Estimation of latent behavioral states and relationships with environmental covariates in giant armadillos (*Priodontes maximus*)

**Background**

While there has been an increased effort to understand the movement and spatial ecology of giant armadillos within the Brazilian Pantanal, quantitative measures of latent behavioral states and their relationships with environmental covariates have not yet been evaluated. The estimation of latent behavioral states from animal telemetry data is useful to elucidate what behaviors individuals are performing over time and how they correspond with different environmental features over space and time. Latent behavioral states are often estimated by evaluating the measures of step lengths and turning angles of a set of tracks using hidden Markov models (HMMs), although there are some limitations to this approach. A recent method developed by Valle et al. (in review) uses a non-parametric Bayesian mixture model with a penalized prior to estimate the optimal number of behavioral states, the associated state-dependent distributions for each state, and the state estimates for each observation of the tracks. These can be evaluated using any set of data streams recorded by the tag or metrics derived from the track itself (e.g., step lengths, turning angles, accelerometer data). This mixture model was used to estimate the latent behavioral states from giant armadillos to assess how this species interacts with its environment in the Brazilian Pantanal.

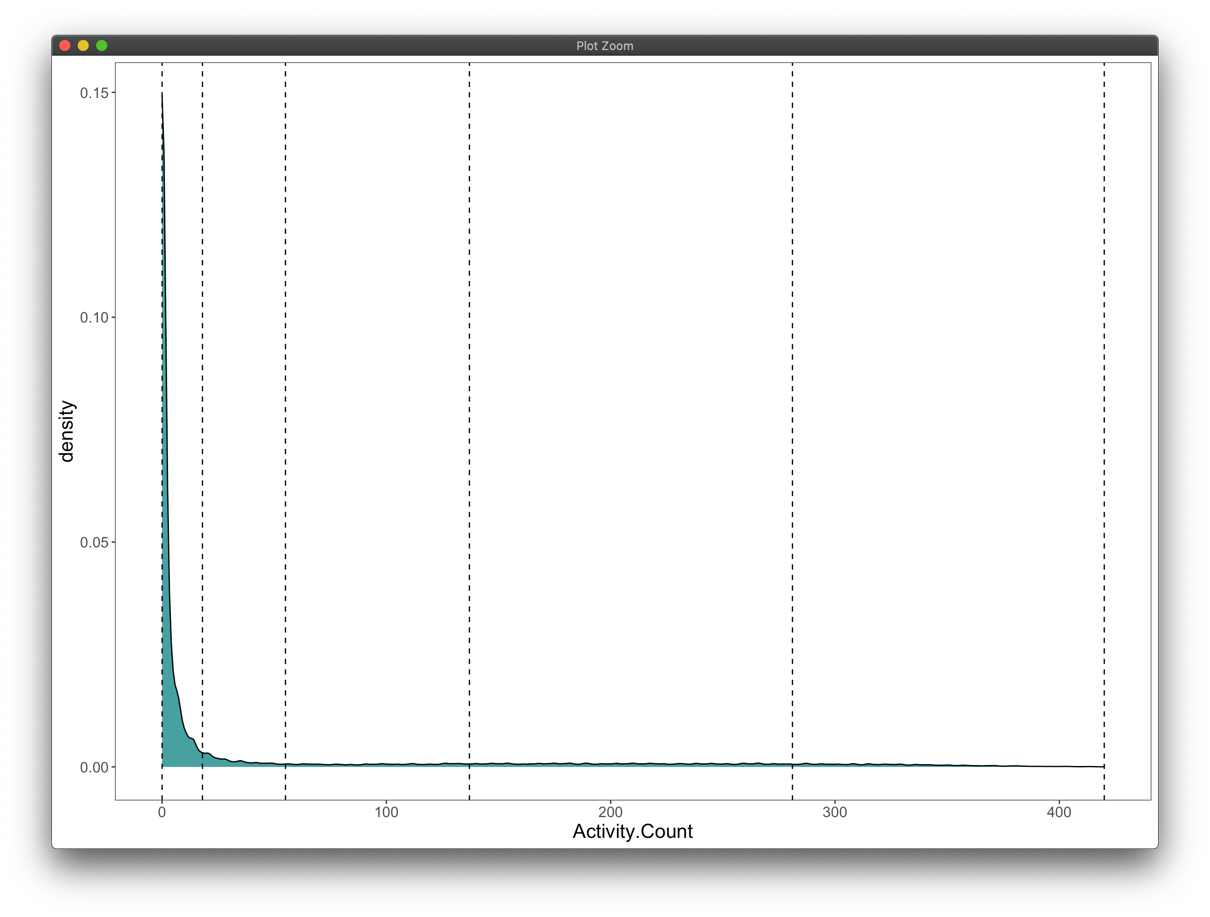
**Methods**

Movement data were available for 7 individuals (blanca, emanuel, gala, mafalda, mazeboti, sara, tex) from 2019 – 2020. A sample size of N = 18,433 total GPS locations and N = 140,162 total activity count measurements were analyzed from all 7 IDs. The tags recorded activity counts (i.e., the changes in the x, y, or z axes over a 5 min duration) and GPS positions, which were used to calculate step lengths and turning angles. These movement data were heavily filtered to remove error-prone measurements that could bias the results. This included the removal of aberrant locations due to low satellite coverage (≤ 4 satellites), step lengths that were larger than possible (> 800 m; 99th quantile), and when activity counts were equal to zero. Since activity counts were recorded at 5 min intervals and GPS locations were recorded at 7 min intervals, activity counts were only retained when sampled at a time ≤ 1 min different from the GPS tag. Additionally, GPS positions were filtered to only include observations sampled at 7 ± 1 min. Step lengths were then converted to speed by dividing by the time step (i.e., 7 min). After all filtering steps, a total of 9,948 observations were retained for further analysis. Next, speed, turning angles, and activity counts were discretized into bins for analysis by the non-parametric Bayesian mixture model.

Activity counts ranged from 1 to 300 and were discretized into 6 bins of equal width. Speed ranged from 0 to 2.13 m/s and was discretized into 7 bins. Since 99% of speed observations were recorded below 1 m/s, 6 bins of equal width were used to discretize speed from 0 to 1 m/s. All observations greater than 1 m/s (up to 2.13 m/s) were assigned to the 7th bin. Turning angles ranged from to radians and were discretized into 10 bins of equal width. Plots of continuous and discretized distributions are shown below. Additionally a plot of filtered tracks for each ID is also included for spatial context.

Chart, scatter chart

Description automatically generated



A picture containing chart

Description automatically generated

A picture containing water, boat, table, sitting

Description automatically generated

Figure 3 This turning angle distribution for all armadillos is biased by a single individual (‘mazeboti’), that possess one of the largest sample sizes and includes a high density of values near 0 radians.

*Segmentation model*

Each unique day was used to pre-specify breakpoints for the model per ID since there was a clear temporal pattern of activity each day. Based on traceplots of the log marginal likelihood for each individual, the model appeared to converge after 20000 iterations.

A picture containing building, window

Description automatically generated

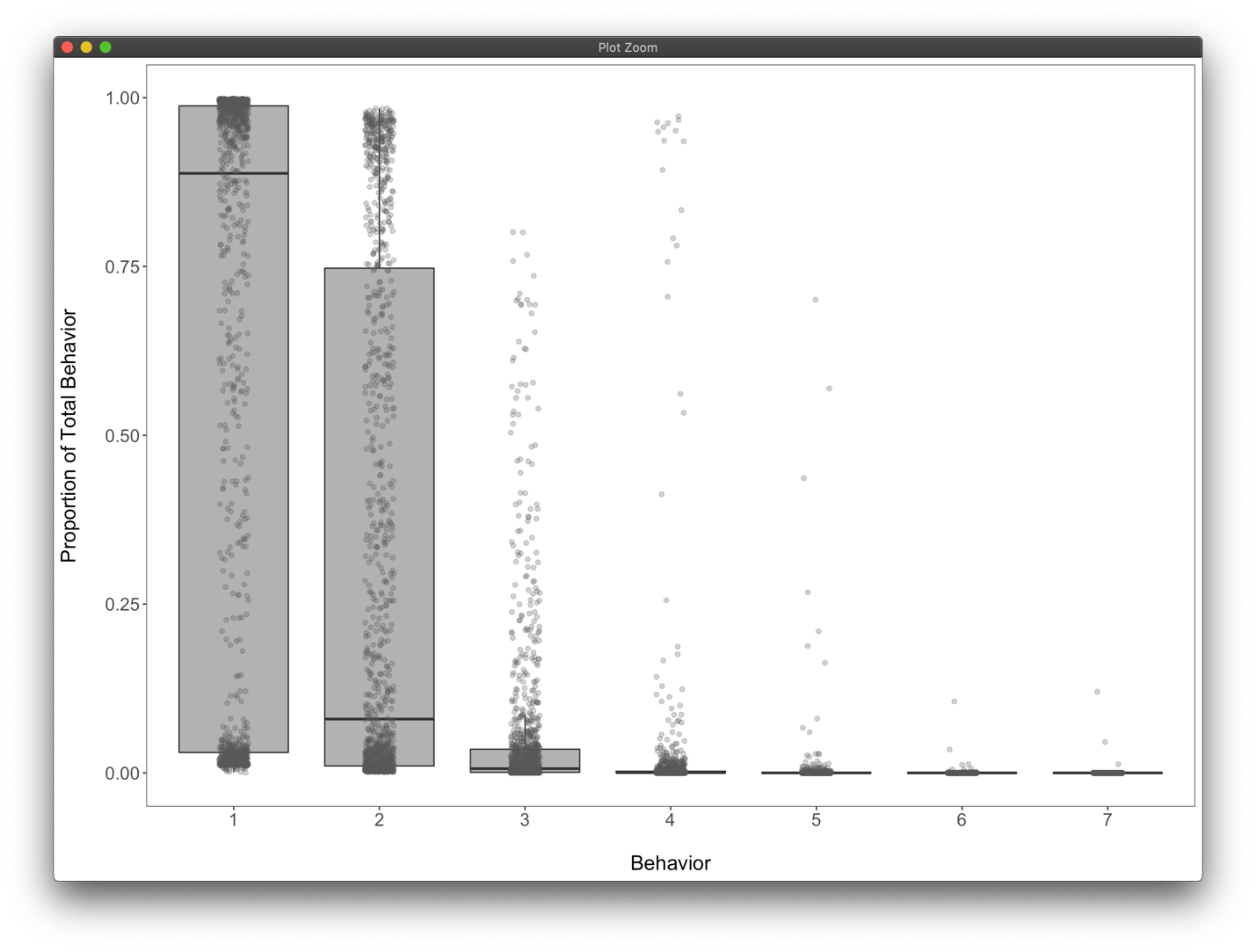
A close up of a window

Description automatically generated

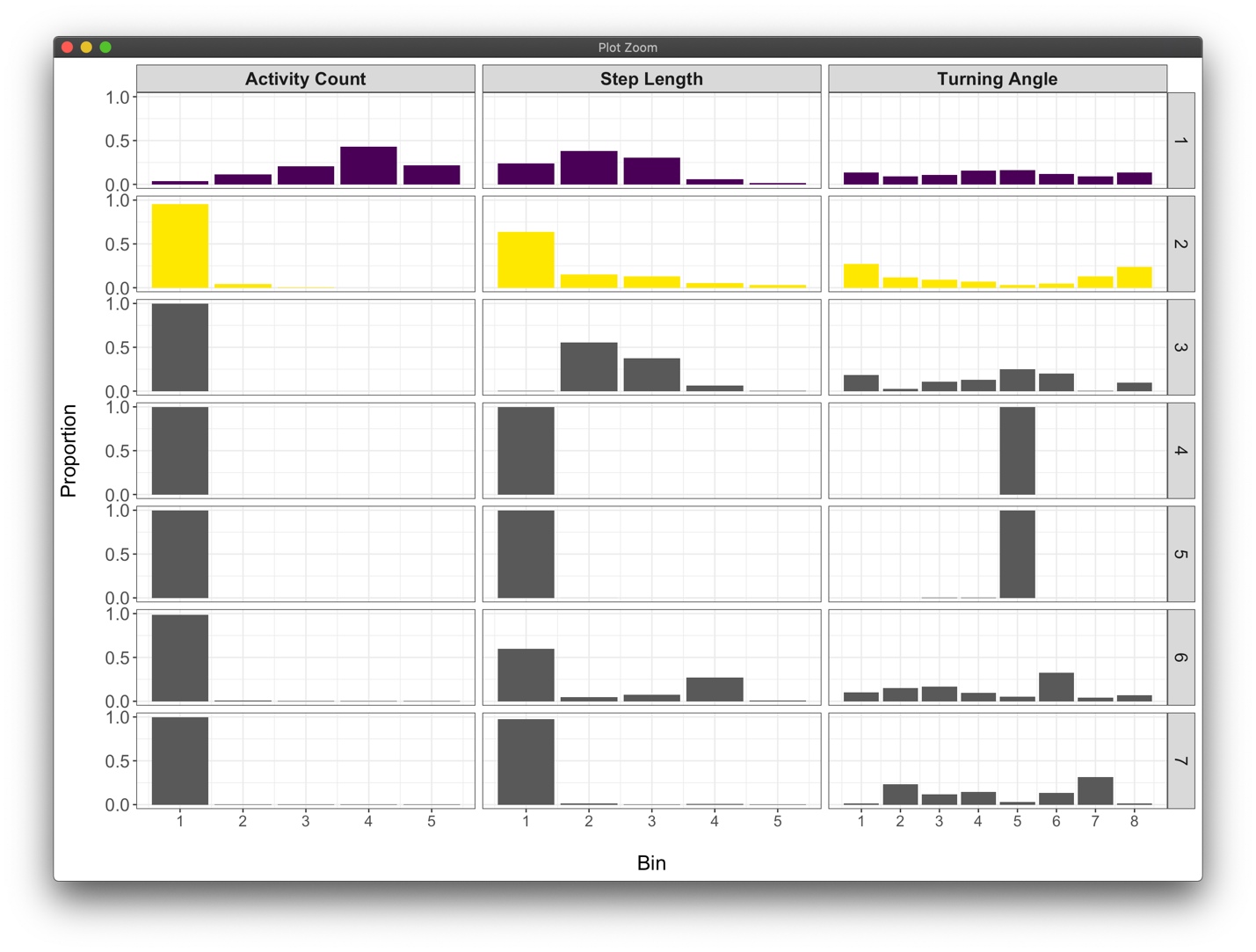
Only a couple examples of the segmentation results are shown to demonstrate breakpoint estimation. Most IDs (e.g., ‘gala’) estimated breakpoints at least once a day if not more frequently. However, ‘mazeboti’ has 3 segments in the middle of its time series that are each quite long (i.e., comprised of multiple days/weeks). This appears to be due to the lack of available Activity Count data during these segments, but that still has yet to be confirmed. These data did not include time spent resting in the burrow during the day, so results presumably reflect active behavioral states only. Additionally, thousands of observations of Activity Count data (w/o corresponding GPS data) were omitted from these analyses since they overwhelm the step lengths and turning angles that were analyzed.

*LDA model*

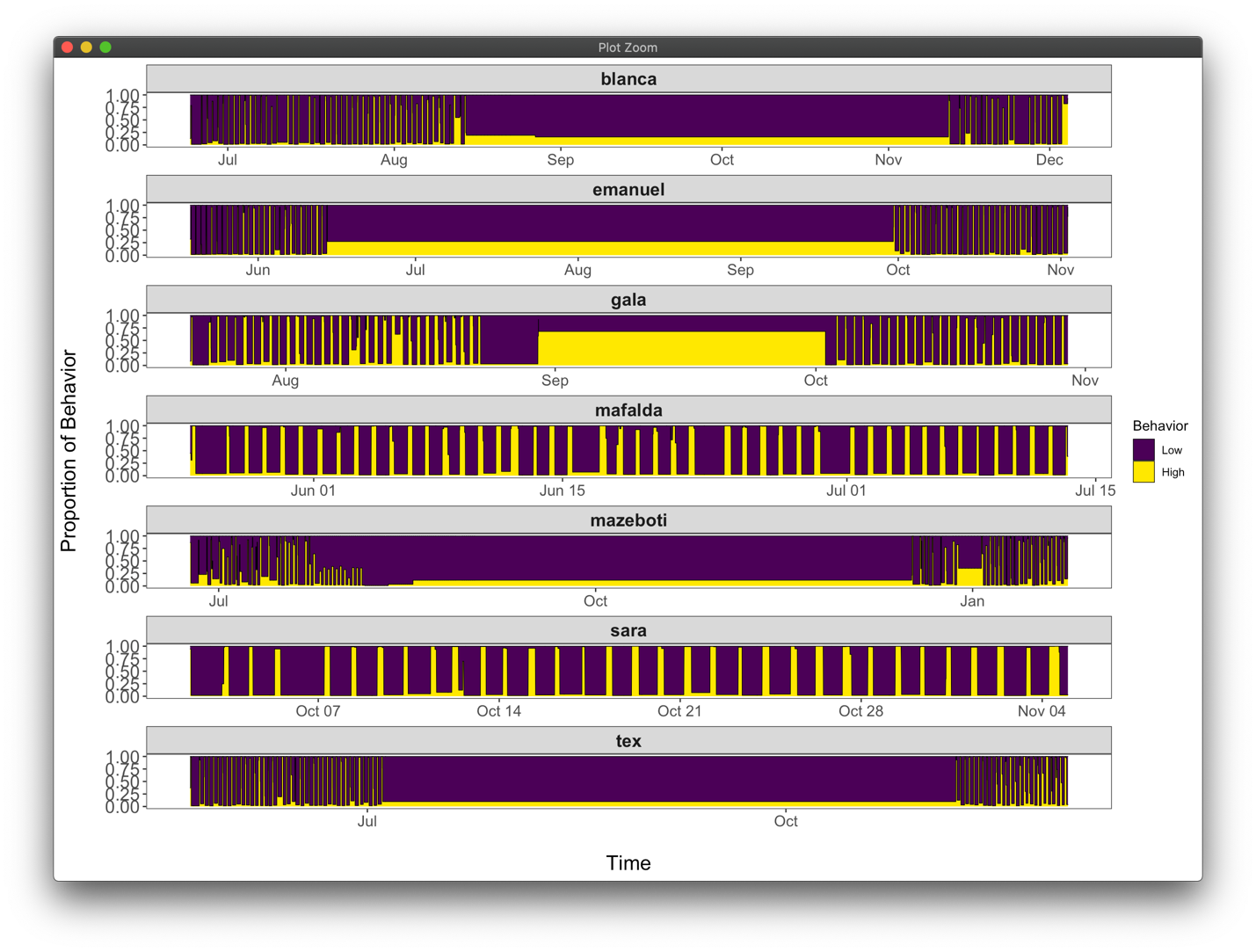
Following the clustering of the segments across all 7 individuals, plots were inspected to determine the optimal number of behavioral states as well as the state-dependent distributions.



Based on this boxplot of results from the theta matrix, the first 2 states/clusters accounted for 93.8% of all observations on average. Therefore, these clusters were retained for further analysis.



Plots of the bin distributions for each of the 7 possible states indicate that these first 2 states are biologically distinct and relevant. Based on these distributions, state 1 appears to represent a high activity “transit” state, whereas state 2 represents a low activity “foraging” state. It should be mentioned that presumed “burrow” behavior was not identified by the model since GPS data were not available during the day (and Activity Counts were only kept when matching GPS observations).



Time series plots of these behavior proportions for each individual armadillo are shown next and clearly depict regular intervals of higher and lower activity. The long periods in the middle of the time series with minimal change in behavior proportion is likely an artefact of the visualization function used, meaning that there are large gaps of missing data during these periods.

Graphical user interface, website

Description automatically generated

*Conclusions*

Based on these findings, it appears that our model performs well at characterizing latent states from a combination of Activity Counts, Step Length, and Turning Angle data streams. While only this single analysis is included here, previous analyses that only evaluated Step Length and Turning Angle or that tried to generate fake Step Lengths and Turning Angles during “burrow” behavior resulted in breakpoints that did not identify daily patterns and therefore improperly classified behavioral states. Therefore, “burrow” behavior may need to be classified for observations outside of the model.

The next potential step, if wanting to classify all Activity Count data, would be to make predictions based on our model results. This would depend on how closely the Activity Count distributions match from the “held out” data compared to those that were classified.