Ecography 40: 788–799, 2017
doi: 10.1111/ecog.02587

Subject Editor: Brody Sandel. Editor-in-Chief: Miguel Araújo. Accepted 11 November 2016

'MigrateR': extending model-driven methods for classifying and quantifying animal movement behavior

Derek B. Spitz, Mark Hebblewhite and Thomas R. Stephenson

D. B. Spitz (spitzderek@gmail.com), Oregon State Univ. Dept of Fisheries and Wildlife, USA. – M. Hebblewhite, Univ. of Montana, College of Forestry and Conservation, USA. – T. R. Stephenson, California Dept of Fish and Wildlife, Sierra Nevada Bighorn Sheep Recovery Program, USA.

To be useful, definitions of animal movement behavior (e.g. migration) should be quantitatively rigorous, flexible enough to accommodate variation in species biology (e.g. latitudinal vs elevational movement) and sufficiently general to allow comparison among different species. Recent studies have applied a model-driven approach to classifying and quantifying animal movement from global positioning system (GPS) location data. We improve upon these methods by 1) revising model structure to provide a simple biologically-defensible basis to reduce misclassification; 2) introducing a data-efficient tool that can be used to quantify and circumvent model sensitivity to starting location; and 3) illustrating how existing models can be adapted to describe short-distance migration, using elevational migration as an example. These improvements are included in 'migrateR', an open source R package that expands and automates model-driven classification and quantification of animal movement behavior. We demonstrate the software and these improved methods using GPS-collar location data from a long-distance migrant, elk *Cervus elaphus*, and a short-distance elevational migrant, Sierra Nevada bighorn sheep *Ovis canadensis sierrae*. We provide in-text example code and a supplementary script illustrating how default options can be revised to meet several common challenges in fitting movement models.

Animal movement plays a central role in ecology, linking the discipline's twin concerns: distribution and abundance (Van Moorter et al. 2016). Recent improvements in animal-tracking technology have allowed researchers to monitor the movements of an increasingly diverse array of taxa in unprecedented detail, spurring near exponential growth in the use of global positioning system (GPS) transmitters and opening new avenues for research (Kays et al. 2015). The open source platform MOVEBANK, for example, now hosts 249 million locations from over 522 species in over 2000 individual studies (Kranstauber et al. 2011, <www.movebank.org/>). This influx of data on animal movement has brought migration to the fore as an area of research interest and conservation concern (Wilcove and Wikelski 2008). Migration is taxonomically pervasive, found in every major vertebrate and many invertebrate groups (Milner-Gulland et al. 2011). Furthermore, migratory species are often extremely abundant exerting a strong influence on ecosystem processes and biodiversity (Bauer and Hoye 2014). Unfortunately, migratory species currently face global declines that remain poorly understood (Wilcove and Wikelski 2008).

Studying migration first requires identifying migratory behavior, but defining migration remains challenging (Dingle and Drake 2007). Migratory behavior is broadly understood as a strategy for exploiting spatial resources that vary cyclically with time. Most research has focused on long-distance migration, but many species migrate short distances. Migration along steep elevation gradients, for example, is common in insects (Gutierrez and Wilson 2014), birds (Boyle et al. 2010), bats (McGuire and Boyle 2013) and ungulates (Albon and Langvatn 1992). Vertical migration of aquatic taxa is also common, but occurs on a daily rather than an annual cycle (e.g. zooplankton, Lampert 1989; fish, Beamish 1966; sea turtles, James et al. 2006; and sharks, Sims et al. 2006). Despite similarities, definitions of migration have remained inconsistent making migratory behavior difficult to compare across populations or taxa (Dingle 2006). Many studies rely on ad hoc classifications that are impossible to generalize or extend. The study of animal migration would benefit from a consistent definition that allows comparison across taxa while still accommodating species-specific differences in behavior.

Since its introduction, the model-driven approach of Bunnefeld et al. (2011; hereafter 'Bunnefeld et al.') has received increasing attention as a promising alternative to ad hoc classification. This method relies on net squared displacement (NSD), a metric based in movement theory, which measures the square of the straight-line distance between an animal's starting point and each subsequent location (Turchin 1998). To determine which movement behavior is best supported, researchers fit a family of a priori non-linear models, each representing a different behavior, to NSD data and then compare these models using Akaike information criteria (AIC, Burnham and Anderson 2002, Bunnefeld et al.). The structure of these models gives each parameter a biological meaning (e.g. the distance, duration and timing of migratory movement), allowing easy comparison and interpretation. Although most applications have focused on terrestrial mammals, this approach has been successfully applied to a variety of other taxa including reptiles, birds and fish (Blake et al. 2013, Nielsen et al. 2014, Beatty et al. 2015).

Nonetheless, model-driven methods for movement classification remain underused. The approach has been hampered by several technical challenges, largely the result of imperfect correspondence between mathematical models and animal behavior. Bunnefeld et al. noted bias against classifying resident behavior even in simplified simulations. This may be explained in part by seasonal home-range expansion, which is common among species living in seasonal environments (Wiktander et al. 2001) and creates temporal changes in the variation in NSD, which can easily be misidentified as migratory or dispersal movements. NSD models can also be sensitive to starting location because the calculation of all NSD values depends on the initial point (Naidoo et al. 2012). Recent work by Singh et al. (2016) focused on quantifying this sensitivity in Swedish moose *Alces alces* but the authors' approach is data intensive, requiring a minimum of two years of continuous location data, and has yet to be applied to another study system. Additionally, NSD models are insensitive to the short distances that characterize many animal migrations. Consequently, researchers often resort to ad-hoc reclassification (Mysterud et al. 2011, Naidoo et al. 2012), reducing the putative benefits of a standardized model-driven approach. Finally, this method is computationally complex, making it challenging to implement, especially for large datasets and populations with flexible migratory behavior.

Here, we introduce the 'migrateR' package for the R statistical environment (R Core Team) to 1) provide a quantitative and biologically-defensible basis for reducing misclassification of resident behavior; 2) quantify and address the sensitivity of existing NSD models to starting location; and 3) adapt NSD models to extend model-driven classification to elevational movement – a common form of short-distance migration. In addition to including these improvements, 'migrateR' makes model-driven movement analysis more accessible by automating these methods and providing graphical tools to assist model evaluation. We include an example analysis, below, which we divide into two parts, first illustrating migrateR's workflow and second comparing the performance of models fit to NSD, relative NSD (defined below) and elevation data from two ungulates: a long-distance migrant, elk *Cervus elaphus*, and an elevational migrant, federally-endangered Sierra Nevada bighorn sheep (*Ovis canadensis sierrae*). We include a supplementary script which fully reproduces our analysis of elk (Supplementary material Appendix 1; all data available in migrateR), while illustrating the code required to analyze bighorn data (a subset of which can be found in migrateR) in text. Further worked examples and a more thorough discussion of inputs can be found in the package's vignette. The full migrateR package is available on GitHub (<https://github.com/dbspitz/migrateR>).

Methods

Adapted NSD models

Following Bunnefeld et al. we used five a priori statistical models each representing a different behavior (Fig. 1). To classify the movement behavior of an animal-year of

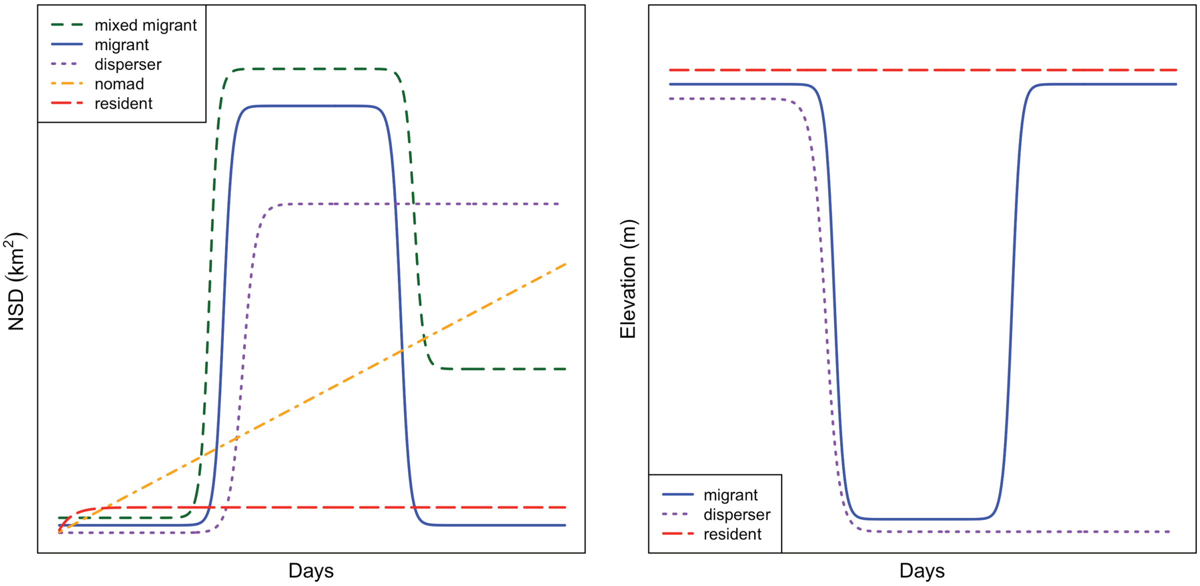

Figure 1. Conceptual illustration of net squared displacement (NSD) and elevation-based model families. The a priori model set fit to NSD covers a wider range of movement behaviors, than does the elevation model set. Elevation models shown illustrate downward movement, but can also accommodate upward migration (i.e. models can also appear as rotated 180 degrees around the x axis). See text for explanation of differences between models.

location data, we identified the best-supported behavior by comparing the fit of these a priori models using AIC (Burnham and Anderson 2002). Our first improvement to the methods of Bunnefeld et al. was to rearrange their notation to directly estimate the duration of migratory-range occupancy. This estimate is useful because it provides a consistent quantitative basis for excluding exploratory out-and-back movements from classification as migratory behavior. Many researchers have defined a minimum time of occupancy as a criterion of migratory behavior, e.g. individuals needed to spend >21 d on a separate seasonal range to be defined as a migrant (Cagnacci et al. 2011, Spitz 2015, Eggeman et al. 2016). Our models allow this additional criterion to be included either as an a priori restriction on model fit or, as previously, as an a priori decision rule implemented after model fitting. Directly estimating migratory-range occupancy required us to omit a direct estimate of the timing of return migration (θ_2; Bunnefeld et al.), which can instead be calculated as a derived parameter ($\theta_2 = \theta + 2 * \varphi_1 + 2 * \varphi_2 + \rho$, see the `theta2` function, below). Our NSD models are otherwise equivalent to those used by Bunnefeld et al. (and their improvement by Börger and Fryxell 2012, noted below). Thus our modified NSD model of migration (solid blue line in Fig. 1) was the double sigmoid:

$$NSD = \frac{\delta}{1+exp\left(\frac{\theta-t}{\varphi}\right)} + \frac{\delta}{1+exp\left(\frac{\theta+2*\varphi+2*\varphi_2+\rho-t}{\varphi_2}\right)} \quad (1)$$

where δ represents the distance separating seasonal ranges, t is time from start, θ is the midpoint of departing movement, φ is the time required to complete 1/2 to 3/4 of the migration (quantifying the duration of movement) and ρ is the length of time spent on the migratory range. Where included, subscripts differentiate estimates for return movements. Similarly we represent mixed migration (Fig. 1 dashed green line) with:

$$NSD = \frac{\delta}{1+exp\left(\frac{\theta-t}{\varphi}\right)} + \frac{\delta*\zeta}{1+exp\left(\frac{\theta+2*\varphi+2*\varphi_2+\rho-t}{\varphi_2}\right)} \quad (2)$$

which includes the addition of ζ allowing the distance traveled between ranges to vary by season (i.e. representing migration in which the individual doesn't return to the original range). Including ζ improves model convergence, but requires us to omit the δ_2 parameter (representing the distance traveled during return movement) included by Bunnefeld et al. This parameter can be derived from our model as $\delta_2 = \delta * \zeta$ (see the `delta2` function, below).

We represented dispersal (Fig. 1 dotted purple line) with a single sigmoid (i.e. movement with no return):

$$NSD = \frac{\delta}{1+exp\left(\frac{\theta-t}{\varphi}\right)} \quad (3)$$

where parameters are interpreted identically as in the migrant model, but θ represents the midpoint of dispersal rather than migration (Bunnefeld et al.). Following improvements made by Börger and Fryxell (2012) we represented residency (Fig. 1 dashed red line) as:

$$NSD = \gamma * \left[1 - exp(\kappa * t)\right] \quad (4)$$

where γ represents the mean NSD of locations in an individual's range and κ is the logarithm of the rate constant (quantifying the initial period of increase required for NSD to reach γ; Börger and Fryxell 2012). Finally, we represented nomadic behavior (dashed yellow line in Fig. 1) with the linear model:

$$NSD = \beta * t \quad (5)$$

where β is a constant and linearly increasing NSD represents the expectation under diffusion-based movement (Turchin 1998).

Adapted elevation models

NSD models only include spatial information from the two horizontal dimensions of a Cartesian plane. In many species, however, migratory movements follow a third vertical dimension. Therefore, we modified three movement models to replace NSD with vertical distance (hereafter 'elevation'; Fig. 1). We chose elevation rather than difference in elevation so that, in contrast to NSD models, our elevation models would be insensitive to an individual's initial position. We omitted nomad and mixed-migrant models because expectations for these behaviors with respect to elevation are unclear. Thus we represent migration (Fig. 1 solid blue line) with the double sigmoid:

$$\text{elevation} = \gamma - \frac{\delta}{1+exp\left(\frac{\theta-t}{\varphi}\right)} + \frac{\delta}{1+exp\left(\frac{\theta+2*\varphi+2*\varphi_2+\rho-t}{\varphi_2}\right)} \quad (6)$$

where γ represents the mean value for the starting range, now based on elevation rather than NSD. The interpretation of all remaining parameters is identical to our NSD models, except that δ represents vertical distance rather than NSD. Additionally, to allow analysis of incomplete animal-years of data, we also included a 'one way' model (analogous to the NSD disperser model, Eq. 3) to quantify unidirectional elevation movements. This one-way model (Fig. 1 dotted purple line; hereafter 'disperser') was parameterized as the single sigmoid:

$$\text{elevation} = \gamma - \frac{\delta}{1+exp\left(\frac{\theta-t}{\varphi}\right)} \quad (7)$$

with parameters interpreted as above. Finally, we represented residency (Fig. 1 dashed red line) with the horizontal linear model:

$$\text{elevation} = \gamma \quad (8)$$

Sensitivity to start date

We developed and automated a method for applying model selection criteria to test for and reduce sensitivity of NSD models to start date. An implicit assumption of the NSD method is that the starting point from which NSD is calculated occurs within an individual's starting range (Bunnefeld et al.). Many migrants, however, make exploratory movements either outside of their home range or between seasonal home ranges prior to migration (e.g. visiting a winter range during summer or fall). This behavior may be especially common among short-distance migrants and can confuse the classification of animal movement. Our approach is based on the calculation of what we term Relative Net Squared Displacement (rNSD), the net squared displacement relative to a chosen reference point (i.e. other than the starting location; NSD is equivalent to rNSD when the first point is used as the reference). rNSD can thus be directly compared to NSD as both terms are in the same units (km^2) and are calculated from identical location data. We then test for sensitivity to start date by calculating rNSD across a range of dates, fitting NSD models to the rNSD from each of these dates, comparing the minimum AIC across dates and choosing the reference date that results in the top model with the lowest overall AIC score. Relying on the best-supported rNSD should relax the need for assumptions about the first location, thereby improving model fit and ensuring a more stable and consistent classification of individual migration behavior. Divergence between rNSD and NSD thus suggest sensitivity of NSD models to starting location and indicate potentially problematic cases that may warrant further examination. We illustrate the resulting changes in classification and parameter estimates between rNSD and NSD models in our examples below.

Data requirements

The functions in migrateR are designed to capitalize on the animal movement trajectories (objects of class 'ltraj') introduced in the popular adehabitat packages (Calenge 2006). Multiple trajectories, for example from different individuals, called 'bursts', can be stored together in a single object with each burst including spatial coordinates, timestamp and, optionally, other values associated with each location (elevation models can not be fit unless elevation values are included as an 'infolocs' field titled 'elev'). For assistance in organizing data in this format, please refer to the very thorough adehabitat documentation (Calenge 2006).

Even with migrateR, fitting movement models can be difficult and users will benefit from first taking the time to carefully organize their data. For best results we recommend separating bursts by animal and study year such that each burst is less than or equal to a year in length. If individuals make more than one migratory movement annually, dividing these movements into shorter separate bursts can improve model fit. Nonetheless, fitting movement models to trajectories shorter than a year or with gaps in data can be difficult and will often require a closer attention to starting parameter estimates and constraints. When possible we also recommend defining study years such that migratory and other movements are contained within single bursts (i.e. all movements are unlikely to either begin before a burst's first or end after a burst's last location). Where previous research on a species' movement behavior is available, this may be accomplished a priori, but in other cases experimentation may be necessary to arrive at a definition of migratory year that minimizes segmentation of movement events. Finally, following Bunnefeld et al. we recommend subsampling ('thinning') each burst to one point/day.

MigrateR includes two example datasets that illustrate our recommendations for data formatting and organization. The `elk` dataset includes 23 animal-years of data from adult female elk in Alberta, Canada collected from 23 individuals between 2003–2005 (the complete dataset, i.e. before subsampling, is available online at doi: 10.5441/001/1.k8s2g5v7; Hebblewhite and Merrill 2016). The dataset `bighorn`, a subset of the larger dataset used in our comparative example, below, includes four-animal years of data (2007–2011) from a single Sierra Nevada bighorn sheep female (we have limited the Sierra bighorn data included due to the species' Federally-endangered status and have offset bighorn coordinates by a constant to mask their location). Thus the analysis of `elk` we provide below can be reproduced using a script we include as Supplementary material Appendix 1 and for purposes of comparison we also include results from the full dataset from which `bighorn` was drawn. The in-text example code we provide below to illustrate our workflow relies primarily on `bighorn`, whose small size makes it more convenient to readers inclined to reproduce these examples while reading.

Example: workflow

Functions in migrateR can be divided by task into three categories: fitting models, checking model fit, and extracting the resulting parameter estimates (Table 1). Before using these function, however, the user must first decide on an appropriate model family. When possible, we recommend that this choice be made a priori and be based on available information concerning a species' migratory behavior and/or the motivation(s) underlying the analysis. While migration may involve both long-distance movement and changes in elevation, this need not be the case. Even when movements do include both long distances and changes in elevation, these transitions are unlikely to occur synchronously. Thus, even where either NSD or elevation models could be appropriate, they are likely to offer distinct characterizations of migratory behavior. In most cases, however, we expect that one model will more clearly match a species' ecology (i.e. either NSD/rNSD or elevation data will show a clearer pattern; regrettably, because NSD/rNSD models and elevation models rely on different underlying data, their fit can not be compared directly using AIC). The strength and nature of inference will depend on the chosen model family's correspondence to data and as such this step warrants careful consideration.

We fit NSD, rNSD and elevation models using the `mvmtClass` function in the package migrateR. This function takes animal movement trajectories (i.e. an object of class 'ltraj' containing one or more movement bursts) as input (with spatial coordinates, timestamp and, optionally, elevation values, for each location; Calenge 2006) and outputs movement models fit separately to each burst of location data. To store these results efficiency, migrateR introduces the 'mvmts' and 'mvmt' classes. A 'mvmts' object is a named list with one element of class 'mvmt' for each burst analyzed. These 'mvmt' objects each contain any models successfully fit to data from that burst, the data used to fit these models (e.g. t and NSD, rNSD or elevation, Eq. 1–8) and other information pertaining to the call to `mvmtClass"`. (MigrateR also introduces novel methods, supporting the application of several common R functions to 'mvmt(s)' class objects, which we describe below.) Finally, like many functions in

Table 1. Contents of migrateR arranged by task and approximate order of workflow. Names followed by asterisks indicate preexisting R functions, for which migrateR includes novel methods. Examples illustrating the application of these functions are included in text, the supplementary script and the packages' vignette. See migrateR's help documentation and vignette for further details.

	Name	Description
Example datasets		
	`bighorn`	4 animal-years of location data, divided by study year, from a single adult female Sierra bighorn sheep
	`elk`	23 animal-years of location data, divided by study year, from 23 adult female elk
Fitting models		
	`mvmtClass`	Fits suite of NSD, rNSD or elevation models to location data
	`findrloc`	Finds the best supported reference location for rNSD (required to fit rNSD models)
	`pEst`	Organizes starting parameter estimates and constraints as required by mvmtClass
	`refine`	Attempts to improve model fit using new starting parameter estimates and constraints
Checking model fit		
	`fullmvmt`	Checks whether models were successfully fit
	`topmvmt`	Identifies the best-supported movement model for each burst
	`print*`	Organizes output from mvmtClass
	`summary*`	Summarizes output from mvmtClass (called by print)
	`plot*`	Plots movement models and the data used to fit them
Extracting estimates		
	`mvmt2df`	Organizes parameter estimates from a list of (top) models
	`mvmt2dt`	Estimates start and end time for movement events (e.g. migration) from chosen model
	`theta2`	Derives timing estimates for return movement (θ_2) from 'migrant' or 'mixed-migrant' models
	`delta2`	Derives distance estimates of return movement (δ_2) from 'mixed-migrant' models

migrateR, `mvmtClass` relies on default inputs to facilitate its use. Users can instead choose to specify these inputs manually. Where possible, we include explanation of these inputs with our in-text examples. For further details and sample code we recommend consulting the package vignette (`'vignette(migrateR)'`).

NSD models based on Cartesian (x, y) coordinates are fit by default, but if elevation values are included, elevation models can be fit instead by setting the optional argument `fam="elev"`. To fit both NSD and elevation models to the example dataset `bighorn`, we used the commands:

```
bighorn.nsd <- mvmtClass(bighorn, stdt=
"10-31")
bighorn.elev <- mvmtClass(bighorn, fam=
"elev", stdt="10-31")
```

The 'stdt' argument (formatted as 'month-day') ensures that estimates of migratory timing are comparable by defining a start date from which these estimates (i.e. θ and θ_2, whose units are decimal days) are calculated. If a start date is not specified, the date of the first location in each burst will be substituted with the result that θ estimates may not be directly comparable among bursts. If all bursts begin on the same calendar day and time, this argument becomes superfluous.

Fitting rNSD models also depends on the `mvmtClass` function, but first requires finding the best-supported reference location using the function `findrloc`. This function calculates rNSD for each location included in the first 15 d of the burst, then fits NSD models to NSD and the rNSD calculated relative to each of these locations, compares the results using AIC and returns a 'data.frame' including the record number of the reference location that results in the top model with the lowest overall AIC. Thus `findrloc` can be used to test for sensitivity to starting location, but does not itself output movement models. To fit models to the best-supported rNSD we first used `findrloc` to find the best supported reference locations and then used the optional 'rloc' argument in `mvmtClass` to fit models to rNSD for these locations:

```
bhs.rlocs <- findrloc(bighorn, stdt=
"10-31")
bhs.rnsd<-mvmtClass(bighorn, rloc=bhs.
rlocs$rloc, stdt="10-31")
```

The range of reference locations considered can be changed (from locations within the first 15 d) using `findrloc`'s optional argument 'max.rloc'. There are two cases to be made against increasing the number of reference locations to include all points in a burst: first, such increases have a multiplicative effect on runtime (e.g. approximately 2 s to fit NSD models to all 23 bursts in `elk` and approximately 30 seconds to find the best supported reference date for each of these bursts using 'max.rloc's' default value of 15; runtimes calculated from a machine running a 2.7 GHz intel core i7 with 8Gb RAM) and second, including a reference date that overlaps a dispersal movement will compromises the fit of the disperser model (in these cases the best supported reference date will be the midpoint of dispersal movement; the first burst in `elk` provides a good example of this). Thus, in some cases it may be appropriate to specify a narrower range of possible reference location to ensure that the range of candidate reference locations precedes any movement event included in the burst. While we recommend the default of 15 d as a starting point appropriate to most datasets, choosing an appropriate range of possible reference locations ultimately depends on the definition of study year and on a species' ecology.

To fit models, `mvmtClass` relies on starting parameter estimates and restrictions provided by the `pEst` function. This function can be used to manually specify starting values or restrict the range of terms included in movement models. Running `pEst()` returns a 'data.frame' containing default values. The missing values in this 'data.frame' are filled dynamically by `mvmtClass`. By default, `pEst` specifies that migration can't occur before the first location observed

($\theta > 0$) and restricts the duration of migratory movements ($1 \leq \varphi \leq 21$ d; total duration of migration ~ 4 φ, approximately 4 to 84 d). Users can change these defaults using `mvmtClass'` 'p.est' argument (we provide an example of these changes below). Adjusting restrictions on θ is especially important if the first location in a burst falls between the beginning and end of a migratory or dispersal movement (i.e. if a burst depicts a movement that began before its first point). Where possible, we recommend choosing the start of the study year to avoid this complication. Increasing φ increases the flexibility of migrant, mixed migrant and disperser models, but may lead to over-smoothing. Highly restrictive parameter constraints, on the other hand, can prevent some models from converging. We therefore generally recommend applying such a priori constraints as decision rules following model fit (e.g. vertical movement by bighorn sheep must be a minimum of 500 m – $|\delta| \geq 500$ – to qualify as a migration). Poor choice of starting values can also impede model convergence.

Any models that fail to converge are considered unsuccessful and are therefore omitted from `mvmtClass'` output. If models are omitted, `mvmtClass` issues a warning indicating the unsuccessful models and the bursts to which they belong. Missing models can also be identified by testing output from `mvmtClass` with the `fullmvmt` function. This function can return a variety of information about missing models, depending on which value is selected for the 'out' argument. Thus applying `fullmvmt` to the elevation models we fit to `bighorn` reveals that the complete suite of models was successfully fit to three of the four bursts ('`fullmvmt(bhs.elev)`'), that the remaining bursts is missing a single model ('`fullmvmt(bhs.elev, out="numer")`'), and that the missing model is 'migrant' ('`fullmvmt(bhs.elev, out="name")`').

Most problems with model fitting are a result of poor correspondence between bursts and parameter constraints or starting values. For datasets containing variable behavior, like `elk` and `bighorn`, a single set of parameter estimates may be insufficient for fitting all models and multiple sets may be required. A related but distinct problem is that poorly chosen starting values or constraints don't always cause convergence failure, but can instead force `mvmtClass` to fit suboptimal models. The `refine` function provides an answers to both of these challenges. This function requires two arguments, output from `mvmtClass` and a new set of starting parameter values and constraints (supplied by `pEst`), attempts to refit each model using the new constraints, and compares these models to the original results, keeping only the single model of each type with the lowest AIC. So, for example, running

```
p.est2 <- pEst(u.d=0)
bhs.elev2   <-   refine(bhs.elev,   p.est=
p.est2)
```

adds a new constraint (an upper limit on migratory distance, δ), forcing all migrant models to represent downward movement (i.e. $\delta < 0$), and returns a 'mvmts' class object which differs from the input ('bhs.elev') by adding a previously omitted migration model to the second burst and replacing the migrant model in the third burst with a new model with lower AIC. Note that both combinations of constraints are needed to fit all models; relying only on the second set of constraints, i.e. `mvmtClass(bighorn, fam="elev", p.est=p.est2, stdt="10-31")`, will also result in omitted models (instead omitting 'disperser' models for the second and third bursts). Because the output of `refine` is equivalent to that of `mvmtClass`, it can be fed back into `refine` to iteratively test for improvements in model fit. The greater the variety of behaviors included in a dataset, the greater the number of iterations likely necessary to fit all models. Plotting a burst often provides valuable clues towards specifying parameter constraints and starting values that will improve model fit (particularly in choosing appropriate starting values for the timing of movement, θ).

Once models are fit, the top model for each trajectory can be easily identified using the function `topmvmt`, which compares AIC values by burst to determine which movement models received the greatest support. We determined the best-supported movement models using, e.g.:

```
top.bhs.elev2 <- topmvmt(bhs.elev2)
```

which provides a list of named models. We then extracted the classifications using

```
table(names(top.bhs.elev2))
```

which returns a table featuring the number of bursts classified as belonging to each movement strategy. The `topmvmt` function also includes several optional arguments that can be used to restrict how a top model is selected, for example by allowing categorical exclusion of a model (e.g. 'nomad') or excluding models whose parameters fail to meet a minimum threshold (i.e. for distance traveled, δ, or length of migratory-range occupancy, ρ). Further worked examples illustrating these options and their application are included in the vignette.

In addition to new functions, migrateR also includes novel methods that increase the utility of several functions already familiar to R users. The default classifications, described above, are also output by `summary`, shown whenever a 'mvmt(s)' class object is printed (e.g. '`summary(bhs.elev2)`' or simply '`bhs.elev2`'). Furthermore, 'mvmt(s)' class objects can easily be inspected visually using the function `plot`. 'Mvmt' objects will plot directly, while 'mvmts' objects introduce a prompt between each burst to allow easy cycling through multiple plots. Thus, `plot(bhs.elev2[[1]])` returns a single plot (the first burst, Fig. 2B), while `plot(bhs.elev2)` introduces a series of prompts allowing the sequential plotting of each bust in 'bhs.elev2'. Figure 2 and 3 are direct outputs of this function. Note that while the x-axis for these plots is held constant, the range of the y-axis will vary by burst unless a consistent y range is specified using `plot`'s 'ylim' argument (e.g. Fig. 3). Models are listed in the legend in order of AIC (lowest to highest), but caninstead be shown alphabetically by setting the optional argument `ranked=FALSE`.

We strongly recommend that all models be examined visually. Model-driven movement classification is subject to a common criticism of AIC-based model selection; a top model will always be identified even when all models inadequately represent the data (Burnham and Anderson 2002).

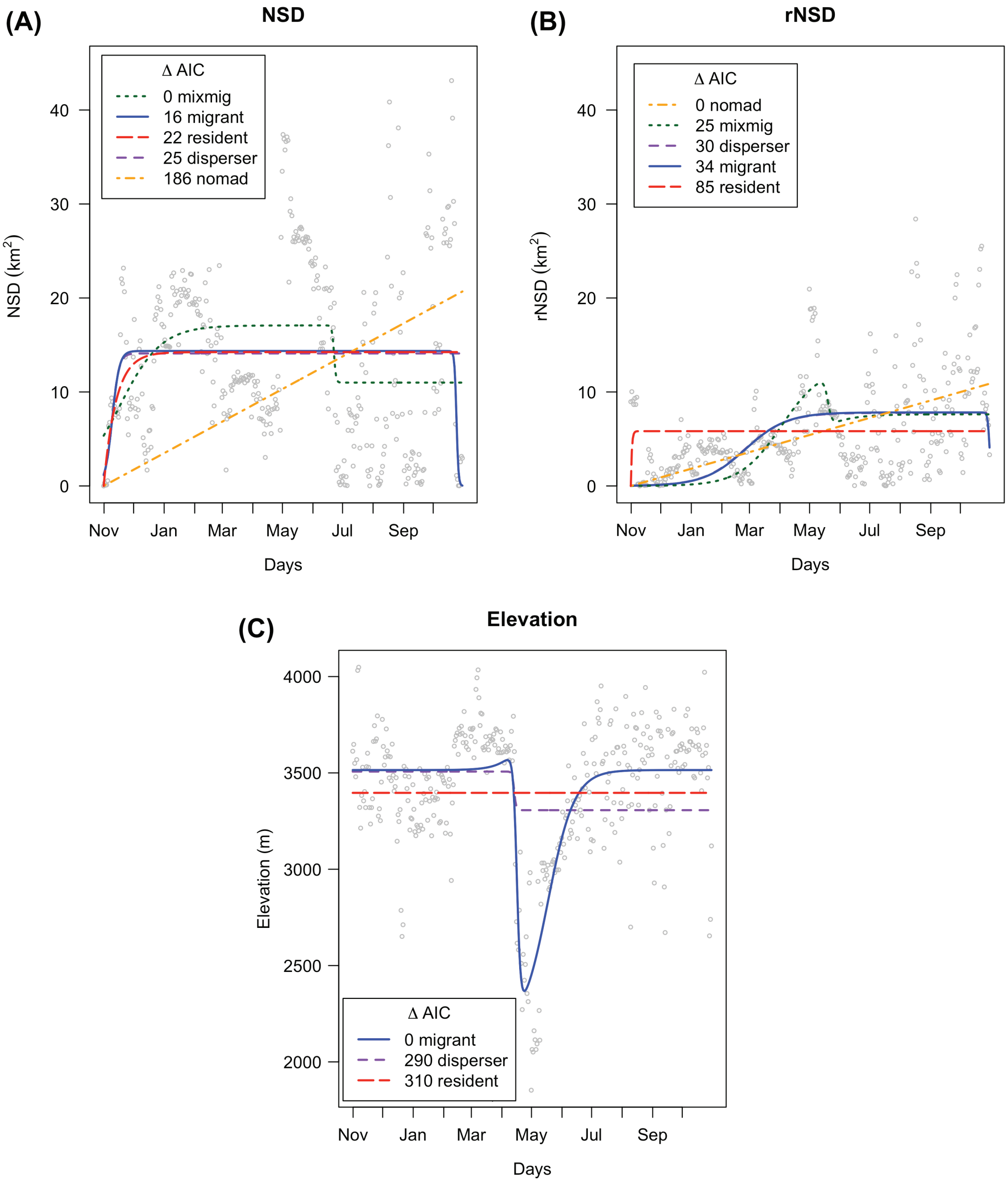

Figure 2. Illustration of differences in the model-driven characterization of movement by net squared displacement (NSD), relative NSD (rNSD; reference location = 11) and elevation methods for Sierra Nevada bighorn sheep *Ovis canadensis sierrae*. NSD data show no clear pattern and rNSD data display a temporal shift in variance (beginning around 1 July) that advantages the most complex model (mixed migration). Elevation values from the same location data show a comparatively regular pattern. Consequently, the elevation models follow these data more closely. Although all three methods classified the movement as migratory (or mixed-migratory), their characterizations of migratory behavior differed in every characteristic we considered. Estimation of the period of winter-range residency, for example, varied from nearly 11 months (A) to less than one (B), but there were also large differences in the speed and timing of departing and returning movements.

Thus even the best supported model may represent a poor choice. Visual comparison remains an important tool for identifying this problem and is therefore key to applying this method critically (Mysterud et al. 2011; Fig. 2, Fig. 3). If all models fit the data poorly, it may be worth revisiting earlier steps in the analysis (e.g. definition of the study year, choice of model family) to see if changes can improve the correspondence between models and data.

For some analyses, quantitative definition of movement behavior may be more important than movement classification. MigrateR has four functions that help either to extract or derive these estimates. The `mvmt2df` function takes a list of movement models as its argument, as output by `topmvmt`, and returns a list containing a 'data.frame' of parameter estimates for each type of movement model included in the input. Thus

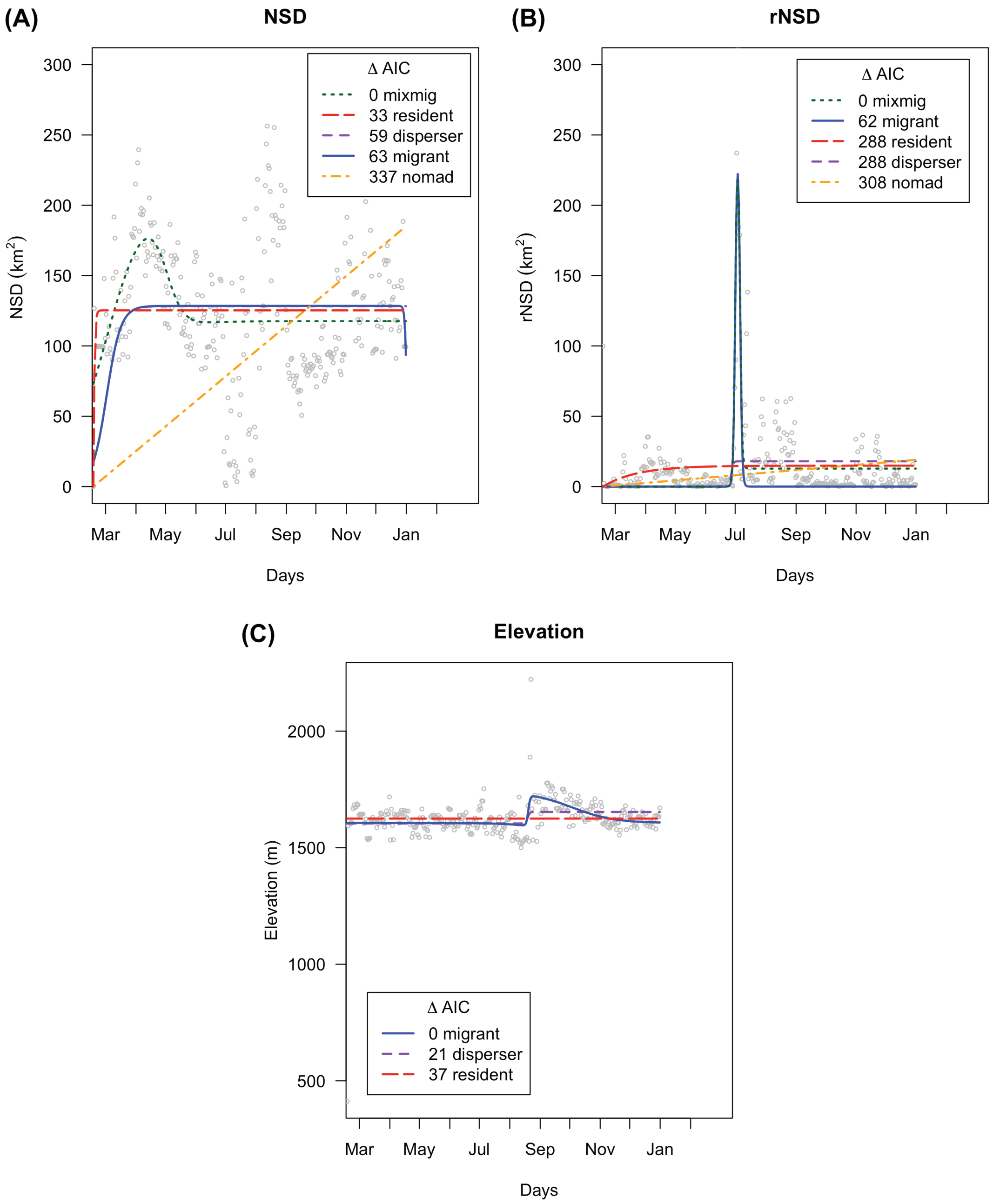

Figure 3. Illustration of differences in the model-driven characterization of movement by net squared displacement (NSD), relative NSD (rNSD; reference location = 4) and elevation methods for elk *Cervus elaphus*. Our rNSD method reduces the influence of outlying points, thus improving correspondence between the location data and the model receiving the greatest support. All NSD models show a relatively poor fit to summer, fall and early winter locations (A). In contrast, rNSD models more closely fit the data with the top rNSD model showing a concerted out-and-back movement in late June/early July (B). The elevation model (C) provides comparatively little information, indicating a fairly constant elevation interrupted by a slight upward movement (< 150 m) August–November. (After running section I and II of the supplementary script, these plots can be reproduced using ‘plot(nsd[[6]], ylim = c(0, 300)’, ‘plot(rnsd[[6]], ylim = c(0, 300)’ and ‘plot(elev[[6]])’.)

```
top.bhs.nsd <- topmvmt(bhs.nsd)
mvmt2df(top.bhs.nsd)
```

returns a list of two ‘data.frames’, containing parameter estimates for ‘mixed migrant’ models (‘mixmig’), and ‘nomad’, respectively. The `delta2` and `theta2` functions both take ‘mvmt(s)’ objects as input and can be used to derive estimates and standard errors for the parameters from Bunnefeld et al. that our models omit (δ_2 and θ_2, respectively). For example, `delta2(bhs.rnsd)` returns a ‘data.frame’ including δ_2 estimates and standard errors for every burst in ‘bhs.rnsd’ with a ‘mixed migrant’ model. Because θ_2 can be derived from both ‘migrant’ and ‘mixed migrant’ models, `theta2` has an additional argument, ‘mod’ specifying the desired model (either ‘migrant’ – the default – or ‘mixmig’). For example, `theta2(bhs.rnsd, mod=“mixmig”)` returns esti-

Table 2. Comparison of classification using NSD and rNSD approaches. Results shown are for n = 88 Sierra Nevada bighorn sheep *Ovis canadensis sierrae*; A) and n = 23 elk *Cervus elaphus*; B). We show the classifications we regard as consistent in bold. Elk show fewer differences in classification (B, 2 of 23) than do bighorn sheep (A, 13 of 88).

				rNSD			
A) Bighorn		disperser	migrant	mixed migrant	nomad	resident	total
NSD	disperser	**0**	0	3	1	0	4
	migrant	0	**1**	**3**	0	0	4
	mixed migrant	0	**0**	**70**	1	2	73
	nomad	0	0	4	**0**	0	4
	resident	0	0	1	1	**1**	3
	total	0	1	81	3	3	88

				rNSD			
B) Elk		disperser	migrant	mixed migrant	nomad	resident	total
NSD	disperser	**1**	0	0	1	0	2
	migrant	0	**8**	**1**	0	0	9
	mixed migrant	0	**0**	**11**	0	1	12
	nomad	0	0	0	**0**	0	0
	resident	0	0	0	0	**0**	0
	total	1	8	12	1	1	23

mates for all 'mixed migrant' models in 'bhs.rnsd'. Finally, the `mvmt2dt` function estimates the times at which movements begin and end (as opposed to their midpoints, θ and/or θ_2). The argument 'p', the percent of the total movement distance (δ), defines the thresholds used to calculate these dates. Start and end dates are calculated as the time at which model predictions reach $p * \delta$ and $(1 - p) * \delta$, respectively. Like `theta2`, `mvmt2dt` has the 'mod' argument, which specifies the model from which values should be calculated. Thus setting `mvmt2dt(bhs.elev2, p=0.1, mod="disperser")` returns a list of the start and end of movement from all 'dispersal' models in 'bhs.rnsd' (i.e. the time at which these movements are 10 and 90% complete). These dates are returned in two forms, decimal day ('dday', calculated either from a burst's first location or from 'stdt' if previously specified) and as a date formatted as POSIXct.

Example: comparing elk and bighorn sheep

To illustrate the performance of NSD, rNSD and elevation models, we compare the fit of these models to data from two ungulates, a long distance migrant, elk and an elevational migrant, Sierra bighorn. We classified movement behavior for 23 animal years of elk location data from Alberta, Canada (each animal year from a different individual; all data included in migrateR, see help documentation for `elk`), and 88 animal-years of location data collected from 65 female Sierra Nevada bighorn sheep (of which a small illustrative subset of n = 4 animal years is included in migrateR, see help documentation for `bighorn`) in three ways, based on: 1) NSD; 2) rNSD; 3) elevation. After comparing classification of animal-years across these methods, whenever NSD classification agreed with either rNSD- or elevation-based classification in identifying migratory behavior, we also calculated differences in the absolute value of estimates of timing parameters. We calculated summary statistics for these differences and tested for directional differences using paired t-tests (Bolker 2008). For purposes of comparison we considered migrant and mixed-migrant classification as equivalent (i.e. both strategies represent a type of migratory behavior).

Our results reveal novel differences in the effect of analytic approach on the classification and quantification of movement behavior. Classification by NSD, rNSD and elevation models differed for bighorn sheep and elk. NSD and rNSD classifications disagreed for 15% of bighorn and 8.7% of elk animal-years (95% confidence intervals 0.084–0.243 and 0.015–0.295, respectively; Table 2). Classification by NSD and elevation differed in 12.5% of animal-years for bighorn sheep (95% confidence interval: 0.067–0.218; Table 3) and 8.7% of elk animal-years (95% confidence interval: 0.015–0.295).

For animal years that were classified as 'migrant' by more than one method we found large differences in parameter estimates between NSD and rNSD and between NSD and

Table 3. Comparison of classification between NSD and elevation models. Results shown are for n = 88 Sierra Nevada bighorn sheep *Ovis canadensis sierrae*; A) and n = 23 elk *Cervus elaphus*; B). The cells we consider as representing consistent classification among methods are shown in bold. As with the comparison to rNSD, elk show fewer differences in classification (B, 2 of 23) than do bighorn (A, 11 of 88).

		Elevation			
A) Bighorn		disperser	migrant	resident	total
NSD	disperser	0	4	0	4
	migrant	**0**	**4**	0	4
	mixed migrant	**1**	**72**	0	73
	nomad	0	4	0	4
	resident	0	3	**0**	3
	total	1	87	0	88

		Elevation			
B) Elk		disperser	migrant	resident	total
NSD	disperser	1	1	0	2
	migrant	**0**	**9**	0	9
	mixed migrant	**0**	**12**	0	12
	nomad	0	0	0	0
	resident	0	0	**0**	0
	total	1	22	0	23

Table 4. Comparison of the differences between NSD-, rNSD- and elevation-based characterizations of migratory behavior. Here we include only animal-years that were classified by both methods as either 'migrant' or 'mixed-migrant'. Results are shown for n = 88 Sierra Nevada bighorn sheep *Ovis canadensis sierrae*; and n = 23 elk *Cervus elaphus*. The mean difference and it's standard deviation were calculated by subtracting the absolute value of either rNSD or elevation-based estimates from their NSD counterparts. Directional differences are shown as the mean of NSD estimates minus the mean of rNSD or elevation estimates, such that negative values indicate that the NSD model's estimate was smaller than that of the model to which it was being compared. Our analysis failed to find evidence of directional differences in estimates between NSD and rNSD models, with the single exception of the duration of returning movements in bighorn sheep (φ_2). We did, however, find directional differences between NSD- and elevation-based estimates for the timing of migration (θ), although support for this difference was weaker in bighorn sheep. We also found directional differences between elevation and NSD model estimates of the duration of migratory movements for bighorn sheep (φ and φ_2) and returning elk (φ_2). Finally, for bighorn we found a directional difference in the duration of migratory-range occupancy (ρ). All units are in days.

A)	NSD vs rNSD	Mean difference	SD	Directional difference
Bighorn sheep				
	Migratory-range occupancy (ρ)	20.5	42.26	3.02 ($p = 0.58$, DF = 73)
	Timing of migration (θ)	18.73	44.91	–1.29 ($p = 0.82$, DF = 73)
	Duration of departure (φ)	3.31	5.56	0.69 ($p = 0.36$, DF = 73)
	Duration of return (φ_2)	2.44	5.35	1.25 ($p = 0.06$, DF = 73)
Elk				
	Migratory-range occupancy (ρ)	8.73	24.97	–0.82 ($p = 0.89$, DF = 19)
	Timing of migration (θ)	14.98	37.94	1.15 ($p = 0.9$, DF = 19)
	Duration of departure (φ)	2.86	6.79	–0.72 ($p = 0.67$, DF = 19)
	Duration of return (φ_2)	1.42	3.97	–0.39 ($p = 0.68$, DF = 19)

B)	NDS vs Elevation	Mean difference	SD	Directional difference
Bighorn sheep				
	Migratory-range occupancy (ρ)	54.79	57.50	–12.61 ($p = 0.1635$, DF = 76)
	Timing of migration (θ)	55.05	70.92	–21.91 ($p = 0.0306$, DF = 76)
	Duration of departure (φ)	5.76	6.47	2.79 ($p = 0.0038$, DF = 76)
	Duration of return (φ_2)	7.97	7.16	–5.99 ($p < 0.0001$, DF = 76)
Elk				
	Migratory-range occupancy (ρ)	42.19	40.52	8.34 ($p = 0.5219$, DF = 20)
	Timing of migration (θ)	30.35	42.15	–26.93 ($p = 0.0118$, DF = 20)
	Duration of departure (φ)	6.99	8.19	0.64 ($p = 0.7896$, DF = 20)
	Duration of return (φ_2)	6.99	7.13	–5.46 ($p = 0.0075$, DF = 20)

elevation models (Table 4; Fig. 2; Fig. 3). We found large mean differences in the estimates from models fit to NSD and rNSD as well as between NSD and elevation models. NSD and elevation models also showed consistent directional differences, with NSD models estimating earlier dates of migration and shorter return movements (Table 4). Finally, elevation models for bighorn sheep also estimated shorter departing movements and longer migratory-range occupancy (by approximately three weeks) compared to NSD models (Table 4). Thus, even where these methods agreed in their classification, they differed in their characterization of migratory movement.

Discussion

By making NSD-, rNSD- and elevation-based methods more accessible through migrateR we hope to foster the further refinement of methods for the classification and quantification of migratory behavior and encourage further comparison across populations and taxa. The approach we propose highlights current ambiguity in the treatment of migratory behavior, but also provides promising avenues for further improvement. Increasing the correspondence between movement models and location data reduces need for ad-hoc corrections and thereby increases the rigor with which animal movement can be classified and quantified thus enhancing our ability to identify and understand the mechanisms that generate these behaviors.

Our examples illustrate migrateR's ability to allay three common problems with previous model-driven approaches of classifying and quantifying movement behavior. First, our re-parameterization of movement models to directly estimate the length of an individual's occupancy (ρ) on the migratory range provides a simple quantitative and biologically-meaningful basis for constraining definitions of migration, thus reducing the misclassification of resident behavior. Second, rNSD provides an efficient means of identifying and reducing NSD-based models' sensitivity to starting location. Third, our elevation-based models illustrate how alternative model families can complement existing NSD-based models, allowing characterization of migratory behavior that would remain invisible to NSD modeling alone. Through these improvements, migrateR provides a framework for repeatable workflow that will increase the success and transparency with which future studies classify movement behavior.

The case studies we provide comparing the application of these methods to data from elk and bighorn sheep also illustrate the variability in animal migration within and between species. These comparisons highlight variation in the classification and quantification of migratory behavior between different approaches that has not previously been directly acknowledged. For example, in the original work, Bunnefeld et al. claimed that migratory classifications in their dataset were insensitive to starting location, a conclusion our results clearly contradict. This sensitivity may be system- or species-specific, offering an explanation of why our results differ from those of previous authors primarily focused on moose (Bunnefeld et al., Singh et al. 2016). The shared framework migrateR provides encourages the development of model-driven classification methods by facilitating the further comparison of differences between approaches and species.

There will always be cases where appropriate classification of an animal's movement remains unclear. Although it may be tempting to adjudicate these conflicts with simulations (e.g. Bunnefeld et al.), simulating movement data presupposes precisely the underlying mechanisms we are seeking to identify and understand. Instead, investigating the differences among classification methods – facilitated by our approach – may provide valuable clues to the ecological drivers underlying this behavior and thus prove a more productive route (Cagnacci et al. 2016). To conserve migration we must first develop a fuller understanding of the causes and consequences of migratory behavior (Bolger et al. 2008).

MigrateR is free and open source, available under version 2 of the GNU general public license. All source code can be found on GitHub (<https://github.com/dbspitz/migrateR>). This R package will be useful to ecologists analyzing animal location data and is timely in providing tools tailored to the challenges attendant on the growing volume and taxonomic diversity of GPS datasets.

To cite migrateR or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for 'version 0':

Spitz, D., Hebblewhite, M. and Stephenson, T. 2016. 'MigrateR': extending model-driven methods for classifying and quantifying animal movement behavior. – Ecography 40: 788–799 (ver. 0).

Acknowledgements – We would like to especially thank Wibke Peters for her help in testing earlier versions of this package and the improvements she suggested. We also appreciate assistance from Evelyn H. Merrill, Nick DeCesare and Scott Eggeman with the elk data and helpful comments from Winsor Lowe, Michael Mitchell and L. Scott Mills.

References

Albon, S. and Langvatn, R. 1992. Plant phenology and the benefits of migration in a temperate ungulate. – Oikos 65: 502–513.

Bauer, S. and Hoye, B. J. 2014. Migratory animals couple biodiversity and ecosystem functioning worldwide. – Science 344: 1242552.

Beamish, F. W. H. 1966. Vertical migration by demersal fish in the northwest Atlantic. – J. Fish. Res. Board Can. 23: 109–139.

Beatty, W. S. et al. 2015. An empirical evaluation of landscape energetic models: mallard and American black duck space use during the non-breeding period. – J. Wildl. Manage. 79: 1141–1151.

Blake, S. et al. 2013. Vegetation dynamics drive segregation by body size in Galapagos tortoises migrating across altitudinal gradients. – J. Anim. Ecol. 82: 310–321.

Bolger, D. T. et al. 2008. The need for integrative approaches to understand and conserve migratory ungulates. – Ecol. Lett. 11: 63–77.

Bolker, B. M. 2008. Ecological models and data in R. – Princeton Univ. Press.

Börger, L. and Fryxell, J. M. 2012. Quantifying individual differences in dispersal using the net squared displacement statistics. – In: Colbert, J. et al. (eds), Dispersal ecology and evolution. Oxford Univ. Press, pp. 222–230.

Boyle, W. A. et al. 2010. Storms drive altitudinal migration in a tropical bird. – Proc. R. Soc. B 277: 2511–2519.

Bunnefeld, N. et al. 2011. A model-driven approach to quantify migration patterns: individual, regional and yearly differences. – J. Anim. Ecol. 80: 466–476.

Burnham, K. P. and Anderson, D. R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. – Springer Science and Business Media.

Cagnacci, F. et al. 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. – Oikos 120: 1790–1802.

Cagnacci, F. et al. 2016. How many routes lead to migration? Comparison of methods to assess and characterize migratory movements. – J. Anim. Ecol. 85: 54–68.

Calenge, C. 2006. The package "adehabitat" for the R software: a tool for the analysis of space and habitat use by animals. – Ecol. Model. 197: 516–519.

Dingle, H. 2006. Animal migration: is there a common migratory syndrome? – J. Ornithol. 147: 212–220.

Dingle, H. and Drake, V. A. 2007. What is migration? – BioScience 57: 113–121.

Eggeman, S. et al. 2016. Behavioral flexibility in migratory behavior in a long-lived ungulate. – J. Anim. Ecol. 85: 785–797.

Gutierrez, D. and Wilson, R. J. 2014. Climate conditions and resource availability drive return elevational migrations in a single-brooded insect. – Oecologia 175: 861–873.

Hebblewhite, M. and Merrill, E. H. 2016. Data from: Behavioral flexibility in migratory behavior in a long-lived large herbivore. – Movebank Data Repository, <doi:10.5441/001/1.k8s2g5v7>.

James, M. C. et al. 2006. Changes in diel diving patterns accompany shifts between northern foraging and southward migration in leatherback turtles. – Can. J. Zool. 84: 754–765.

Kays, R. et al. 2015. Terrestrial animal tracking as an eye on life and planet. – Science 348, doi: 10.1126/science.aaa2478

Kranstauber, B. et al. 2011. The Movebank data model for animal tracking. – Environ. Model. Softw. 26: 834–835.

Lampert, W. 1989. The adaptive significance of diel vertical migration of zooplankton. – Funct. Ecol. 3: 21–27.

McGuire, L. P. and Boyle, W. A. 2013. Altitudinal migration in bats: evidence, patterns, and drivers. – Biol. Rev. 88: 767–786.

Milner-Gulland, E. J. et al. 2011. Animal migration: a synthesis. – Oxford Univ. Press.

Mysterud, A. et al. 2011. Partial migration in expanding red deer populations at northern latitudes – a role for density dependence? – Oikos 120: 1817–1825.

Naidoo, R. et al. 2012. Home on the range: factors explaining partial migration of african buffalo in a tropical environment. – PLoS One 7: e36527.

Nielsen, J. K. et al. 2014. Characterizing Pacific halibut movement and habitat in a Marine Protected Area using net squared displacement analysis methods. – Mar. Ecol. Prog. Ser. 517: 229–250.

Sims, D. W. et al. 2006. Hunt warm, rest cool: bioenergetic strategy underlying diel vertical migration of a benthic shark. – J. Anim. Ecol. 75: 176–190.

Singh, N. J. et al. 2016. Quantifying migration behaviour using net squared displacement approach: clarifications and caveats. – PLoS One 11: e0149594.

Spitz, D. B. 2015. Does migration matter? Causes and consequences of migratory behavior in Sierra Nevada bighorn sheep. – PhD thesis, Univ. of Montana.

Turchin, P. 1998. Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants. – Sinauer.

Van Moorter, B. et al. 2016. Movement is the glue connecting home ranges and habitat selection. – J. Anim. Ecol. 85: 21–31.

Wiktander, U. et al. 2001. Seasonal variation in home-range size, and habitat area requirement of the lesser spotted woodpecker (*Dendrocopos minor*) in southern Sweden. – Biol. Conserv. 100: 387–395.

Wilcove, D. S. and Wikelski, M. 2008. Going, going, gone: is animal migration disappearing? – PLoS Biol. 6: e188.

Supplementary material (Appendix ECOG-02587 at <www.ecography.org/appendix/ecog-02587>). Appendix 1.