density of the control distances that are sampled according to (3.2). The offset is needed only when the distances are sampled uniformly. Using the weak law of large numbers, the denominator of $p_t^{(k)}$ is then approximatively equal to a tilted version of the denominator of the time t contribution to conditional likelihood, for a given

state, of the multi-state random walk model of Nicosia et al. (2017). Thus in (3.2), the parameter $\eta^{(k)}$ is replaced by $\eta^{(k)} + \tilde{\eta}$. Hence if $\hat{\eta}^{(k)}_{\text{SSF}}$ is the SSF estimator for $\eta^{(k)}$, then the corresponding random walk model estimator is

$$\hat{\eta}^{(k)} = \hat{\eta}^{(k)}_{\text{SSF}} + \tilde{\eta}. \tag{3.5}$$

A detailed proof is provided in Supplementary Material section B . We have established that the multi-state random walk model is a special case of the multi-state SSF since any multi-state random walk model can be fitted by a multi-state SSF. Note however that every multi-state SSF cannot be fitted using a multi-state random walk model, as a random walk model cannot account for the local features of the pixels selected.

3.2. *Simulation Studies.* In Section 3.1 we established the relationship between the HMM-SSF and the multi-state random walk model models when J, the number of control locations sampled, is large using the law of large numbers. We now investigate whether this relationship holds for a finite value of J under both the uniform and the parametric sampling schemes for the control distances. We also assess the adequacy of the equation (3.5) for the estimators of the distance parameters when control distances are sampled from a parametric model. We follow the simulation studies of Nicosia et al. (2017) that investigated the statistical properties of a general random walk model. We intend to demonstrate that if we simulate a trajectory from the general multi-state random walk model then we can estimate its parameters using the multi-state SSF model proposed in Section 2.

We simulated the movement of one animal in the plane. The simulation procedure includes one target and it was placed at the center of the map and the covariate θ_t represents the direction from the animal at position P_t to this target at time step t. The simulation scenario consisted in repeating the following steps $N = 500$ times: (i) a time-homogeneous two-state Markov chain $S_{0:T}$ with transition matrix $\mathcal{P}$ is generated; (ii) at time 0, the animal is placed at a random position close to the south west corner of the map; (iii) at each time step t, $t = 1, 2, \ldots$, the location of the animal is obtained by simulating a direction ϕ_{t0} and a distance h_{t0} from the proposed general random walk model of Section 3.1 with ϕ_{0t} generated according to a consensus von Mises model with parameters $\kappa_0^{(k)}, \kappa_1^{(k)}$, $k = 1, 2$ and explanatory angles $\phi_{0,t-1}$ and θ_t and h_{0t} is simulated from a gamma distribution with shape parameter $\lambda_1^{(k)}$ and rate parameter $1/\lambda_2^{(k)}$; (iv) the simulation stops when the animal is within 30 distance units from the target. The gamma distribution belongs to the exponential family (3.2) with

sufficient statistics $T(d) = (\log(d), -d)$ and vector of natural parameters $\eta^{(k)} = \left(\lambda_1^{(k)} - 1, 1/\lambda_2^{(k)}\right)^T$ since its density can be written as

$$f(d; \lambda_1^{(k)}, \lambda_2^{(k)}) = \exp\left\{ \begin{pmatrix} \log(d) & -d \end{pmatrix} \begin{pmatrix} \lambda_1^{(k)} - 1 \\ 1/\lambda_2^{(k)} \end{pmatrix} - A(\lambda_1^{(k)}, \lambda_2^{(k)}) \right\},$$

where $A(\lambda_1^{(k)}, \lambda_2^{(k)}) = \log\left(\Gamma(\lambda_1^{(k)})\right) + \lambda_1^{(k)} \log(\lambda_2^{(k)})$.

The values of the parameters used in the simulations are given in Table 1. The scenario is one where the animal shows high directional persistence and high attraction to the target when in state 1, and high directional persistence and a moderate repulsion from the target in state 2.

TABLE 1
Parameters for the simulation scenario

$\mathcal{P} = \begin{pmatrix} 1-q_1 & q_1 \\ q_2 & 1-q_2 \end{pmatrix}$	$\begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$
$\kappa_0^{(1)} = \beta_3^{(1)}$	20
$\kappa_1^{(1)} = \beta_4^{(1)}$	15
$\lambda_1^{(1)} - 1 = \beta_1^{(1)}$	4
$1/\lambda_2^{(1)} = \beta_2^{(1)}$	10/7
$\kappa_0^{(2)} = \beta_3^{(2)}$	10
$\kappa_1^{(2)} = \beta_4^{(2)}$	-2
$\lambda_1^{(2)} - 1 = \beta_1^{(2)}$	0
$1/\lambda_2^{(2)} = \beta_2^{(2)}$	2

with q_1 the switch probability $\mathbb{P}(S_t = 2 | S_{t-1} = 1)$ while q_2 is associated with the reverse switch $\mathbb{P}(S_t = 1 | S_{t-1} = 2)$.

Once an animal's trajectory has been simulated, two sets of $J = 500$ control locations, for each visited location, are sampled. In the first one, the control distances are sampled uniformly over $[0, 15]$, where $M = 15$ is large enough to cover the supports of the gamma distributions in the two states up to their 99.9th percentiles. In the second set, the control distances are sampled from (3.2) with parameter $\tilde{\eta} = (0, 1)$, which actually corresponds to the exponential distribution with rate 1. The correspondence equation (3.5) for the SSF estimators $(\widehat{\eta_{1,SSF}^{(k)}}, \widehat{\eta_{2,SSF}^{(k)}}) = \left(\left(\widehat{\lambda_{1,SSF}^{(k)}} - 1\right), 1/\widehat{\lambda_{2,SSF}^{(k)}}\right)$

and the parameters of model (3.2) are:

$$\begin{aligned}\widehat{\eta_1^{(k)}} &= \widehat{\eta_{1,SSF}^{(k)}}, k = 1, 2.\\ \widehat{\eta_2^{(k)}} &= \widehat{\eta_{2,SSF}^{(k)}} + 1, k = 1, 2.\end{aligned}$$

When sampling control distances with a unit exponential distribution the covariates (3.3) are

$$\mathbf{x}_{it} = (\log(h_{it}), -h_{it}, \cos(\phi_{it} - \phi_{0,t-1}), \cos(\phi_{it} - \theta_t))^\top.$$

and the offset is $\log b(h_{0t}) = \log 1 = 0$. With this definition of $\mathbf{x}_{it}$, the parameters of the SSF are $(\beta_1^{(k)}, \beta_2^{(k)}, \beta_3^{(k)}, \beta_4^{(k)}) = (\lambda_1^{(k)} - 1, 1/\lambda_2^{(k)}, \kappa_0^{(k)}, \kappa_1^{(k)})$ for $k = 1, 2$.

To evaluate the sampling properties of the HMM-SSF estimators, the following statistical indicators were calculated:

$$b(\hat{\beta}) = \frac{1}{500}\sum_{i=1}^{500}(\hat{\beta}^{(i)} - \beta) \tag{3.6}$$

$$\mathrm{Sd}(\hat{\beta}) = \sqrt{\frac{1}{499}\sum_{i=1}^{500}(\hat{\beta}^{(i)} - \bar{\beta})^2}, \tag{3.7}$$

where β is a true parameter presented in Table 1, $\hat{\beta}^{(i)}$ the parameter estimate in the i-th simulation and $\bar{\beta}$ the mean of the estimates over the 500 simulations. Equation (3.6) gives the bias of the estimator, (3.7) its standard deviation. When the control distances are generated from the unit exponential, the conversion of SSF estimates into estimates of the gamma distributions for the travelled distances used (3.5) with $\tilde{\eta} = (0, 1)$. Maximum likelihood estimators of the parameters listed in Table 1 were also calculated using the algorithm in Nicosia et al. (2017). As these estimators were unbiased, Table 2 only reports their standard errors calculated using (3.7). This is useful to evaluate the loss of information associated with an estimation through a HMM-SSF.

Table 2 shows that the HMM-SSF recovers well the parameters of the random walk model when $J = 500$ under the two sampling schemes for the distances. Indeed the standard errors are comparable to those obtained when fitting the model by maximum likelihood using the algorithm in Nicosia et al. (2017). The biases are important when $J = 20$, especially for the distance parameters β_1 and β_2, in both states. This was expected since the relationship between HMM-SSF and random walk model estimators was

TABLE 2
Result of the $N = 500$ simulations with $J = 20, 500$. The bias $b(\hat{\beta})$ is presented together with standard deviation, $Sd(\hat{\beta})$, between parentheses.

Parameters	Multi-state random walk model	multi-state SSF estimation			
		Uniform sampling		Parametric (3.2) sampling	
True parameter	$\mathrm{Sd}(\hat{\beta})$	$J = 20$	$J = 500$	$J = 20$	$J = 500$
q_1	0.0194	-0.015 (0.03)	-0.000 (0.02)	0.013 (0.02)	0.002 (0.02)
q_2	0.0323	-0.028 (0.03)	0.002 (0.03)	-0.010 (0.04)	0.007 (0.03)
$\beta_3^{(1)}$	1.7449	2.160 (3.19)	0.177 (1.70)	1.682 (5.47)	0.305 (1.73)
$\beta_4^{(1)}$	1.3679	2.291(2.67)	0.212 (1.33)	0.644 (4.83)	0.212 (1.29)
$\beta_1^{(1)}$	0.4419	0.031 (0.98)	0.055 (0.43)	-1.691 (2.13)	0.045 (0.47)
$\beta_2^{(1)}$	0.0808	-0.008 (0.27)	0.014 (0.13)	-2.701 (0.77)	0 .007 (0.15)
$\beta_3^{(2)}$	1.2740	1.575 (2.53)	0.341 (1.27)	0.385 (1.66)	0.258 (1.11)
$\beta_4^{(2)}$	0.5531	0.688 (1.16)	-0.050 (0.58)	-0.244 (0.87)	-0.072 (0.49)
$\beta_1^{(2)}$	0.1044	-0.485 (0.35)	-0.006 (0.12)	-0.148 (0.23)	0.010 (0.10)
$\beta_2^{(2)}$	0.0638	-1.046 (0.53)	-0.040 (0.32)	-2.346 (0.80)	0.062 (0.29)

established using the law of large numbers which fails to hold when only $J = 20$ control locations are used.

An additional simulation study was carried out to investigate the impact of J, the number of control locations, and M, the radius of the circle used to draw the control distances, on the sampling properties of the SSF estimators. Nine scenarios, obtained by crossing three values of J and three values of M, were investigated. The three values of M were set equal to $q_{.6}$, $q_{.8}$, and $q_{.99}$, respectively the 60%, the 80% and the 99% percentile of the gamma distribution used to generate the distances in the traveling state (state 1). The results are presented in Table 3.

An additional illustration of the relationship between the two methods, either HMM-SSF or maximum likelihood, when fitting a multi-state random walk model is presented in Supplementary Material A wherea random walk model is fitted to data on bison movement and nearly identical estimates with the two estimation methods.

3.3. *State identification.* In this section, we investigate wether the underlying state S_t can be identified using estimates of the smooth probabilities $\mathbb{P}(S_t = k|\mathcal{F}_T)$, $t = 1, \ldots, T$, $k = 1, \ldots, K$ calculated in Appendix B .

The HMM-SSF was fitted to 25 simulated trajectories from the model with parameters given in Table 1 and the smooth probability estimates of being in state 1 where obtained for each step. Table 4 compares the true states and the predicted state when a step is predicted to be in state 1 when

TABLE 3
Result of the $N = 500$ simulations with $J = 20, 100$ and 500 and M as the $0.6, 0.8$ and 0.99 maximum quantiles of both travelled distances distribution. The bias $b(\hat{\beta})$, and standard deviation $Sd(\hat{\beta})$ between parentheses, are presented.

	$J = 20$	$J = 100$	$J = 500$
M	$M = q_{0.6}$	$M = q_{0.6}$	$M = q_{0.6}$
q_1	0.019 (0.25)	0.007 (0.02)	-0.012 (0.02)
q_2	-0.004 (0.04)	0.015 (0.03)	0.012 (0.03)
$\beta_3^{(1)}$	0.038 (2.80)	0.578 (2.17)	0.570 (1.85)
$\beta_4^{(1)}$	0.480 (2.43)	0.493 (1.81)	0.229 (1.53)
$\beta_1^{(1)}$	-4.615 (0.24)	-5.526 (0.14)	-4.811 (0.09)
$\beta_2^{(1)}$	-2.765 (0.16)	-2.954 (0.11)	-3.362 (0.14)
$\beta_3^{(2)}$	0.516 (1.81)	0.228 (1.46)	0.219 (1.34)
$\beta_4^{(2)}$	-0.331 (0.87)	-0.171 (0.65)	0.052 (0.66)
$\beta_1^{(2)}$	0.0184 (0.24)	0.034 (0.15)	0.015 (0.11)
$\beta_2^{(2)}$	0.077 (0.56)	0.034 (0.38)	-0.120 (0.33)
M	$M = q_{0.8}$	$M = q_{0.8}$	$M = q_{0.8}$
q_1	0.014 (0.02)	0.097 (0.02)	0.009 (0.02)
q_2	-0.011 (0.03)	-0.017 (0.03)	-0.021 (0.03)
$\beta_3^{(1)}$	0.509 (2.53)	0.108 (1.99)	-0.073 (2.02)
$\beta_4^{(1)}$	0.185 (2.12)	0.079 (1.60)	0.094 (1.61)
$\beta_1^{(1)}$	-4.049 (0.33)	-4.011 (0.22)	-4.202 (0.16)
$\beta_2^{(1)}$	-1.967 (0.14)	-2.002 (0.11)	-2.152 (0.12)
$\beta_3^{(2)}$	0.798 (2.25)	0.215 (1.47)	0.217 (1.33)
$\beta_4^{(2)}$	-0.157 (0.97)	-0.121 (0.66)	-0.128 (0.63)
$\beta_1^{(2)}$	-0.112 (0.26)	-0.004 (0.15)	- 0.001 (0.12)
$\beta_2^{(2)}$	0.318 (0.57)	-0.048 (0.38)	-0.077 (0.37)
M	$M = q_{0.99}$	$M = q_{0.99}$	$M = q_{0.99}$
q_1	-0.015 (0.03)	-0.0111 (0.02)	-0.000 (0.02)
q_2	-0.028 (0.03)	0.0032 (0.04)	0.002 (0.03)
$\beta_3^{(1)}$	2.160 (3.19)	0.3611 (2.10)	0.177 (1.70)
$\beta_4^{(1)}$	2.291 (2.67)	0.3018 (1.70)	0.212 (1.33)
$\beta_1^{(1)}$	0.031 (0.98)	-0.3679 (0.56)	0.055 (0.43)
$\beta_2^{(1)}$	-0.008 (0.27)	-0.1376 (-0.17)	0.014 (0.13)
$\beta_3^{(2)}$	1.575 (2.53)	0.4311 (1.43)	0.341 (1.27)
$\beta_4^{(2)}$	0.688 (1.16)	0.0776 (0.73)	-0.050 (0.58)
$\beta_1^{(2)}$	-0.485 (0.35)	0.0306 (0.16)	-0.006 (0.12)
$\beta_2^{(2)}$	-1.046 (0.53)	-0.0724 (0.73)	-0.040 (0.32)

its smooth probability estimate is larger than 0.75

Table 4 shows that the smooth probabilities provide and accurate predictor of the true state since the error rate is less than 7%. The same message

TABLE 4
Cross-classification table of the true know states and of the smooth probability estimates, calculated with a threshold of 0.75, for 25 trajectories.

True State:	1	2
Predicted State =1 0.75	6362	308
Predicted State =2 0.75	351	2979

is conveyed by the ROC curve presented in Figure 3 since the area under the curve is 0.98.

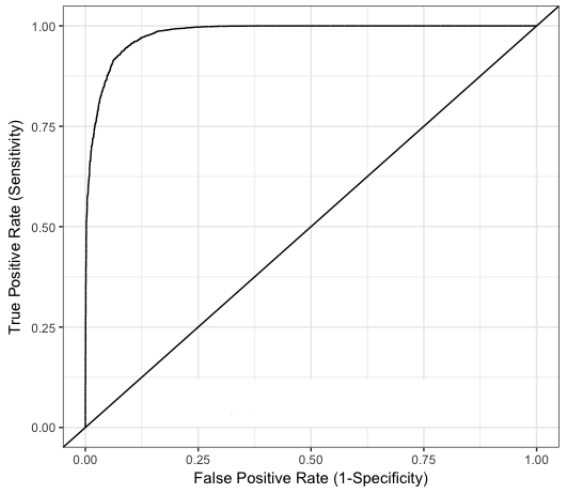
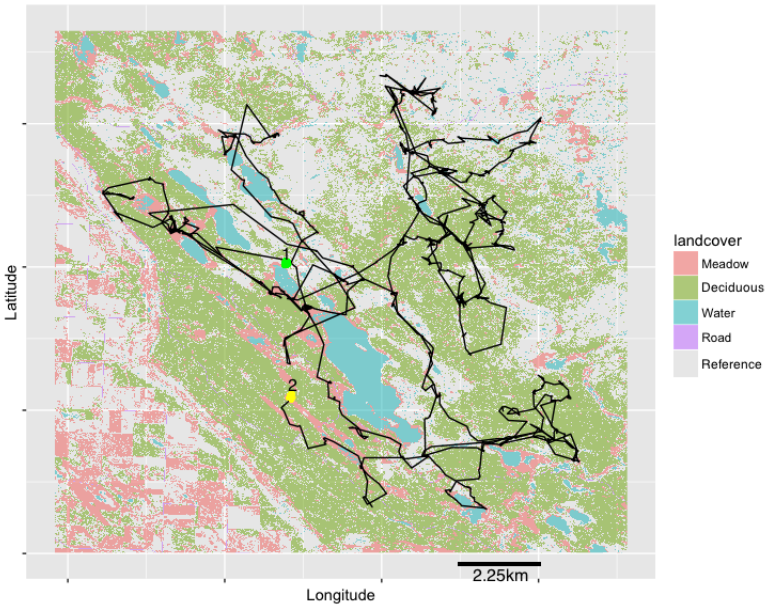


Fig 3: ROC curve for the association between an indicator variable taking the value 1 when the animal is in state 1 and the smooth probability estimates for being in state 1 calculated in the HMM-SSF.

4. Multi-state SSF model with animal movement and resource selection. Latombe et al. (2014) showed how a single state SSF can integrate both movement (angles and distances) and resource selection. Avgar et al. (2016) further studied the properties of the approach; they emphasized that an integrated analysis for both movement and local pixel selection was feasible through an SSF. When a large number of control locations are sampled, we have shown that the multi-state random walk model can be fitted thorugh an HMM-SSF. Thus the proposed model can be viewed as a multi-state version of the integrated Step Selection Analysis of Avgar et al. (2016). In this section we analyse the trajectory of an individual bison from November 2013 to April 2014 ($T = 3073$ hourly steps) in Prince Albert National Park, Saskatchewan, Canada. We show that two states can be distinguished

and that the movement and selection parameters can vary between states. Figure 4 depicts this trajectory.



copy2.png

Fig 4: Trajectory of a radio-collared bison from November 2013 to April 2014 in Prince Albert National Park. The green point (marked 1) is the departure and the yellow one (marked 2) the end. The landcover type 'Reference' denotes the reference landscape type which mainly consists of a mixture of deciduous and conifer forests.

During the winter season the bison tends to select more locations among meadows, water, roads or deciduous stands (Dancose, Fortin and Guo, 2011). We can therefore treat all other types of landscape as the baseline landscape level and fit the model with the following linear predictor:

$$
\begin{aligned}
\mathbf{x}_{jt}^T \beta^{(k)} &= \beta^{(k)}_{\text{cos.persis}} \cos(\phi_{jt} - \phi_{0,t-1}) \\
&+ \beta^{(k)}_{\text{dist.neg}}(-h_{jt}) + \beta^{(k)}_{\text{dist.log}} \log(h_{jt}) \\
&+ \beta^{(k)}_{\text{water}} z_{jt,\text{water}} + \beta^{(k)}_{\text{dec}} z_{jt,\text{dec}} \\
&+ \beta^{(k)}_{\text{meadow}} z_{jt,\text{meadow}} + \beta^{(k)}_{\text{road}} z_{jt,\text{road}}, \quad k = 1, 2
\end{aligned} \tag{4.1}
$$

$j = 0, \ldots, J$, where the explanatory variable $z_{jt,*}$ is the indicator that the location j at time step t is of type *, with * denoting one of the four types of landscape; for example $z_{jt,\text{water}} = 1$ if the j-th sampled location at time

step t is in the water and 0 otherwise. The $K = 2$ states was validated by exploratory analyses similar to those presented by Nicosia et al. (2017). Table 5 presents the estimates of the parameters by (4.2) along with their standard errors obtained when fitting the proposed SSF model with $J = 500$ uniformly sampled control locations. M was choosen to be the 99%-percentile of the observed distances and the control distances were drawn from the uniform distribution over $[0, M]$.

TABLE 5
Estimated parameters of the multi-state SSF model with $J = 500$ uniformly control locations. The estimated parameters are presented and the standard errors are given between parentheses.

Parameter	State 1: encamped	State 2: exploratory
q	0.247 (0.02)	0.161 (0.01)
$\beta_{\text{cos.persis}}$	-0.550 (0.05)	0.315 (0.05)
$\beta_{\text{dist.neg}}$	7.746 (0.57)	0.285 (0.02)
$\beta_{\text{dist.log}}$	0.475 (0.07)	-0.138 (0.04)
β_{water}	-0.132 (0.30)	-1.212 (0.37)
β_{dec}	0.057 (0.10)	-0.504 (0.10)
β_{meadow}	0.412 (0.09)	1.670 (0.09)
β_{road}	-0.247 (1.49)	1.533 (0.39)

As was the case in Section 3.1, the first state ($k = 1$) corresponds to an encamped state while the second state ($k = 2$) corresponds to an exploratory state, i.e., a traveling mode with a moderate significant directional persistence ($\hat{\beta}^{(2)}_{\text{cos.persis}} = 0.31,\ s.e. = 0.05$) and a larger average speed $(\hat{\beta}^{(2)}_{\text{dist.log}} + 1)/\hat{\beta}^{(2)}_{\text{dist.neg}} \approx 3$ km per hour. In the encamped regime, the animal is almost stationary, moving by about 0.19 km per hour. The directional persistence parameter $\hat{\beta}^{(1)}_{\text{cos.persis}}$ is strongly negative and significant, which means that the bison tends to move back and forth. In this general model the states are also related to the type of habitat. In the encamped state ($k = 1$) the bison prefers meadows, whereas in the exploratory state ($k = 2$), it selectively travels in meadows or roads while avoiding water and deciduous stands.

Table 5 presents the parameter estimates of the Markov chain model that governs the transitions between states. The stationary distribution of this fitted Markov chain gives a probability of being in state 1 of $\hat{q}_1/(\hat{q}_2 + \hat{q}_1) = 0.6053$, suggesting that the bison was in the travelling regime about 1860 out of $T = 3073$ steps. The model can be used to "predict" the state of the bison at time step t using the smooth probabilities $\mathbb{P}(S_{tk} = 1|\mathcal{F}_T; \hat{\theta}_{\text{MLE}})$, $t = 1, \ldots, T$, $k = 1, 2$ calculated in the filtering-smoothing part of the E-step of the EM algorithm. These predictions are depicted in Figure 5 with

a color gradient from red (state 2, $\mathbb{P}(S_{t1} = 1|\mathcal{F}_T; \hat{\theta}_{\text{MLE}}) = 1$) to blue (state 1, $\mathbb{P}(S_{t2} = 1|\mathcal{F}_T; \hat{\theta}_{\text{MLE}}) = 1$).

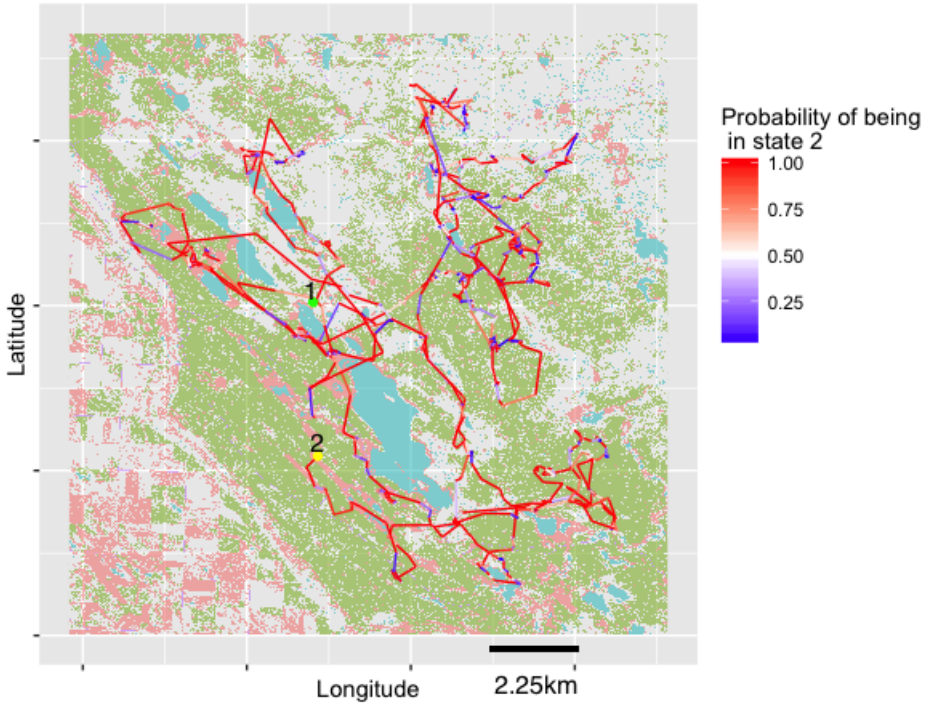


Fig 5: Estimated smooth probability for the trajectory of the bison in the landscape presented in Figure 4

5. Conclusion. This paper proposes a new multi-state version of the SSF model, the HMM-SSF, to describe the movement of an animal. It improves on classical multi-state random walk models by letting two important features of the movement evolve according to multiple behaviors: a global movement strategy and a local discrete habitat selection. The multi-state random walk model only considered the former. As such, the proposal generalizes the single-state integrated analysis proposed by Avgar et al. (2016) and used recently in Prokopenko, Boyce and Avgar (2016). By using recent techniques for the implementation of the EM algorithm in complex settings, we provide new statistical tools to fit a HMM-SSF and to identify the hidden behaviors of the animals.

We have proven that the multi-state random walk model of Nicosia et al. (2017) can be fitted using the proposed HMM-SSF. The HMM-SSF allows

to include explanatory variables that are more general than angles and distances, such as the type of land cover, in the analysis, thereby increasing the power of the method to characterize hidden states. When applied to the analysis of bison movement, the model successfully identified 1) foraging areas and 2) preferred trajectories when the bison moved between foraging areas. First, the strong selection for meadows in the encamped mode is consistent with bison spending more time where forage is most abundant. Indeed, bison consume grasses and sedges (plants) that are at least three times more abundant in meadows than forest stands (Fortin, 2007). Second, the association between the exploratory mode and habitat features provides valuable information on landscape connectivity. Landscape connectivity involves structural and functional components; structural connectivity depends the physical arrangement of habitat patches, such as their Euclidean distance (Tischendorf and Fahrig (2000), Kindlmann and Burel (2008)), whereas functional connectivity accounts for the movements within the patch network (Dancose, Fortin and Guo (2011), Courbin et al. (2014)). The exploratory mode model reveals that landscape functional connectivity for bison largely depends on their selective use of roads and meadows for travel, as well as their avoidance of water and deciduous forests relative to mixed and conifer forests. Our study thus demonstrates that a HMM-SSF can provide a mechanistic understanding of animal distribution dynamics in heterogeneous landscapes.

In our application, a land animal is followed using the GPS collar technology, which is relatively accurate when compared to other satellite telemetry, such as Argos archival data loggers that are used to track animal movements in environments like marine systems (Patterson et al. (2008), Albertsen et al. (2015)). Measurement error is therefore not an issue in the present example, but adding measurement error in the model could be an interesting future development of our method. About measurement, another interesting extension would be to consider irregular space time measures. This point seems difficult since it is not included in the classical theory of HMM. There are other possibilities of extension of the methods presented here. Because animals tend to exhibit heterogeneity in their movement behavior, it would be interesting to carry out the combined analysis of the movement of many individuals using a model featuring random effects. The approach based on combining individual models proposed by Murtaugh (2007) could be considered in the meantime. Defining a multi-state model based on a more complex hidden process could also be potentially interesting, for instance when trying to model the behavior of an animal over a long period of time (e.g., more than one "biological season", see Basille et al. (2012)) where the time

homogeneity assumption becomes questionable.

Acknowledgements. The authors are grateful to Marie-Caroline Prima for her help with the bison data and for insightful discussions. Financial support was provided by a grant from the Fonds de recherche du Québec - Nature et technologie to Louis-Paul Rivest, Thierry Duchesne and Daniel Fortin, and a scholarship from the Institut des sciences mathématiques awarded to Aurélien Nicosia. Field work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Parks Canada.

SUPPLEMENTARY MATERIAL

A : Relationship between methods when applied to bison trajectory
(doi: COMPLETED BY THE TYPESETTER; .pdf). Comparison of the multi-state SSF and random walk models using real bison movement data.

B : Proofs of the relationship between the proposed multi-state SSF model and the multi-state random walk model
(doi: COMPLETED BY THE TYPESETTER; .pdf). Theoretical proofs that the multi-state random walk model can be fitted using the proposed multi-state SSF model.

C : R code
(doi: COMPLETED BY THE TYPESETTER; .R). R code to fit the proposed HMM-SSF and the multi-state random walk model

D : Bison data set
(doi: COMPLETED BY THE TYPESETTER; .csv). Data of the bison trajectory and its habitat attributes used in Section 4

References.

Albertsen, C. M., W., K., Yurkowski, D., Nielsen, A. and Flemming, J. M. (2015). Fast fitting of non-Gaussian state-space models to animal movement data via Template Model Builder. *Ecology* **96** 2598–2604.

Avgar, T., Potts, J. R., Lewis, M. A. and Boyce, M. S. (2016). Integrated step selection analysis: bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution* **7** 619–630.

Avriel, M. (2003). *Nonlinear Programming: Analysis and Methods. Dover Books on Computer Science Series*. Dover Publications.

Basille, M., Fortin, D., Dussault, C., Ouellet, J.-P. and Courtois, R. (2012). Ecologically based definition of seasons clarifies predator-prey interactions. *Ecography* **36** 220–229.

Basille, M., Fortin, D., Dussault, C., Bastille-Rousseau, G., Ouellet, J.-P. and Courtois, R. (2015). Plastic response of fearful prey to the spatiotemporal dynamics of predator distribution. *Ecology* **96** 2622–2631.

Baum, L. E. and Petrie, T. (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Ann. Math. Statist.* **37** 1554–1563.

Biernacki, C., Celeux, G. and Govaert, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis* **41** 561–575.

Courbin, N., Fortin, D., Dussault, C. and Courtois, R. (2014). Logging-induced changes in habitat network connectivity shape behavioral interactions in the wolf–caribou–moose system. *Ecological Monographs* **84** 265–285.

Dancose, K., Fortin, D. and Guo, X. (2011). Mechanisms of functional connectivity: the case of free-ranging bison in a forest landscape. *Ecological Applications* **21** 1871–1885.

Duchesne, T., Fortin, D. and Rivest, L.-P. (2015). Equivalence between Step Selection Functions and Biased Correlated Random Walks for Statistical Inference on Animal Movement. *PLOS ONE* **10** e0122947.

Forester, J. D., Im, H. K. and Rathouz, P. J. (2009). Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology* **90** 3554–3565.

Fortin, M. E. (2007). Effets de la taille du groupe sur la sélection de l'habitat à plusieurs échelles spatio-temporelles par le bison des plaines (Bison bison bison). *MSc Thesis. Université Laval, Quebec, Quebec, Canada.*

Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T. and Mao, J. S. (2005). Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park. *Ecology* **86** 1320–1330.

Fryxell, J. M., Hazell, M., Borger, L., Dalziel, B. D., Haydon, D. T., Morales, J. M., McIntosh, T. and Rosatte, R. C. (2008). Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proceedings of the National Academy of Sciences* **105** 19114–19119.

Frühwirth-Schnatter, S. (2013). *Finite Mixture and Markov Switching Models (Springer Series in Statistics)*. Springer-Verlag New York.

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.

Holzmann, H., Munk, A., Suster, M. and Zucchini, W. (2006). Hidden Markov models for circular and linear-circular time series. *Environmental and Ecological Statistics* **13** 325–347.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied logistic regression. Wiley series in probability and statistics*. John Wiley & Sons, Inc. A Wiley-Interscience Publication, New York, Chichester, Weinheim.

Jonsen, I. D., Flemming, J. M. and Myers, R. A. (2005). Robuts state-space modeling of animal movement data. *Ecology* **86** 2874–2880.

Kindlmann, P. and Burel, F. (2008). Connectivity measures: a review. *Landscape Ecology* **23** 879–890.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D. and Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology* **93** 2336–2342.

Latombe, G., Parrott, L., Basille, M. and Fortin, D. (2014). Uniting Statistical and Individual-Based Approaches for Animal Movement Modelling. *PLoS ONE* **9** e99938.

Lehmann, E. L. and Casella, G. (2003). *Theory of Point Estimation. Springer Texts in Statistics*. Springer New York.

Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*. Wiley.

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.

Murtaugh, P. A. (2007). Simplicity and complexity in ecological data analysis. *Ecology* **88** 56–62.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. and Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences* **105** 19052–19059.

Nicosia, A., Duchesne, T., Rivest, L.-P. and Fortin, D. (2017). A general hidden state random walk model for animal movement. *Computational Statistics & Data Analysis* **105** 76 - 95.

Patil, G. P. (2005). *Weighted Distributions*. John Wiley and Sons, Ltd.

Patterson, T., Thomas, L., Wilcox, C., Ovaskainen, O. and Matthiopoulos, J. (2008). State–space models of individual animal movement. *Trends in Ecology & Evolution* **23** 87–94.

Prokopenko, C. M., Boyce, M. S. and Avgar, T. (2016). Characterizing wildlife behavioural responses to roads using integrated step selection analysis. *Journal of Applied Ecology* n/a–n/a.

Rivest, L.-P., Duchesne, T., Nicosia, A. and Fortin, D. (2016). A general angular regression model for the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society, series C* **65** 445-463.

Therneau, T. M. (2015). A Package for Survival Analysis in S version 2.38.

Tischendorf, L. and Fahrig, L. (2000). On the usage and measurement of landscape connectivity. *Oikos* **90** 7–19.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Turchin, P. (1998). *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Sinauer Associates Sunderland, Massachusetts, USA.

Vanak, A. T., Fortin, D., Thaker, M., Ogden, M., Owen, C., Greatwood, S. and Slotow, R. (2013). Moving to stay in place: behavioral mechanisms for coexistence of African large carnivores. *Ecology* **94** 2619–2631.

APPENDIX A: NUMERICAL IMPLEMENTATION OF MAXIMUM LIKELIHOOD ESTIMATION

Notation:

- $S_{k,t}$ is an indicator equal to 1 if $S_t = k$, i.e., if the animal is in state k at time step t
- π_{hk} is the transition probability from state h to k.

The EM algorithm is generally used for the maximization of likelihood functions when data are missing or unobserved. The EM algorithm only requires evaluation of the complete data log-likelihood function, which in

our case is easily derived from (2.3):

$$\begin{aligned}\log \mathrm{L}_{\mathrm{complete}}(\theta; \mathcal{F}_T^c) &= \sum_{t=1}^{T}\sum_{h=1}^{K}\sum_{k=1}^{K} S_{h,t-1}S_{k,t}\log \pi_{hk} \\ &+ \sum_{t=1}^{T}\sum_{k=1}^{K} S_{k,t}\log p_t^{(k)}.\end{aligned}$$

The EM algorithm consists of iterating an expectation (E) and a maximization (M) step. Let us denote by $\hat{\theta}_s$ the value of the estimate of θ after the s-th iteration of the algorithm. Then the $(s+1)$-th iteration of the algorithm starts with one application of the E-step, which evaluates the expectation of $\log \mathrm{L}_{complete}$ with respect to the conditional distribution of the missing values given the observed data, as follows:

$$\begin{aligned}Q(\theta|\hat{\theta}_s) &= \mathbb{E}_{S_{0:T}}\left[\log \mathrm{L}_{\mathrm{complete}}(\theta; \mathcal{F}_T^c)|\mathcal{F}_T^o, \hat{\theta}_s\right] \\ &= \sum_{t=1}^{T}\sum_{h=1}^{K}\sum_{k=1}^{K} \mathbb{E}(S_{h,t-1}S_{k,t}|\mathcal{F}_T^o, \hat{\theta}_s)\log \pi_{hk} \\ &+ \sum_{t=1}^{T}\sum_{k=1}^{K} \mathbb{E}(S_{k,t}|\mathcal{F}_T^o, \hat{\theta}_s)\log p_t^{(k)}.\end{aligned}$$

Then the value of $\hat{\theta}_{s+1}$ is calculated in the M-step as the value of θ that maximizes $Q(\theta|\hat{\theta}_s)$.

• **E step.** The function $Q(.|\hat{\theta}_s)$ involves two conditional expectations, $\mathbb{E}(S_{k,t}|\mathcal{F}_T^o, \hat{\theta}_s)$ and $\mathbb{E}(S_{h,t-1}S_{k,t}|\mathcal{F}_T^o, \hat{\theta}_s)$. These can be efficiently computed by a forward-backward (filtering-smoothing) algorithm for Markov chains, see Appendix D. The filtering-smoothing algorithm starts from the initial time $t = 0$ and computes the "filtering" probabilities $\mathbb{P}(S_t|\mathcal{F}_t^o)$ by using predictive probabilities $\mathbb{P}(S_t|\mathcal{F}_{t-1}^o)$ (going forward in time). The last filtering probability $\mathbb{P}(S_T|\mathcal{F}_T^o)$ is then used to compute the "smoothing" probabilities $\mathbb{P}(S_t|\mathcal{F}_T^o)$ using Bayes theorem (going backward in time). Details of this implementation of the E-step in a context of random walk model is given in the appendix of Nicosia et al. (2017).

• **M step.** For the M-step, we see that $Q(\theta|\hat{\theta}_s)$ is a sum of two functions that depend on different sets of parameters and can thus be maximized separately:

- When the latent states follow a Markov process, there is a closed form expression for the maximizer of the hidden process part,

$$\hat{\pi}_{hk}^{(s+1)} = \frac{\sum_{t=1}^{T} \mathbb{E}(S_{h,t-1} S_{k,t} | \mathcal{F}_T^o, \hat{\theta}_s)}{\sum_{t=1}^{T} \mathbb{E}(S_{h,t-1} | \mathcal{F}_T^o, \hat{\theta}_s)}, h, k = 1, \ldots, K.$$

- Since $p_t^{(k)}$ has a conditional logistic regression form, the log-likelihood for the observed choice can be maximized with respect to $\beta^{(k)}$ using a weighted maximum likelihood procedure (e.g. the function `coxph` with `weigths` in the survival `R` package, see Therneau (2015)).

Sampling Distributions. Quantities that are usually required for inference such as the value of the maximized log-likelihood for the observed data or an estimation of the variance matrix of $\hat{\theta}$ are not directly computed when using the EM-algorithm. The filtering-smoothing algorithm is used to evaluate the observed data likelihood (2.3). Moreover at each time t, one can evaluate the probability that the animal is in state k using the value of $\mathbb{E}(S_{k,t}|\mathcal{F}_T)$ in the "smoothing" part of the Filtering-Smoothing algorithm. Because we are able to compute $\log \mathrm{L}(\hat{\theta}_{\mathrm{MLE}})$, we can numerically approximate the negative of its hessian matrix, whose inverse, denoted v, is the usual estimate of the variance matrix of the maximum likelihood estimators. A numerical approximation of the Hessian matrix is available under most software implementations of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Avriel, 2003); in the data analysis section we use the one provided in the R function `optim`.

APPENDIX B: FILTERING-SMOOTHING ALGORITHM

- $S_{k,t}$ is an indicator equal to 1 if $S_t = k$, i.e., if the animal is in state k at time step t

In the E-step of the $(s+1)$-th iteration of the EM algorithm we have to compute two posterior expectations involving the hidden $S_{k,t}$, $k = 1, \ldots, K$, $t = 0, \ldots, T$, conditionally on the observed data $\mathcal{F}_T^o$:

$$\mathbb{E}(S_{k,t}|\mathcal{F}_T^o, \hat{\theta}_s) = \mathbb{P}(S_{k,t} = 1|\mathcal{F}_T^o, \hat{\theta}_s) \tag{D.1}$$

$$\mathbb{E}(S_{h,t-1}S_{k,t}|\mathcal{F}_T^o, \hat{\theta}_s) = \mathbb{P}(S_{h,t-1} = 1|S_{k,t} = 1, \mathcal{F}_T^o, \hat{\theta}_s)\mathbb{P}(S_{k,t} = 1|\mathcal{F}_T^o, \hat{\theta}_s), \tag{D.2}$$

where $\hat{\theta}_s$ is the maximized vector of parameters after the s-th step of the EM algorithm. The first probability on the RHS of (D.2) can be computed with

Bayes' theorem because, as we can see from Figure 1. , S_{t-1} is independent of the observed data from time t to T (i.e. $\{\mathcal{F}^o_{t+s}\}_{s\geq 0} \setminus \mathcal{F}^o_{t-1}$) given S_t and $\mathcal{F}^o_{t-1}$:

$$\mathbb{P}(S_{h,t-1}=1|S_{k,t}=1,\mathcal{F}^o_T,\hat{\theta}_s)=\frac{\hat{\pi}^{(s)}_{hk}\mathbb{P}(S_{h,t-1}=1|\mathcal{F}^o_{t-1},\hat{\theta}_s)}{\sum_{j=1}^K \hat{\pi}^{(s)}_{jk}\mathbb{P}(S_{j,t-1}=1|\mathcal{F}^o_{t-1},\hat{\theta}_s)},\ k=1,\dots,K,\ t=0,\dots,T.$$

Finally, to compute the remaining conditional probabilities in the posterior expectations (D.1) and (D.2), we adapt the filtering-smoothing algorithm of Frühwirth-Schnatter (2013).

FILTERING-SMOOTHING ALGORITHM TO IMPLEMENT THE E-STEP OF THE $(s+1)$-TH ITERATION OF THE EM ALGORITHM.

Filter Compute $\mathbb{P}(S_{l,t}=1|\mathcal{F}^o_t,\hat{\theta}_s)$, for every $l=1,\dots,K$:

$$\mathbb{P}(S_{l,t}=1|\mathcal{F}^o_t,\hat{\theta}_s)=\frac{p_t^{(l)}\mathbb{P}(S_{l,t}=1|\mathcal{F}^o_{t-1},\hat{\theta}_s)}{\sum_{k=1}^K p_t^{(k)}\mathbb{P}(S_{k,t}=1|\mathcal{F}^o_{t-1},\hat{\theta}_s)}$$

where $\mathbb{P}(S_{l,1}=1|\mathcal{F}^o_0,\hat{\theta}_s)=\sum_{k=1}^K \hat{\pi}^{(s)}_{kl}(\pi_0)_k$
and

$$\mathbb{P}(S_{l,t}=1|\mathcal{F}^o_{t-1},\hat{\theta}_s)=\sum_{k=1}^K \hat{\pi}^{(s)}_{kl}\mathbb{P}(S_{k,t-1}=1|\mathcal{F}^o_{t-1},\hat{\theta}_s)$$

for $t=2,\dots,T$.

Smooth Compute $\mathbb{P}(S_{l,t}=1|\mathcal{F}^o_T,\hat{\theta}_s)$, for every $l=1,\dots,K$:

S-step 1 For $t=T$, set $\mathbb{P}(S_{l,T}=1|\mathcal{F}^o_T,\hat{\theta}_s)$, the conditional probability computed at the last filtering step.

S-step 2 Recursion: For $t=T-1,\dots,0$, compute:

$$\mathbb{P}(S_{l,t}=1|\mathcal{F}^o_T,\hat{\theta}_s)=\sum_{k=1}^K \frac{\hat{\pi}^{(s)}_{lk}\mathbb{P}(S_{l,t}=1|\mathcal{F}^o_t,\hat{\theta}_s)\mathbb{P}(S_{k,t+1}=1|\mathcal{F}^o_T,\hat{\theta}_s)}{\sum_{j=1}^K \hat{\pi}^{(s)}_{jk}\mathbb{P}(S_{j,t}=1|\mathcal{F}^o_t,\hat{\theta}_s)}.$$

APPENDIX C: NUMERICAL DETAILS

Finding the Global Maximum of the Likelihood Function. Due to the complexity of the model, the EM algorithm may converge to local

or spurious maxima of the likelihood function. To deal with this, we run the EM algorithm with many random starting values for a few iterations and check for spurious and local solutions. We then choose the parameter values that yield the highest likelihood as the starting point of a new EM algorithm that we run until convergence. This strategy of combining short- and long-run EM algorithms to avoid possible local and spurious maxima is known as the 1em-EM algorithm (Biernacki, Celeux and Govaert (2003)). To obtain an estimate of the variance matrix of the maximum likelihood estimator we run one iteration of the quasi-Newton algorithm to obtain the value of the inverse of the Hessian matrix of the observed log-likelihood function evaluated at the maximum likelihood estimates. Here is an algorithmic description of this procedure.

FINDING THE GLOBAL MAXIMUM OF THE OBSERVED LOG-LIKELIHOOD FUNCTION

Preliminary step :

- Let $\theta_1, \ldots, \theta_N$ be N random initial starting values. (In our application of this method, we chose $N = 10$.)
- For $i = 1, \ldots, N$, run the EM algorithm until the first of (i) 50 iterations or (ii) the greatest relative difference in parameter value between successive iterations is less than 1%. Denote the estimators obtained at the end of this step $\hat{\theta}_i$, $i = 1, \ldots, N$.

Avoid spurious maxima :

- For each $\hat{\theta}_i$, $i = 1, \ldots, N$, compute the stationary distribution of the Markov chain, $\hat{\nu}_k^{(i)}$, $k = 1, \ldots, K$.
- Only keep the $\{\hat{\theta}_i, i \in I\}$ such that

$$\min_{k=1,\ldots,K} \hat{\nu}_k^{(i)} > \epsilon \quad \text{and} \quad \max_{j=1,\ldots,p;k=1,\ldots,K} |\beta_j^{(k)(i)}| < M.$$

(In our application of this method, we chose $\epsilon = 0.001$ and $M = 100$.)

Avoid local maxima :

- Put $\theta_0 = \arg\max_{i \in I} \mathrm{L}(\hat{\theta}_i)$.

Long-run EM algorithm :

- Start the EM algorithm at θ_0 and run it until the first of (i) 10 000 iterations or (ii) the greatest relative difference in parameter value between successive iterations is less than 10^{-8}.

Quasi-Newton iteration :

- Run one iteration of the quasi-Newton algorithm with the output of the long-run EM algorithm as initial value to get the final global maximum likelihood estimators of the model parameters and an estimation of their variance matrix.

A Note on the Initial Distribution $(\pi_0)_k, k = 1, \ldots, K$. Calculation of either the observed or complete data likelihood involves the initial distribution of the Markov chain, $(\pi_0)_k$, $k = 1, \ldots, K$. We have decided to fit the model twice. For the first fit we use $(\pi_0)_k = 1/K$, $k = 1, \ldots, K$, the uniform distribution over the K states. Then we fit the model again, but this time with $(\pi_0)_k = \hat{\nu}_k$, the stationary distribution of the chain computed from the first model fit.

A Note on the Identifiability of the Model up to State Label Switching. We can easily see that the value of the likelihood function remains the same if we relabel the states. We therefore define the states at the end of each M-step as follows: we give label i to the state with the i-th smallest $\hat{\beta}_0^{(k)}$.

Département de mathématiques et de statistique, Université Laval, Québec, Canada E-mail: aurelien.nicosia.1@ulaval.ca
E-mail: thierry.duchesne@mat.ulaval.ca
E-mail: louis-paul.rivest@mat.ulaval.ca
Département de Biologie, Université Laval, Québec, Canada E-mail: daniel.fortin@bio.ulaval.ca