CTMM Workshop

* Task: How do we account for
  + Temporal autocorr: Stochastic process model
  + Account for sampling irregularity & mismatch: continuous time models
  + Account for location error, if necessary: calibrated error model
  + Don’t assume models: model selection and testing assumptions
  + Don’t be satisfied with point estimates: estimate uncertainties
  + Propagate uncertainties into population estimates: hierarchical models
* Example : Black Bear movement sampling
  + When ignoring autocorrelation, we end up underestimating things like home range.
    - Autocorrelation is primarily caused by the time between sampling points, we are ignoring some of the landscape autocorrelation that would cause an individual to say in a location
  + KDE is statistically optimal non-parametric method of estimating distributions of independently sampled data
  + Conventional estimates (KDE) can be 2-20 times too small (Noonan et al., 2019)
  + Bias is worse when fewer range crossing are observed
* What do we do to relieve autocorrelation?
  + Thin the data based on time, useful for home range estimation, but less useful for anything else
* What if we don’t? Autocorrelated data can be useful information! It can be used to calculate speeds
* Continuous time bs discrete time
  + Conventional movement analysis may have step link distribution + angle distribution. This is implicitly discreet because things only happen at the measured points
  + Continuous time tries to also estimate what happens during the time between the measured/sampled points. This is analyzed and ends up in a ribbon strand of possible paths between points.
  + Benefits
    - Sampling may be irregular in time
    - Sampling schedules may differ across individuals. Model parameters are biological, not sampling dependent
    - Discree time models can be intrinsically scale dependent. Step selection functions tell you information about selection across the time scales at which it is analyzed.
    - Can accommodate speed, distance, acceleration, power and energy (first derivatives, calculus 1)
    - Location error is easy to model
* Building block continuous-time stochastics process models (ordered in historical commonality
  + Independent locations: KDE, MCP, RSF
  + Brownian Bridge + motion: locations are correlated from one space to another. Assume “infinite speed.” And does not have a home range and can diffuse forever. Cannot estimate home ranges. Can be used with small data because it utilizes less parameters than an integrated OU or OUF.
  + Ornstein-Uhlenbeck (OU) motion: Same as Brownian, but has a home range.
  + Integrated Ornstein-Uhlenbeck (OU) motion: Introduced in the marine context. Diffuses forever and does not have a home range. However, has a persistence of motion rather than jittering. Can estimate speeds, distances, and velocities. (crawl package)
  + Orenstein-Uhlenbeck-Fleming (OUF) motion: Ornstein-Ihlenbeck-Fleming model. Persistence of motion parameters and can estimate speed and distance, has a home range.
  + HOWEVER! Don’t assume you have one of the models, test your assumptions and select. Fit all the models and different variations, select, then estimate speed and home range.
* CTMM Package
  + Has vignettes for long-form examples
  + Has an FAQ
  + Google group to ask questions, search for questions
  + Github is for bugs
  + Recorded workshops and some lectures are in the google group, 2022 up, should have 2023 and 2024 at some point
  + Each method should include references in the help file, all these papers are in a dropbox
  + CRAN will always be a bit out of date, but will be stable
  + Changelog is found in the NEWS file.
  + Moveapps is a neat, no-code option and plays nice with Movebank
  + Older shiny package works, has functionality, but moveapps will be the primary tool going forward.
* Functions of the Package
  + Import Data: Can read from Movebank directly. Use the “as.telemetry” function.
    - Reads in as a list of individuals. Each individual data is within the telemetry class
    - Doesn’t use POSIX time stamps because it is SLOW. Does it retain the timezone?
    - Can estimate a relatively safe projection (equidistant). This is easier to calulate and estimate than an equal area
  + Plotting
    - Some good functions to vuisually separate individuals, show movement over time
    - Can color individuals movement over time to check for home range vs migration
  + Variogram
    - Can get an unbiased view in the autocorrelation
    - Acf function is the correlogram. Checks for autocorrelation in the residuals.
    - Some arguments for the variogram to allow for changes in sampling rates for individuals.
    - Asymptote gives us information about home range size and also about the amount of time required to cover the home range.
    - Quadratic function at the very beginning of the variogram can indicate that the data is sampled finely enough to estimate speed.
    - Variogram only gives us the first half. As you get farther along the variogram, it becomes autocorrelated again since you’re splitting the data in half and then it degenerates to noise as there are less and less data.
  + Model Selection
    - Ctmm.guess gives us some sliders to play with
      * Tau position day gives us information about the time required between days for an individual to cross it’s home range. This can differ from the asymptote because it can be put into an exponential to examine the amount of autocorrelation that is addressed/removed across different periods of time.
      * Tau velocity is compared to the quadratic at the very beginning of the variogram to examine if we can estimate speed.
      * If you’re not learning, run with interactive = F
    - Ctmm.select
      * Fits multiple stochastic process models.
      * Verbose = T keeps all models, otherwise it only gives the best fitting
      * Cores argument only good for MacOS and Linux
    - Can show the variograms between the data and the model. Check for gaps between the modeled and actual variograms for a visual estimation of model fit
* Home Range and Kernel Density Estimation
  + Dt.plot shows all the sampling intervals in the data. Differences in time intervals can be due to collar error or intentional changes. Usually due to error
  + Next step is to to run ctmm.guess and then ctmm.select
  + Conventional KDE (“kde”) can be run from an IID model (either anisotropic or isotropic)
  + Autocorrelated KDE (“akde”) Akde with weights = T can allow the differences in sampling interval to be accommodated
  + Confidence interval for AKDE can be really tight, overly so
  + KDE estimator can also plot the bin size/the resolution at which the density estimator is calculated
  + Oversmoothing Bias correction. Oversmoothing is when too large of a bandwidth is selected.
  + Average Home range of multiple individuals as derived from AKDEs.
    - To propagate the error within the estimates themselves, there is a hierarchical model to combine them.
    - Can use the “Log” function to get approximately normally distributed values for home ranges. These can then be used with the “metaphor” package for metanalyses
* Estimating population range
  + Used to calculate the critical patch size and minimum area requirements for a given population.
  + How do we scale up from individual home ranges to a minimum population range?
  + Traditional Population Range Estimations
    - MCP around all individuals, assumes IID
    - KDE: Assumes IID
  + Biases of traditional estimators
    - Modern tracking assume IID
  + Potential fix to scale up in CTMM, use the akde and then mean these density estimates. However, this ignores any individual variation
  + ACTUAL Fix: uses a hierarchical model to address individual variation
    - Gives weights to individuals based on their sampling amount/density. Makes sure that individuals who are oversampled are not given more weight on the population KDE estimate
* Effective Sample Sizes
  + Uses REML or moderately small sample sizes (4-5)
  + Uses a parametric bootstrap for very small sample sizes (2-3)
* Location error
  + Need to compare the error around a location estimate and the potential movement distance across
  + Do we have calibration data to convert a relative error like HDOP values into meters?
    - Relative standard deviation of the errors at that point in time for all the satellites that you can see.
  + Statistical identifiability issues arise when we have no distinction between location error and movement.
  + Error model parameters derived from a stationary tag are used as a prior on the movement model to separate location error from the variation introduced by an individual moving
  + When looking at the error fitting, we can use the outputs along with uere()<- function to calculate actual error from the HDOP values.
  + There may be instances where a blanket error that is applied to all locations may be more valuable to use than utilizing HDOP values
  + Many of the functions within ctmm can incorporate the error model that comes from the pipeline that converts DOP values to actual error
    - Fishers Z test can incorporate the error model
  + Outlier calculations (“outlie”) can be error model informed, i.e. can examine the point in relation to the geometric median. Based on the error and the distance from the geometric median, a calculation to get the minimum speed required to reach the point is given. Based on the movement capabilities and possible situation , you can make informed decisions
    - Outliers can also include speed data derived from GPS data.
* Afternoon Section: Derivative analyses
  + Movement model is the most important piece since many additional features are dependent upon it (e.g. behavior states, interpolation and forecasting, path length, velocity, home range, etc)
  + When location bias is associated with certain habitat features, models will end up biasing for all of those habitat types
  + All derivative analyses assume the movement model is correct
  + Conditional Simulation?
  + Power analyses and study design
    - **MoveDesign. A power analysis web application for movement ecology**
    - We can run tests to see how long it takes to cross a home range and use it to inform the sampling interval and sampling length possible/needed to address the question of interest. There are always tradeoffs between frequency and duration.
  + Brownian bridge gives the benefit of probability of location during a specific time. Not part of CTMM, but interesting
  + Occurrence Distributions
    - Excludes natal dispersal
    - Excludes occastional sallies outside the area, perhaps exploratory in nature
    - Can shifty, grow, and shrink (non-stationary)
    - What KDE estimates
    - Generalized time series kriging: Rather than thinking about kriging as a spatial process, we see it as a time process
    - Occurrence distribution using the interpolation is dependent on sampling frequency, it estimates where about the individual could have gone between points
  + When to use occurrence distributions?
    - Where did an animal cross a linear feature
    - How likely is it that an animal visited a location of interest
    - When and where did two animals interact
    - Which areas of a landscape contain high priority areas for conservation (for that individual during the period of analysis)? (Also addressed by RSFPs,. RSFPs would be better for populations while the occurrence distribution would be better for figuring out probability of linear feature corssings.)
* Estimation of Speed and Distance Travelled
  + More interpolation of what the animal did
  + Among the most routinely estimated metrics form tracking distance.
  + Traditionally estimated by examining straight line distances,
    - Biased too short
    - Known for decades that this is not great.
    - Solution is to get shorter shorter intervals. Only really works when there is no location error
  + Continuous time speed and distance estimation
    - **Assumes stationary movement model, the individual is always moving, if the animal changes states (foraging/moving/migrating/roosting/etc) this is an assumption violation. If this is a concern. Then the data should be segmented behaviorally**
    - Needs to have velocity autocorrelation. Requires Integrate Ornstein-Uhlenbeck or OUF.
    - Need to simulate, estimate, and ensemble the point estimates in order to get a distribution of estimates on distance traveled and speed
    - Good expect when you are sampling really coarsely (<50 locations per day)
  + Tau Velocity gives you an estimate for about how long an individual stays in similar velocities
  + It is possible to examine how speed changes in different habitat types based on the habitats that the measure is taken in.
  + Function: speeds (with and s!)
    - Creates a dataframe with the timestamp
    - Has the estimated instantaneous movement speed.
    - Can annotate the graphs with the velocities
  + Requires a correlated velocity model.
    - Tauv can be plotted against sampling frequency (fixes per day) to get some sort of model selection? I’m not totally sure what’s going on here
* Interactions: how do pairs of animals overlap?
  + Home range overlap
    - Captures the potential of interactions to occur
    - Need accurate home range
    - Need measure of overlap that is unbiased and easily interpretable. Use Bhattacharyya Distance. Measure of the overlap between two probability distributions.
    - Need a way to propagate uncertainty
  + Encounter Locations Distributions
    - Can calculate the conditional distribution of encoutners, asks the question: where do these encounters happen?
    - Result in a distribution of encounter locations
    - Could be useful for modeling potential disease transmission
    - Says nothing about whether they are in the same space at the same time
    - Makes an assumption of independent movement
  + Pairwise distance
    - Pairwise separation distances tells us attraction/avoidance
    - Shows that some of the buffalo have a fission/fusion dynamic.
    - Most of the time, it will look very messy. Mating seasons may be able to be determined when plotting pairwise distances across time
    - TO test….
      * Simulate two movement tracks
      * Plot and examine
    - Other function: “proximity” which gives unitless ratio. Value less than 1 is closer than expected by random chance. Value of 1 is random. Value greater than 1 is further apart than by random chance.
  + Encounter rates
    - Currently does not include a perceptual range
    - Can define what distance defines an encounter
    - Highly recommend a sensitivity analysis where you examine the number of encounters across ranges of the encounter radius
* Resource Selection Functions
  + “I’m not going to teach you RSFs. I’m going to assume you have been doing [them] wrong and will teach you the right way.”
  + KDE is a non-parametric while the RSF is a parametric
  + Basic RSF assumes independent data, so the RSF in ctmm makes a post-hoc adjustment and utilizes the appropriate effective sample size.
  + RSF in ctmm also includes the weights on the points for disproportionatly sampled areas due to changes in sampling regime or collar failure
  + Keeps sampling until you hit a 1% error rate
  + Normal RSF compares actual points to “available” area. However, this is usually determined by another method and can therefore be bias
  + Ctmm RSF is an integrated model since it estimates the RSF variables in tandem with availability (not sure how this works/does it)
  + Check for convergence for you
    - Numerical Error of the loglik
    - Checks the loglik change in number of points
  + Can use a Riemann integrator rather than a Monte Carlo one to avoid the use of “available” and availability points.
    - Requires that the predictors are lined up
  + Either do this or do integrated step selection models if we are going to talk about habitat selection/suitability from movement data
    - Integrated step selection means you need to thin to the first order of autocorrelation
    - Step selection is for a function of what will be selected I the next step, given the current state
  + Utilization distribution keeps coming up, need to look it up.
* Other notes
  + When doing validation for movement data, split the data temporally (first half/second half)
  + As data get coarser, (time wise) while neglecting autocorrelation we end up estimating shorter movement paths between time periods than actual, and therefore underestimate individual speed
  + The functions are set up in a way where usually, 10x more data means 10x longer processing.
  + Isotropic vs anisotropic. Isotropic results in a circular gaussian estimate while anisotropic can result in a stretched gaussian as the correlation matrix can be elongated.
  + MonteCarlo vs Riemann integration. Monte Carlo is randome sample while Riemann is summing. Riemann only works for temporal stationarity. Does it have any problems when utilizing coefficients that are dependent upon one another? (Proportion within the cell that is ag vs proportion in cell that is forest, vs proportion of x,y,z)
  + Incoming, untested feature non-stationary landscapes that allow RSF to be fit across different time periods.

Noonan, M. J., Tucker, M. A., Fleming, C. H., Akre, T. S., Alberts, S. C., Ali, A. H., Altmann, J., Antunes, P. C., Belant, J. L., Beyer, D., Blaum, N., Böhning-Gaese, K., Cullen Jr., L., de Paula, R. C., Dekker, J., Drescher-Lehman, J., Farwig, N., Fichtel, C., Fischer, C., … Calabrese, J. M. (2019). A comprehensive analysis of autocorrelation and bias in home range estimation. *Ecological Monographs*, *89*(2), e01344. https://doi.org/10.1002/ecm.1344