Appendix S4 – iSSA practical user guide

Here we provide tips and guidelines for conducting a fruitful integrated step selection analysis (iSSA). We refer the reader to Thurfjell *et al.* (2014) for a more general review of applications of SSA.

1. Collect animal positional data
    - To maximize the usefulness of the data in an iSSA, positional data should be collected at a constant fix rate (equal time steps).
    - Generally speaking, high fix rate (short time steps and hence short step-lengths) is expected to increase the reliability of the analysis. This is particularly true when aiming to capture continuous use of small spatial units, such as roads. Note however that fix rate should be adjusted to the typical displacement rate of the study species. Fix rate should be considered too high (and hence wasteful) if during a single step the animal is expected to travel less than the positional error (~ 20-30 m for GPS tags), or the spatial resolution of the focal habitat map (~ 20-250 m for most satellite derived maps).
    - Both habitat selection and movement behavior may depend on time of day and season. Sampling should attempt to capture as much temporal variability as possible during the time of the study. For example, fix schedule that is out of synch with time of day (e.g., every 5 hours) can help capture more (temporal variability) with less (fixes).
2. Tabulate observed (case) steps
    - Clean the data - even good GPS datasets contain erroneous positions. Exclude fixes taken before and shortly after tag deployment and after mortality/drop-off events. Plot the observed trajectories and visually look for potential positional errors. Run a code/script that scans the data for extreme values such as unreasonably long (or fast) steps or return-trips (relocations starting and ending at approximately the same point).
    - Because fixes are not always taken at their designated time, it is useful to define a reasonable temporal tolerance range for the step duration (e.g., 1 hr ± 10 min).
    - We recommend including the following fields in the 'used-step' table: individual ID, unique step ID, step start-point (time, easting and northing), step endpoint (time, easting and northing), step-duration, step-length, and step-heading (relative to the true north).
    - If velocity autocorrelation is to be included in the analysis, one must make sure to only calculate velocity deviations (e.g., turn-angles or step-length differences) between successive valid steps. Any step that does not have a valid step leading to it (e.g., because the previous fix is missing), cannot be characterized by valid velocity deviations and hence must be

excluded from the analysis (note that such a step should still be used to calculate velocity deviations for the proceeding step).

3. Sample lengths of available (control) steps
    - Available step lengths should be sampled based on one of the following probability density distributions (see Appendix S2): a uniform distribution within some maximal distance (e.g., the longest observed step), exponential, normal, gamma, and log-normal.
    - If there is no a priori reason to use a particular distribution, we would recommend using the gamma because it is flexible and includes the exponential as a special case.
    - Whatever the theoretical distribution of choice is, the observed step-lengths should be used to estimate its parameters (using either the method of moments or maximum likelihood). If there is strong a priori biological reason to think that these estimates should differ between distinct portions of the data (individual ID, sex, season, study area, etc.), more efficient model fitting can be gained by partitioning the data accordingly.
    - If the user is interested in determining which theoretical distribution best fits the data, we recommend using the uniform distribution to sample available step-lengths. The step-length, its natural logarithm, its square, and the square of its natural logarithm should then be included as predictors in a set of four competing models (see Appendix S2 for details), and AIC can be used to choose the best one.
4. Sample available (control) step headings
    - In the simplest case, available step headings are sampled from a uniform distribution between 0 and 2π.
    - If directional correlations or bias are evident in the data, increased efficiency may be gained by sampling available step headings from a von Mises distribution where the concentration parameter is estimated from the observed directional persistence/bias distribution.
    - If observed directional persistence/bias is correlated with step-length (e.g., the animal tends to turn less when making short steps), increased efficiency can be gained by accounting for the correlation structure when sampling available step headings.
5. Generate available (control) steps
    - Combine sampled step lengths and headings (with appropriate cross-correlation structure) to generate available steps starting at a used step start-point and ending in random endpoints.

- Identify each cluster, consisting of a single used step and its matched set of available steps, with a unique cluster ID. Code all used steps as '1' (case) and all available steps as '0' (control).

6. Attach step attributes
    - Characterize each step (cases and controls) with the following: step-length, step turn-angle and angular deviation from a preferred direction (if relevant), temporal covariates (e.g., time of day, season), and spatial covariates (e.g., elevation, NDVI, cover, temperature, etc.).
    - Covariates could be matched (in space and/or time) to the step's start-point, to its endpoint, and/or based on some interpolation between the two (e.g., average along the step, within an ellipse bounded by the start and end positions, or along a Brownian bridge). Note that, if covariates are purely temporal or are measured at the step start-point, their value would be identical for all steps belonging to the same cluster and their independent effects are thus statistically unidentifiable. The effects of such covariates are identifiable when interacting with other variables (e.g., an interaction between step-length and season).
7. Fit a conditional logistic regression
    - As long as sample sizes are sufficient (see Appendix S6), we recommend fitting iSSA for each individual independently (rather than using a mixed effects approach). This allows for a straightforward and unbiased evaluation of both inter- and intra-individual variability (Fieberg *et al*. 2010). Population level inference can then be gained by averaging individual model fits.
    - Function *clogit* in R is often used to fit conditional logistic regression. Note that this function (as many other conditional logistic regressions routines) rely on a Cox proportional hazard model to obtain MLEs and hence its output is a *coxph* output.
8. Adjust movement coefficients
    - Once step-length and/or turn-angle coefficient estimates are obtained, those must be combined with the tentative parameter estimates used for sampling available steps (see Appendix 1 and the main text for details). If steps were sampled from a uniform distribution, no adjustments are needed.
9. Simulate space-use
    - Once the integrated step selection function has been fully parametrized, it can be used to simulate space-use across any discrete map of its spatial covariates.

- This requires a simulation model that iteratively calculates the redistribution kernel at each simulated position and then samples from this kernel to select the next position. Note that the parametrized step-length distribution is one-dimensional, it describes the probability density of any particular displacement over the prescribed time step. When calculating the full two-dimensional redistribution kernel, rather than just drawing from such kernel, the basal probability density for any distance, r, from the center of the kernel is proportional to $2\pi r$. One must thus correct for dimensionality by dividing by $2\pi r$. This requires care when the step-length approaches 0, possibly by setting a minimal value for r (e.g., $1/2\pi$), so as to prevent the kernel from collapsing onto a Dirac delta function.
- A Monte Carlo approximation of the utilization distribution can be gained by simulating long and/or multiple trajectories (starting from the same or different positions) and then normalizing space-use across the map (see also Appendix S5).
- Such simulations may be used for cross validation (e.g., using ROC AUC to quantify the predictive power of the model over a validation dataset), as well as for predicting the ecological consequences of habitat loss, fragmentation, and other environmental changes that might affect animal space-use.