**README: Wildlife Activity Pattern Analysis using Camera Trap Data in R**

**By: Samuel Xin Tham Lee (23rd January 2024)**

Background and Aims:

Quantifying wildlife activity or how species distribute their activity the 24-hour cycle is an important question in ecology. Understanding when animals are active is pivotal in providing deeper insights into how species interact with one another and therefore how entire communities are assembled. With more than 70% of world’s surface having measurable levels of anthropogenic disturbance1, it is crucial for ecologists to elucidate how wildlife may temporally respond to such disturbances and the magnitude of these responses. Here, I have created multiple code scripts which serve to examine the change in wildlife activity in response to anthropogenic disturbance using Forest Landscape Integrity Index (FLII) as a disturbance proxy2. I analyzed the change in activity pattern of Southeast Asian wildlife community at multiple levels (*i.e.*, community, guilds, and species) and used two different analytical approaches, namely Kernel Density Estimation and Multinomial Logit Mixed Modelling3, 4.

Summary of the pipeline

Here, I have developed a general pipeline to ensure that each analysis is ran smoothly and visualizations are at publications standards. The pipeline has no steps since most of these analyses are independent from one another, but as always, a data cleaning, wrangling and standardizing process is mandatory before engaging any sort of analysis. All my code is implemented in the R statistical software version 4.3.1 using either regular R scripts or R markdown (Rmd) files to ensure reproducibility and a tidier workflow. I have also inserted detailed descriptions and my own thoughts for each code chunk which hopes to assist users in answering some of their doubts and questions. The code scripts are currently hosted at Github (link) and at Zenodo (link). All datasets are made available at Figshare ([link](https://doi.org/10.6084/m9.figshare.23513412)). Below is a step-by-step guide of the data cleaning and wrangling section as well as the respective analyses conducted and their associated packages.

1. **Data cleaning, wrangling and standardizing**
   * The main goal of this entire process is producing clean captures files and integrating important covariate information that will be used for all our analyses.
   * The process includes the removal of certain taxon from the captures. I removed taxon that do not possess species-level identifications, domesticated taxon (*e.g.,* dogs and cattle etc) that can influence our results, taxon that possess body mass <1 kg due to high inaccuracies in identifying small-sized individuals, and humans.
   * Furthermore, the code chunks also serve to integrate and standardize species trait information of each species in our captures data frame. The species trait that are of interest includes body mass (in kg) and feeding guild (*i.e.*, carnivore, herbivore, and omnivore). Both of this trait information are combined to form the “trophic\_guild” columns which basically describes the trophic guild of a species.
   * In addition, the code chunks also function to integrate our disturbance covariate, namely the Forest Landscape Integrity Index to be part of our captures data frame. The point values of FLII at each of our camera locations are extracted previously using the ArcGIS Software where values are calculated in a geographically consistent manner and are standardized globally. Values ranges from 0 (most disturbed) to 10 (most intact). The disturbance status of each camera site is then defined based on the median FLII or the first and third quartiles of FLII (sensitivity analysis).
   * Moreover, some of the code chunks are also aimed to standardized the time-stamped of each detection to account for daylength variation by using the suntime() function.
   * Lastly, the remaining code chunks are for the generation of our datasets that will be used for our analyses and supplementary tables for publication purposes. The datasets are named based on the level being analyzed (*i.e.*, community, guild, and species) and how disturbance/intact is defined. The datasets also begin with the tag “SEA\_Activity\_” and ends with the date of its production (*e.g.*, “\_20240123”). For species-level kernel density estimation, the datasets only included species with ≥20 detections in both intact and disturbed forests to prevent zero errors.

* The code script for this entire process is saved as a .Rmd file with detailed descriptions for each code chunk. Packages that were used includes “tidyverse (version 2.0.0)”, “sp (version 2.1-2)”, “activity (version 1.3.3)”, “overlap (version 0.3.4)”, and “plyr (version 1.8.8)”.

1. **Kernel Density Estimation** 
   * The purpose of this code script is for the generation of circular kernel density curves (hereafter, “activity curves”) of each community, guild, or species in both intact and disturbed forests.
   * The activity curves are generated using the fitact() function found in the “activity” package which describes the activity of a community, guild, or species throughout the 24-hour cycle.
   * To test whether shifts in activity between intact and disturbed forests are biologically meaningful, I used several criteria, namely p-value (calculated from the compareCkern() test in the “activity” package), peak activity (determined by extracting the highest kernel density value in both forests), and the overlap coefficient (∆) or the shared area between two activity curves (range from 0 to 1) which is estimated using the overlapEST() function in the “overlap” package.
   * All activity curves are visualized using the ggplot() function found within the “tidyverse” package.
   * **Notes for users**: For testing purposes, please specify the number of boostrapping iterations (“reps”) for the fitact(), compareCkern() and overlapEST() functions to a lower number as this can take some time to finish. Also, kernel density estimation fall short in accounting for nested sampling design (random effects) and weighing species equally especially in a guild- and community-level analysis. Therefore, I took a step further by using multinomial logit mixed modelling approach to counteract this (described below).
   * The code scripts for these analyses are saved either as a .Rmd file or as a regular R file with detailed descriptions for each code chunk. Packages that were used includes “tidyverse”, “activity”, “overlap”, “lubridate (version 1.9.2)”and “plyr”. The name of the code scripts is also specified to include how disturbance status is defined (median or quartile) and the level being analyzed.
2. **Multinomial Logit Mixed Modelling**

* To assess the change in wildlife diel activity over a gradient of disturbance, I built multinomial logit mixed models (MNLMMs) with three response variables, namely “day”, “night”, and “twilight” and used FLII as our primary fixed effect. MNLMMs overcome the problems associated with kernel density estimation as mentioned previously.
* Before building the models, I first defined each of the diel category, I defined detections as “day” when detections are found between 0730 – 1630 hrs, “night” when detections are found between 1930 – 0430 hrs, and “twilight” when detections are found between 0430 – 0730 hrs or 1630 – 1930 hrs.
* I fitted the MNLMMs using the mblogit() function found in the “mclogit” package version 0.9.6. To account for nested sampling design, I included the landscapes surveyed as a random effect for all our models. To test whether species traits mediate the change in diel activity for community-level models, I also included body size (*i.e.*, small, medium, and large), body mass (in kg), and feeding guild (*i.e.*, carnivore, herbivore, and omnivore) as fixed effects. To assess whether species turnover contributes to changes in diel activity for guild- and community-level models, I built models with or without species as random effects, where the former weighs each species equally while the latter weighs each detection equally.
* I calculated the Akaike Information Criterion (AIC) score of each model and selected the best model based on the model with lowest AIC score.
* I then calculated the predicted probabilities of the best model using the predict() function and visualized it using the ggplot () function.
* The code scripts for these analyses are saved as a .Rmd file with detailed descriptions for each code chunk. Packages that were used includes “tidyverse” and “mclogit”. The name of the code scripts is also specified to include the level being analyzed

1. **Species Pairwise Overlap Analysis**

* This analysis is quite similar to the kernel density estimation, however, this part of the analysis focuses more on the overlap coefficient (∆) between species pairs.
* To ensure we have enough sample sizes, only species with ≥20 detections in both disturbed and intact forests are selected.
* The activity curves in both intact and disturbed forests for each species are generated using the fitact() function and are paired with other species within the same forest types.
* Species pairs are categorized into either “competition” or “predation”. Species pairs that are highly unlikely to interact are then removed.
* The ∆ of each species pairs are collated and calculated within each forest type and a paired t-test or a Mann-Whitney test is used to determine whether mean overlap changes between forest types are significant. The results are then visualized using ggplot() function.
* The code script for this analysis is saved as a .Rmd file with detailed descriptions for each code chunk. Packages that were used includes “tidyverse”, “activity”, “lubridate”, “overlap”, “data.table (version 1.14.8)”, and “plotrix (version 3.8-2)”. The datasets used for this process are all located [here](https://www.dropbox.com/scl/fo/0ffof3avffs7d2uiytcge/h?rlkey=znfwguai8ptk6eup3mg8j7epe&dl=0) on Dropbox.

1. **Change in Percentage Day, Night and Twilight Detections**

* Essentially, these lines of code serve to calculate the percentage change in day, night, and twilight detections from intact to disturbed forests.
* The formula is as follows: [(day/night/twilight detections in disturbed habitats / total detections in disturbed habitats) - (day/night/twilight detections in intact habitats / total detections in intact habitats)] \*100.
* You can also conduct this quite easily in Excel if you are press for time.

Conclusions

To my knowledge, most of the approaches listed here are currently being used extensively in the field of ecology and camera trap analysis. Recommendations to improve current analytical methods and collaborations on new projects using similar methods above are always welcomed. For more information regarding the code itself, please contact me at [samleext@gmail.com](mailto:samleext@gmail.com).

References

1. Venter, O.*, et al.* Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nat Commun* **7**, 12558 (2016).

2. Grantham, H.S.*, et al.* Anthropogenic modification of forests means only 40% of remaining forests have high ecosystem integrity. *Nat Commun* **11**, 5978 (2020).

3. Melff, M. *mclogit: Multinomial Logit Models, with or without Random Effects or Overdispersion*. R package version 0.9.6. (2022).

4. Rowcliffe, J.M. *activity: Animal Activity Statistics*. R package version 1.3.2. (2016).