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

Josh Cullen

28 January 2021

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

Data prep

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 ($\leq$ 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 $\leq$ 1 min different from the GPS tag. Additionally, GPS positions were filtered to only include observations sampled at 7 $\pm$ 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 $-\pi$ to π 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.

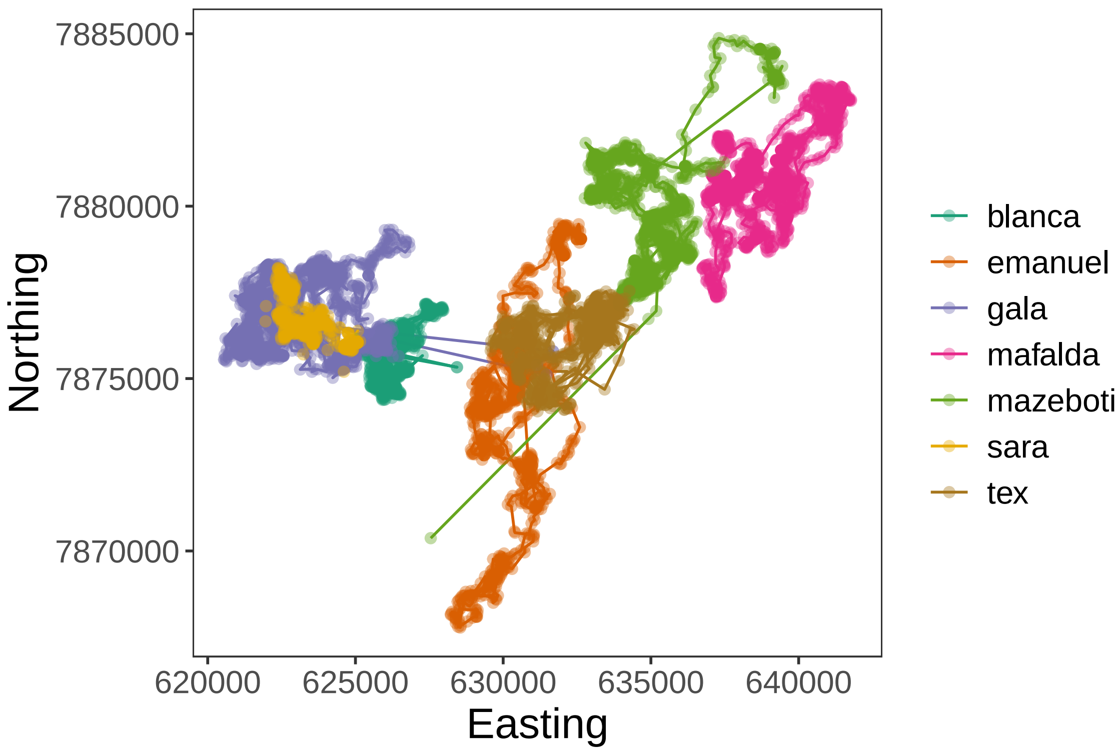


Figure 1: Tracks of retained GPS positions for all 7 giant armadillos.

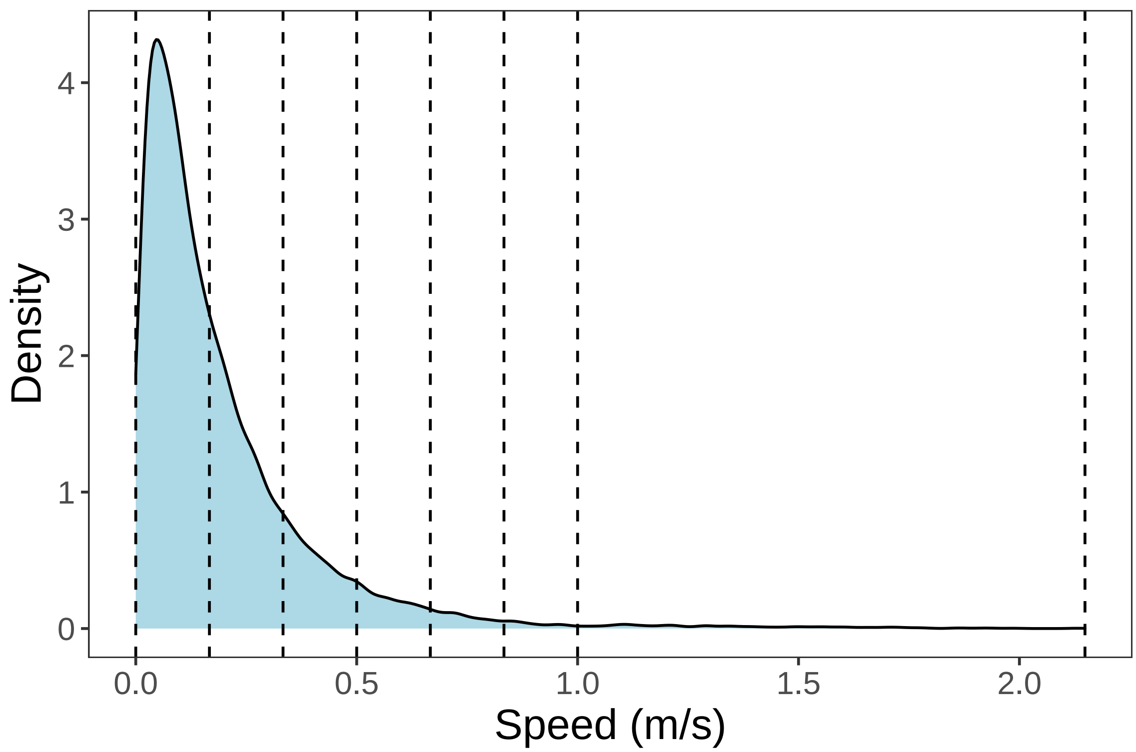


Figure 2: Density plots of speed, turning angles, and activity counts.

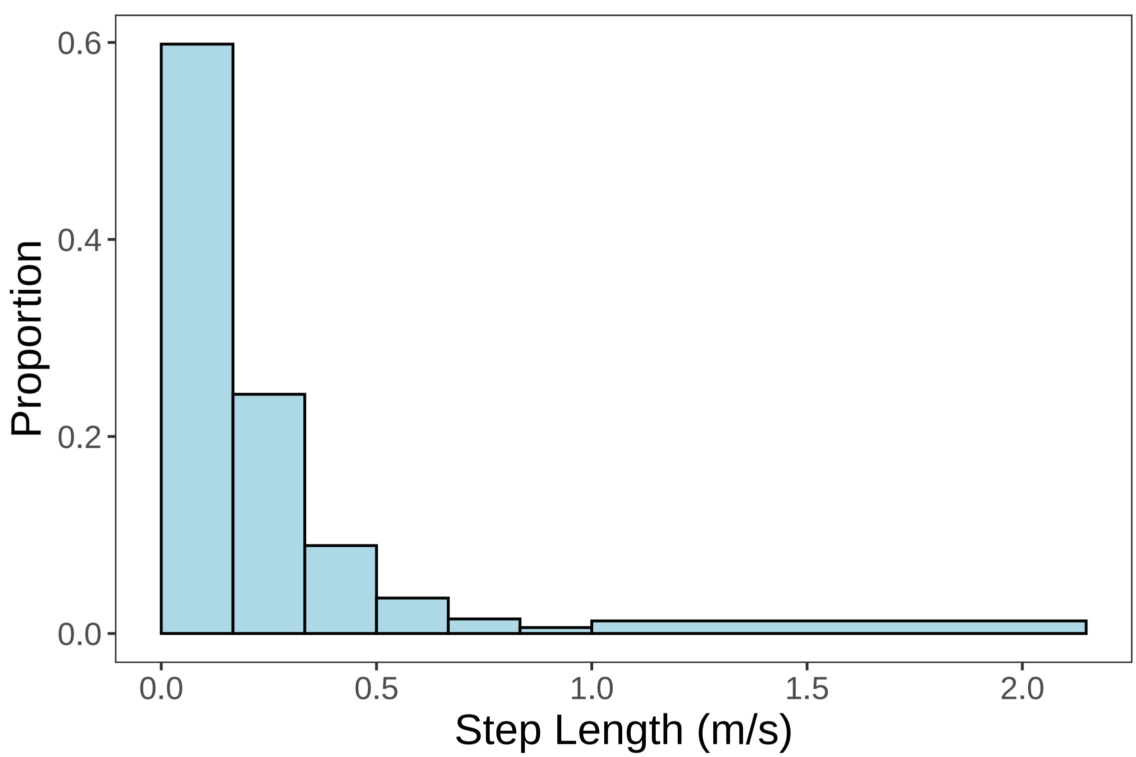


Figure 3: Density plots of speed, turning angles, and activity counts.

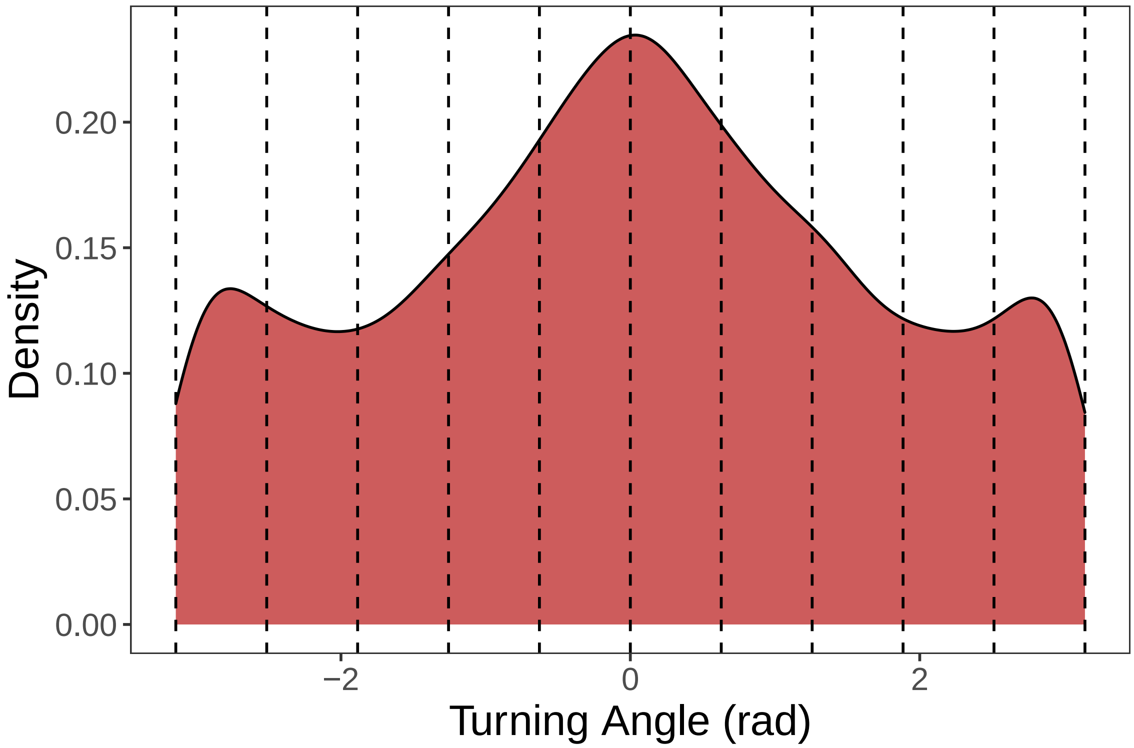


Figure 4: Density plots of speed, turning angles, and activity counts.

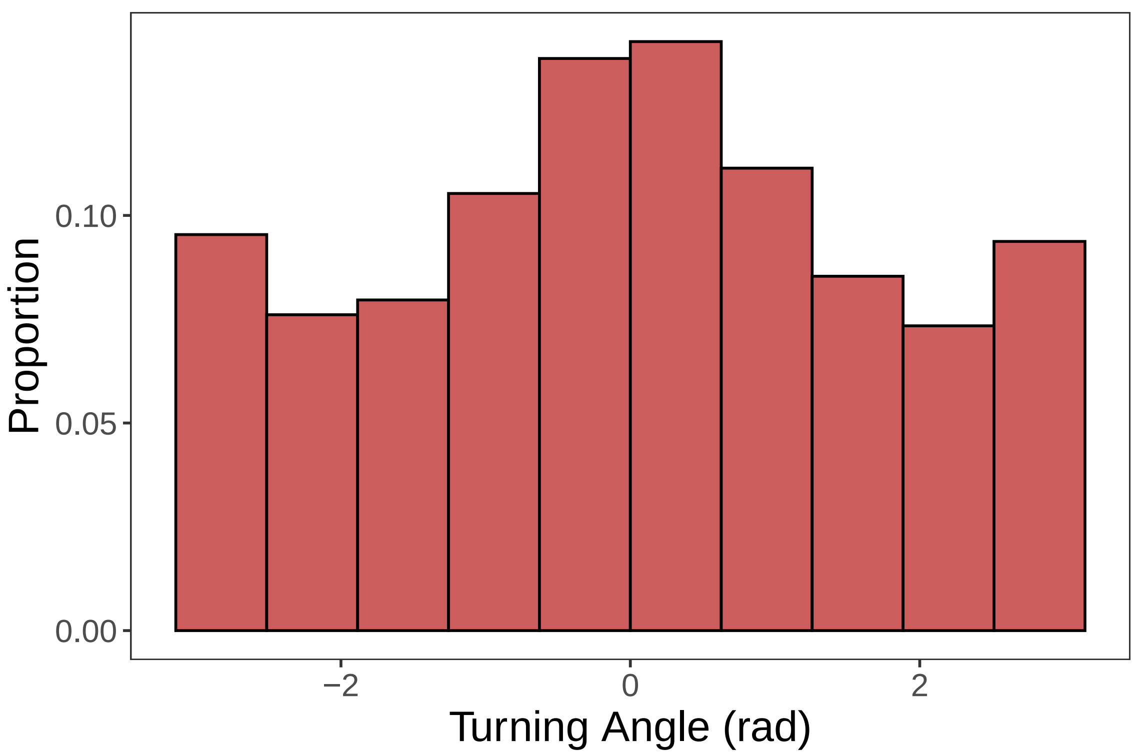


Figure 5: Density plots of speed, turning angles, and activity counts.

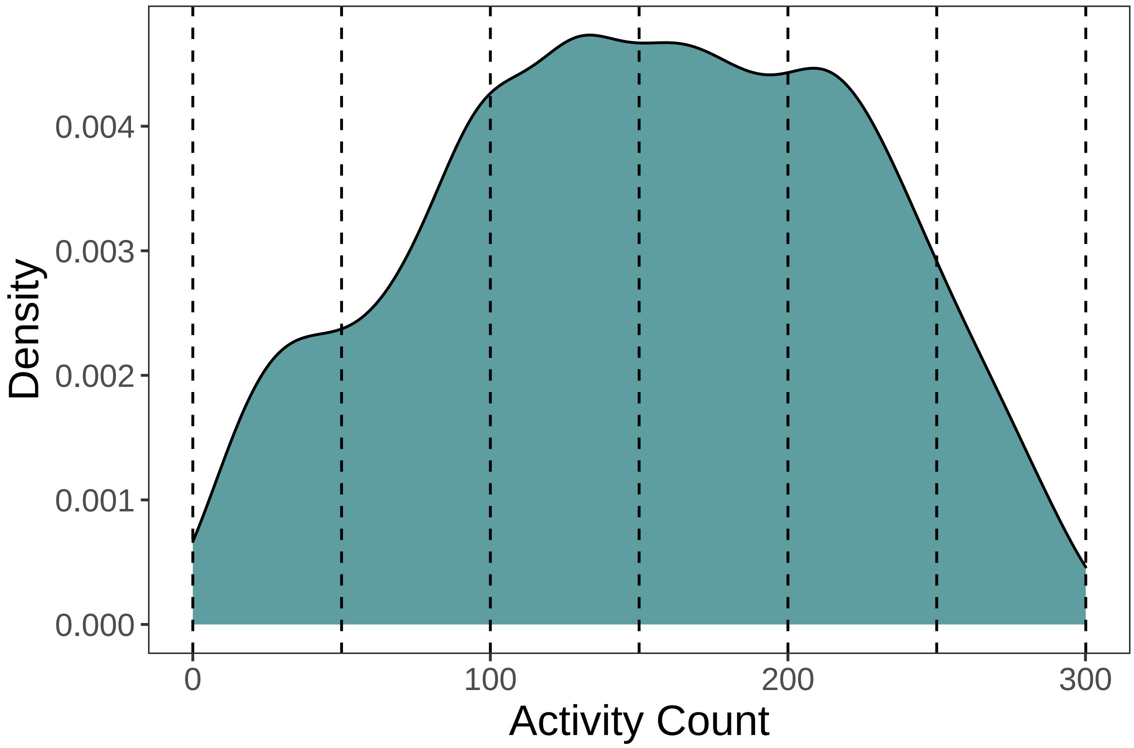


Figure 6: Density plots of speed, turning angles, and activity counts.

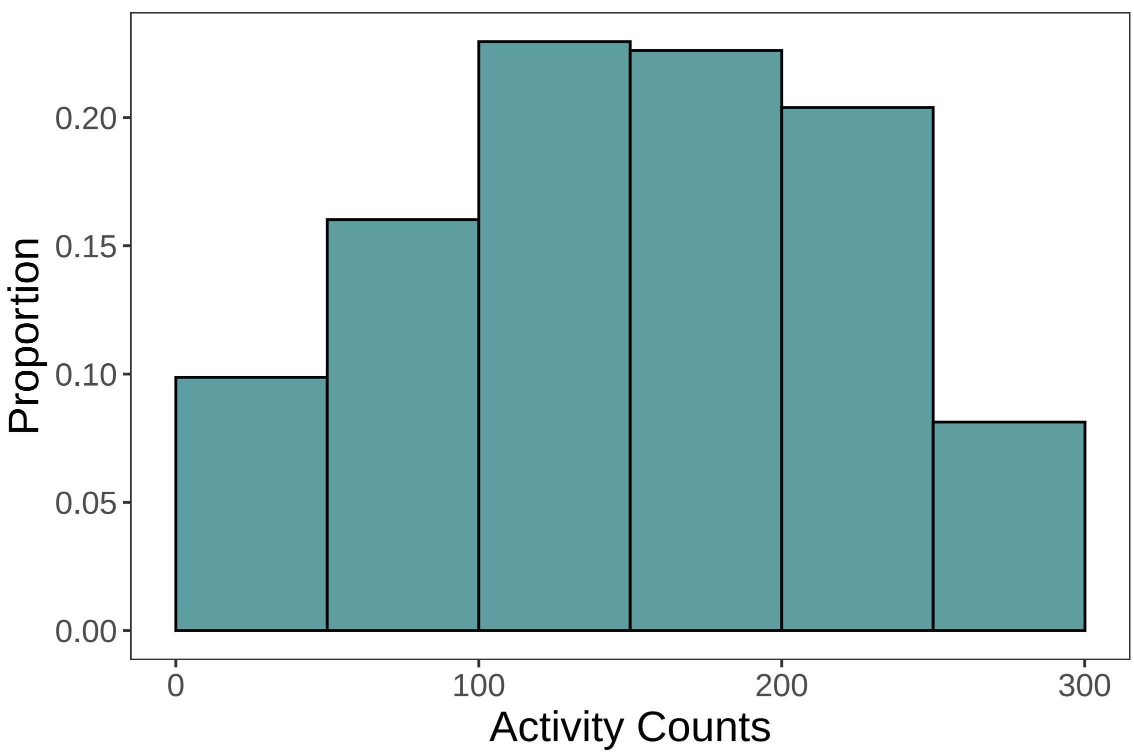


Figure 7: Density plots of speed, turning angles, and activity counts.

Clustering observations into latent behavioral states

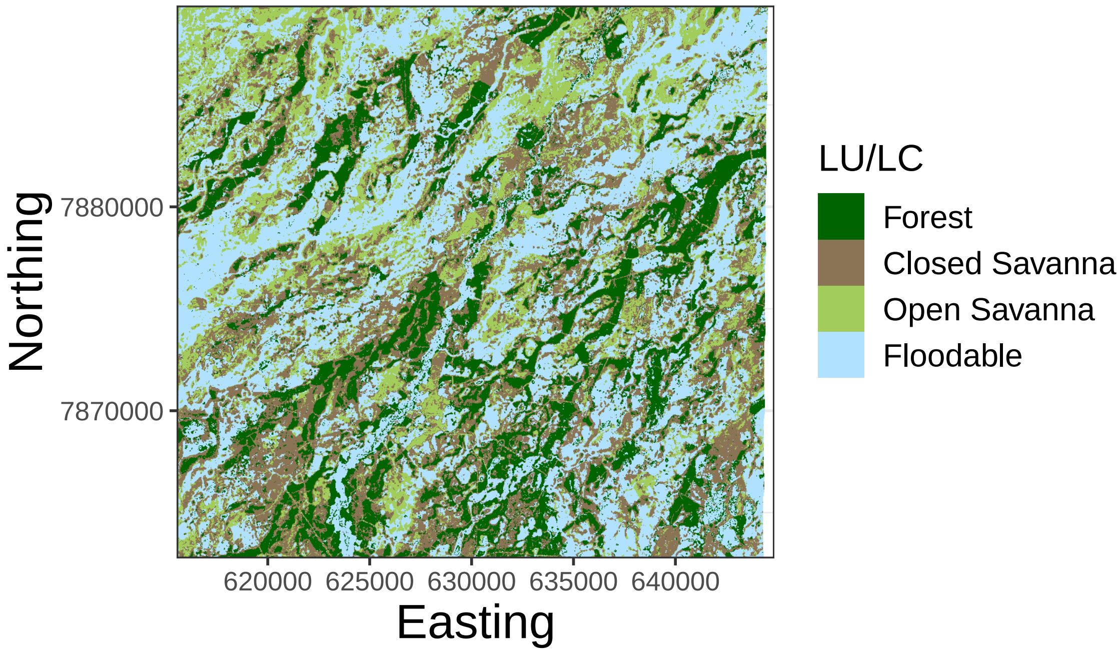
Filtered data were analyzed by a non-parametric Bayesian mixture model using the `cluster_obs()` function in the *bayesmove* package in R. This model clusters observations pooled across all individual tracks into a finite number of latent behavioral states. Through the use of a penalizing prior, we can set the maximum number of possible states and the model will prefer to assign observations to a fewer number of clusters/states. This clustering was performed using the bin labels for all three data streams (i.e., activity counts, speed, and turning angles) from all 9,948 observations where the maximum number of clusters was set equal to 10. We ran a Gibbs sampler using Markov chain Monte Carlo (MCMC) with 20,000 iterations where the first 10,000 iterations were treated as the burn-in and the second 10,000 iterations constituted the posterior.

The number of latent states was identified by inspecting the average proportion of observations assigned to each cluster/state as well as the state-dependent distributions of activity counts, speed, and turning angles. Once the states were defined, the posterior estimates of state assignments for each observation were evaluated. To ensure that there was high certainty with the assigned behavioral state for a given observation, $\geq 75\%$ of all posterior estimates needed to belong to a single state. If none of the states for a given observation fell above this threshold, the observation was labeled as "Unclassified".

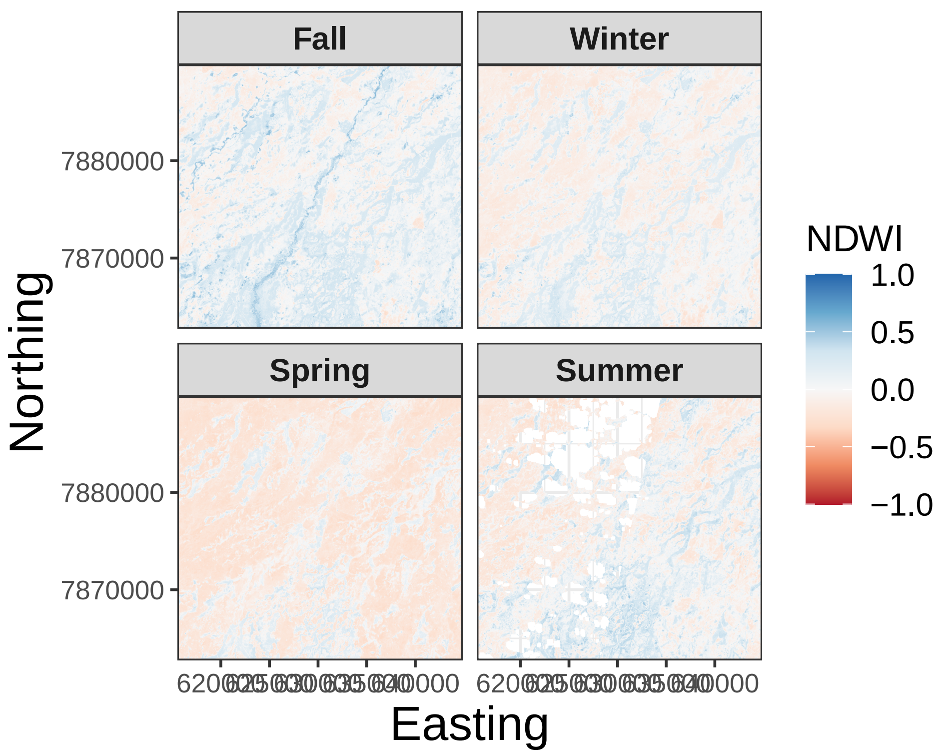
Extraction of environmental covariates

To relate the estimated behavioral states to giant armadillo habitat, a number of environmental covariates were extracted at each location. This included land use/land cover (LU/LC), normalized difference vegetation index (NDVI), normalized difference water index (NDWI), elevation, and Tasseled Cap Transformations of Landsat imagery (brightness, greenness, wetness). Tasseled Cap brightness, greenness, and wetness were calculated as possible alternatives to LU/LC and NDVI/NDWI since they are orthogonal metrics intended to represent bare ground, vegetation productivity, and soil moisture, respectively. The LU/LC data was obtained from ICAS and consisted of four land classes: Forest, Closed Savanna, Open Savanna, and Floodable. These data were stored at a 30 m spatial resolution with as a static variable (i.e., it did not change over time). Elevation was downloaded for the study region using the *elevatr* R package at 18 m resolution, but was resampled to 30 m to match the other covariates. NDVI, NDWI, brightness, greenness, and wetness were derived from Landsat 8 imagery over the time period for which the giant armadillos were transmitting (May 2019 to January 2020). These were downloaded at a spatial resolution of 30 m, but were aggregated to an average seasonal layer to minimize data missingness due cloud cover.

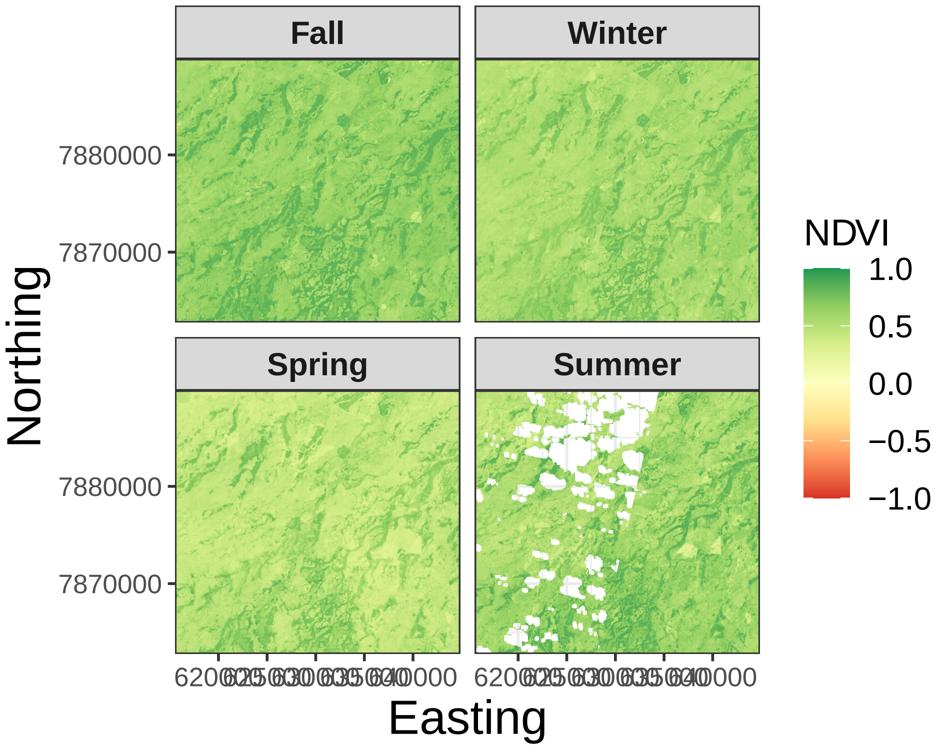
Environmental covariates were extracted using a 30 m buffer around each GPS position. This was performed to account for possible location error from the tag and capture any local heterogeneity in the habitat for a given location. Since LU/LC is a discrete variable, the proportion of land classes was calculated for each position. The mean was calculated for all other covariates since they are continuous variables.



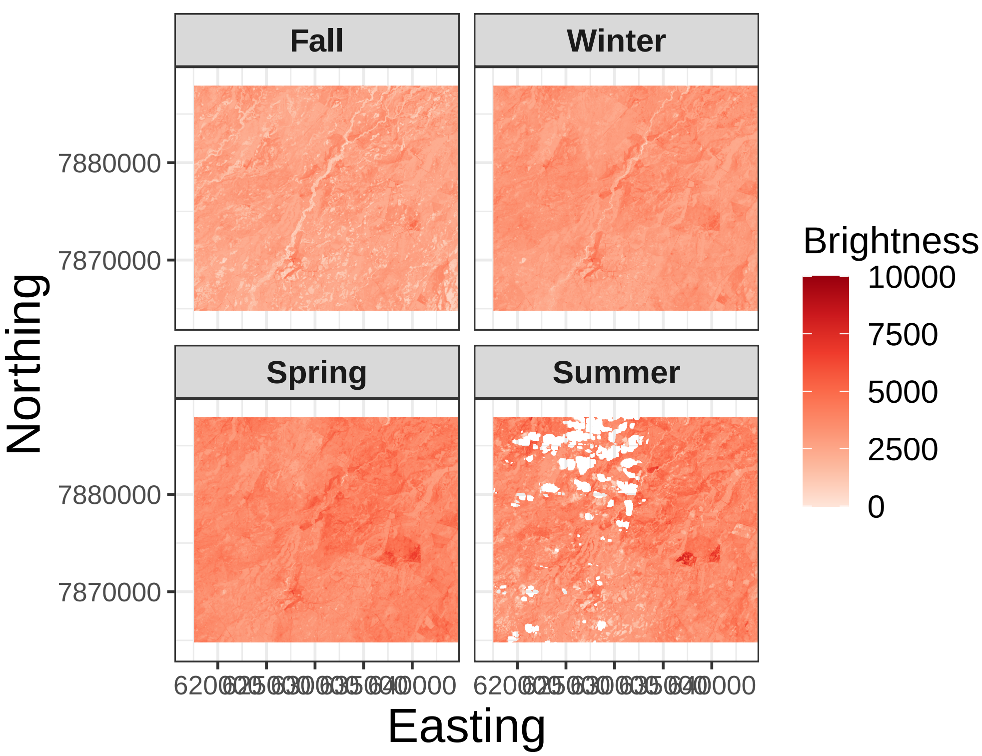
LU/LC
Forest
Closed Savanna
Open Savanna
Floodable
Northing
7880000
7870000
620000
625000
630000
635000
640000
Easting



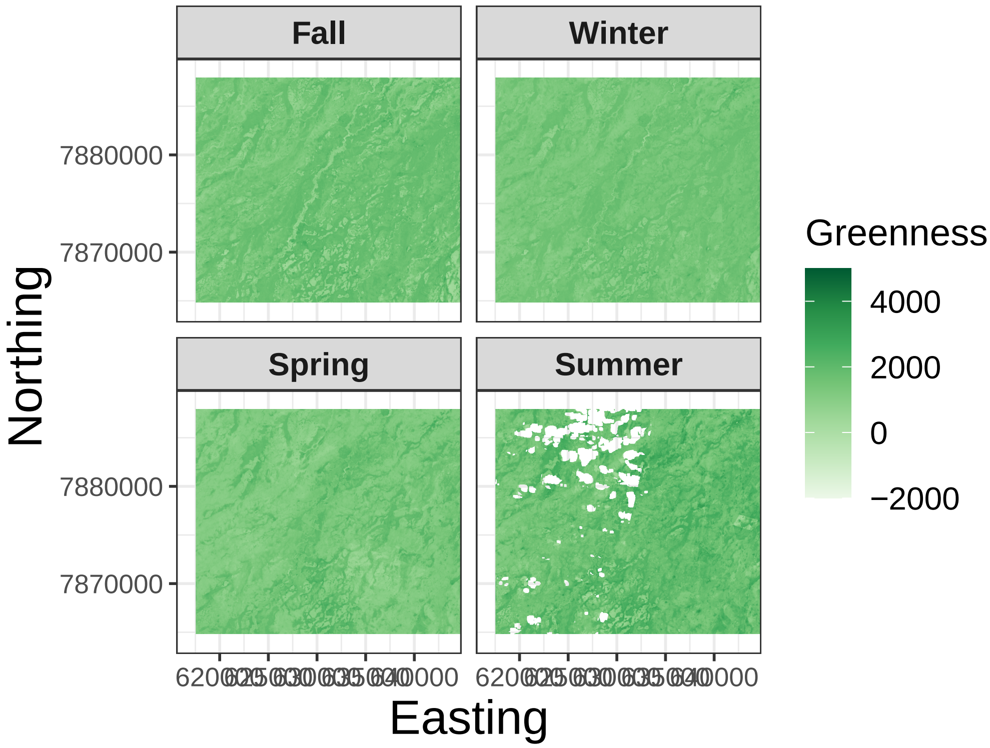
Fall
Winter
Spring
Summer
NDWI
1.0
0.5
0.0
−0.5
−1.0
Northing
7880000
7870000
Easting

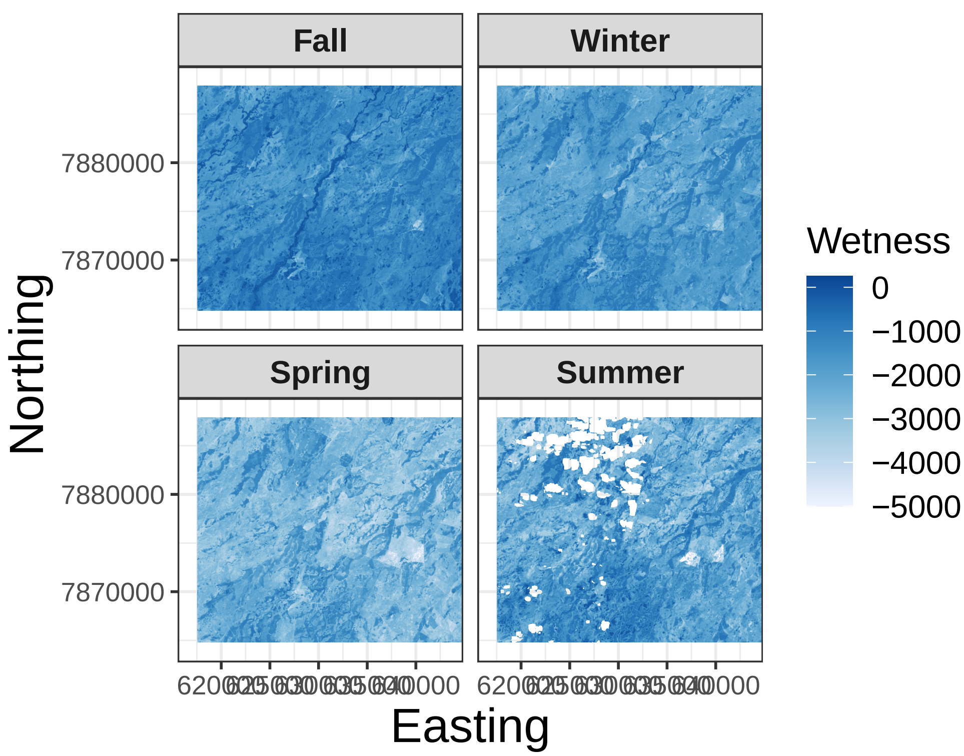


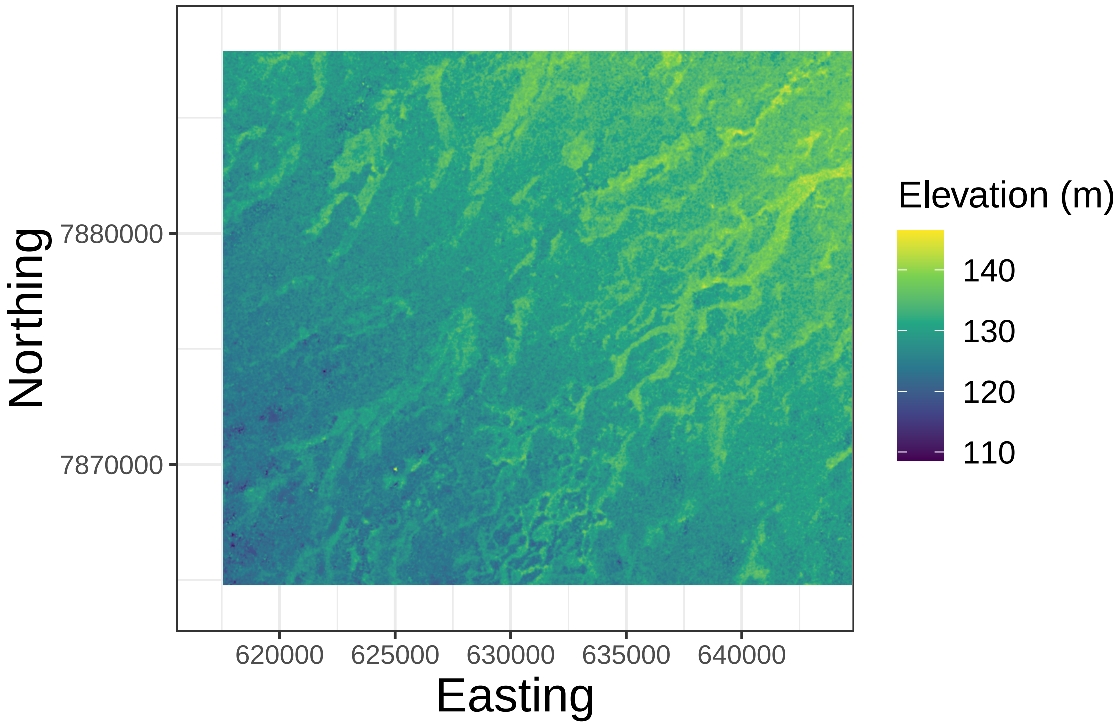
Fall
Winter
Spring
Summer
NDVI
1.0
0.5
0.0
−0.5
−1.0
7880000
7870000
Northing
Easting



Fall
Winter
Spring
Summer
Brightness
10000
7500
5000
2500
0
7880000
7870000
Northing
Easting


Fall
Winter
Spring
Summer
7880000
7870000
Northing
620000 625000 630000 635000 640000
Easting
Greenness
4000
2000
0
−2000


Fall
Winter
Spring
Summer
7880000
7870000
Northing
620000 625000 630000 635000 640000
Easting
Wetness
0
−1000
−2000
−3000
−4000
−5000



Comparison of habitat use among behavioral states

To quantify the differences in relationships with environmental covariates among behavioral states, logistic regressions were performed to evaluate pairwise differences. Due to high correlations among some of the variables and seeming redundancy of others, **only greenness and wetness** were included in subsequent analyses. Additionally, armadillo ID was also included as a covariate within the logistic regressions. Blanca was treated as reference class (part of intercept) for animal ID. For quick visualization of the model results, coefficients from each model comparison were plotted as odds ratios with 95% confidence intervals. This was performed for each pairwise comparison of states.

Results

Estimation of latent behavioral states

The mixture model suggested that four states were optimal based on the average number of observations assigned to each state (top four accounted for 98% of all observations) and an inspection of the state-dependent distributions. By using a threshold of 75% posterior samples assigned to a single state for a given observation, 38% of the 9,948 observations were unclassified.

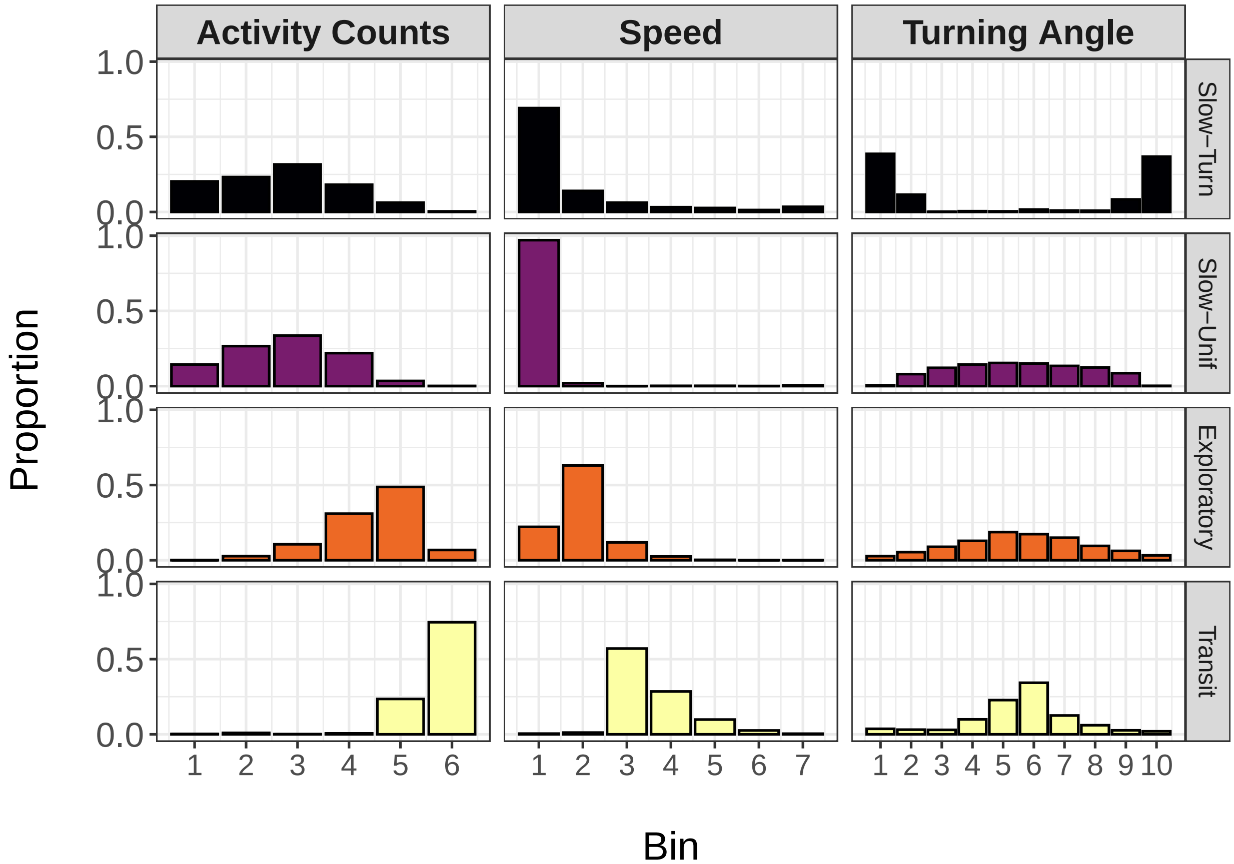


Figure 8: State-dependent distributions for each of the four retained states from the mixture model. Bin numbers reflect discretized variables from previous figures.

Given my limited background in the behavioral ecology of giant armadillos, behavioral states are given preliminary names that can be later modified. For now, they are identified as 'Slow-Turn', 'Slow-Uniform', 'Exploratory', and 'Transit'. The 'Slow-Turn' state is characterized by relatively lower activity counts, low speed, and high turning angles (near $-\pi/\pi$ radians). The 'Slow-Uniform' state is nearly identical to the 'Slow-Turn' state for distributions of activity counts and speed, but has a more uniform distribution of turning angles. The 'Exploratory' state is characterized by higher activity counts, greater speed, and lower turning angles (near 0 radians) than either the 'Slow-Turn' or 'Slow-Uniform' states. Lastly, the 'Transit' state exhibits the greatest activity counts, greatest speeds, and lowest turning angles.

A quick inspection of the plotted behavioral states provides an idea of their spatial configuration for each individual:

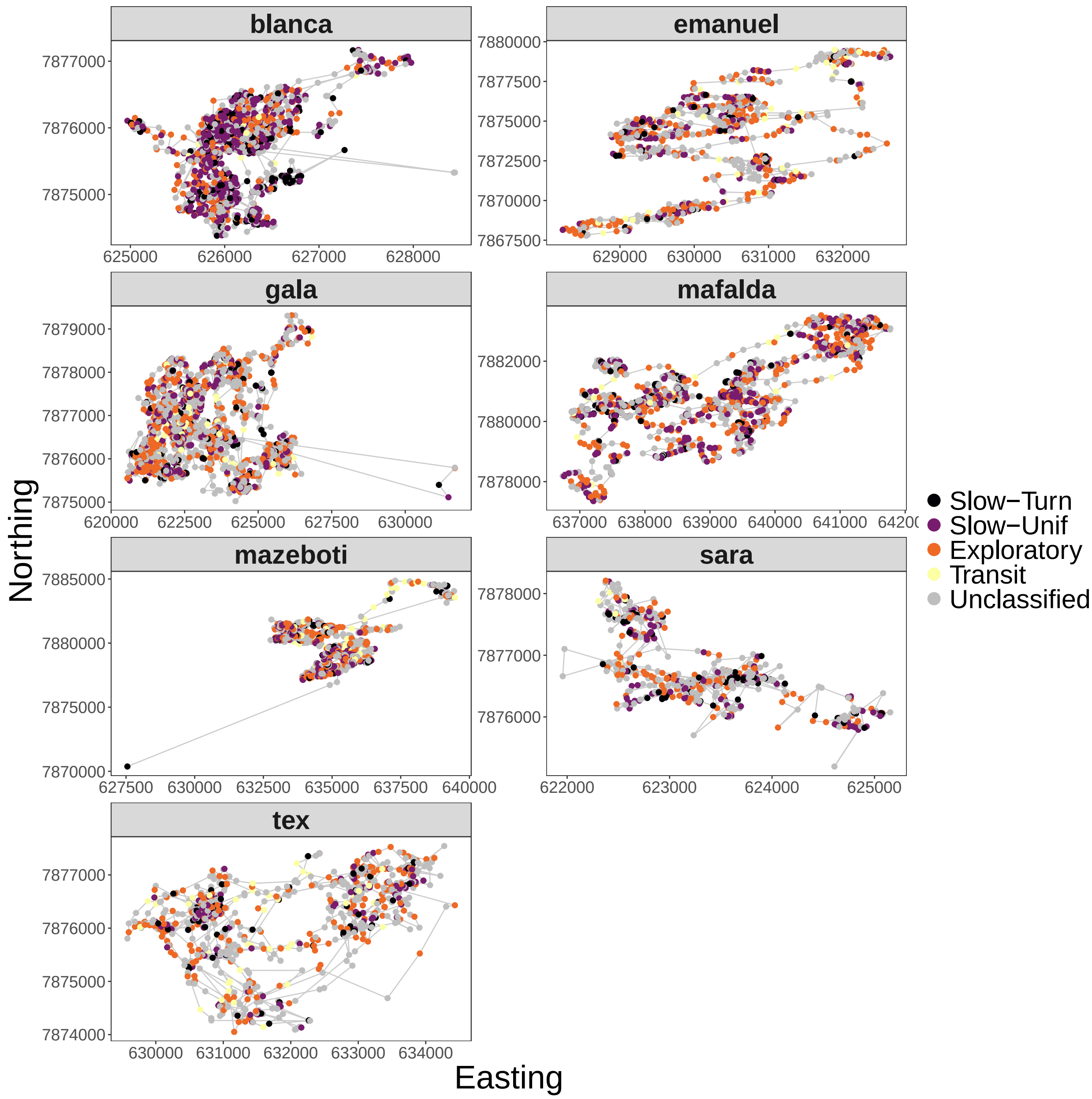


Figure 9: Mapped behavioral states from the mixture model for each armadillo

Exploratory data analysis

Below is an example of a track segment where LU/LC and elevation would be extracted within 30 m of a given observation. Based on these values, as well as other characteristics of the data, we can summarize the data a few different ways to explore any obvious underlying spatiotemporal patterns for the behavioral states.

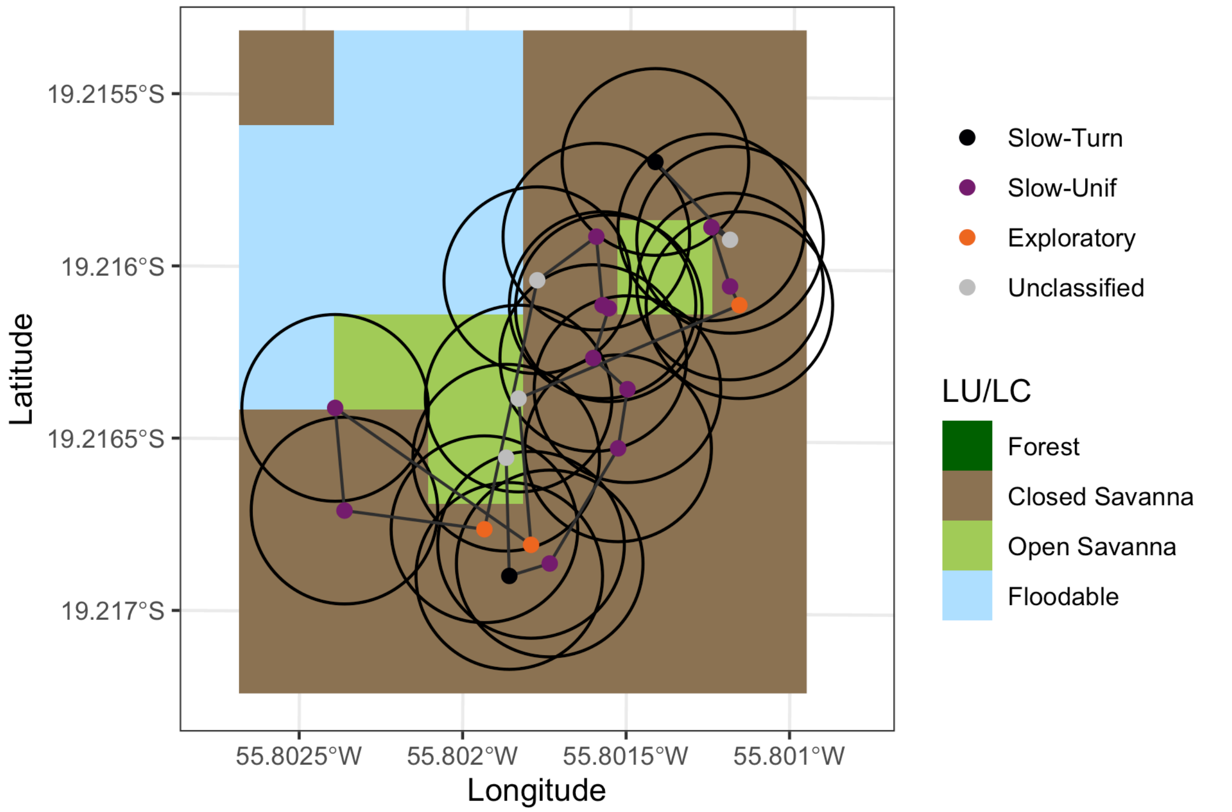


Figure 10: Subset of blanca's first 20 observations. State estimates are color coded over LU/LC and elevation and 30 m buffers used to extract proportions of land classes and mean elevation are shown for each position.

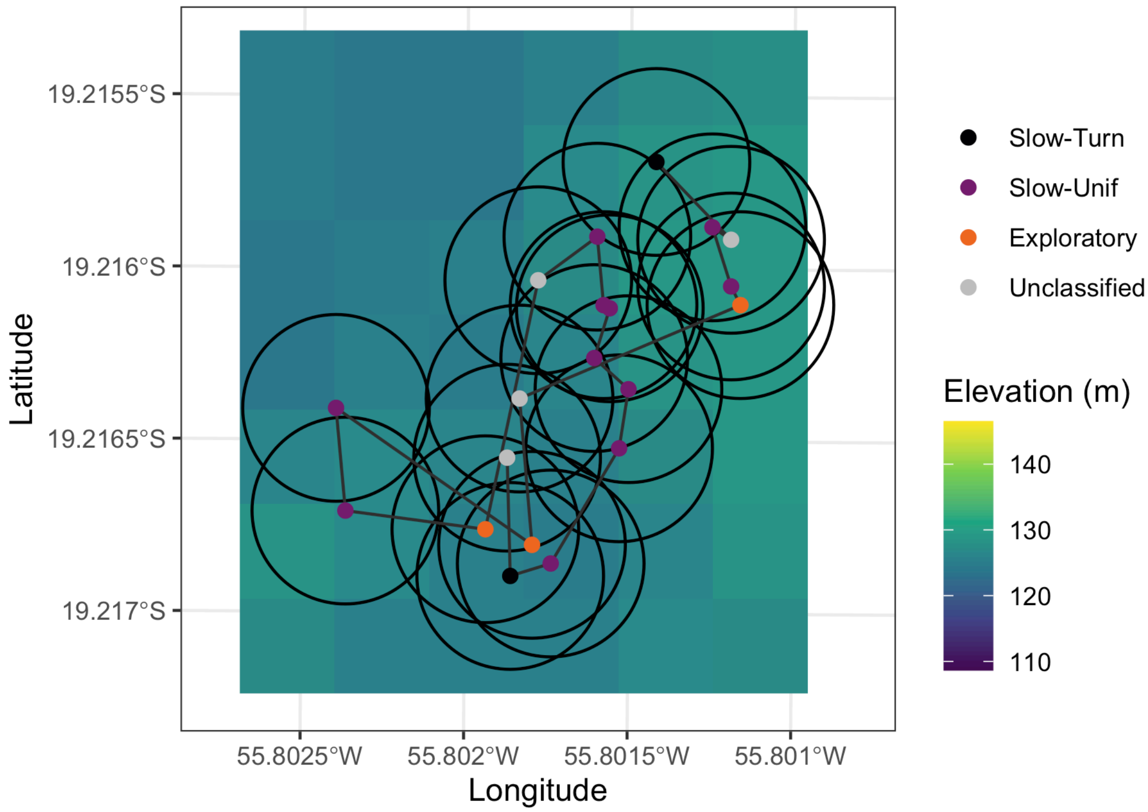


Figure 11: Subset of blanca's first 20 observations. State estimates are color coded over LU/LC and elevation and 30 m buffers used to extract proportions of land classes and mean elevation are shown for each position.

I summarized the behavioral states by ID, season, and hour of day. An overview of the activity budgets for each ID shows that there is very little difference among IDs, where Exploratory and Slow-Uniform states are the most common and occur at approximately the same frequency (except in blanca). The Slow-Turn state is the next most common, followed by the Transit state. These patterns are also observed when visualizing activity budgets by season. However, it appears that Slow-Uniform and Exploratory states occur at the same frequency during the fall and spring, but the Slow-Uniform state occurs more often during winter and summer. While Unclassified observations typically comprised the greatest proportion of observations, they were omitted in some circumstances to facilitate interpretation of the remaining data. An assessment of activity budgets broken down by hour of the day shows greatest activity from midnight to 4 am. However, there does appear to be some activity occurring after sunrise (i.e., 6:30 to 10 am), which is uncommon since this is a nocturnal species. By explicitly evaluating the proportion of observations attributed to each state per hour, there does not appear to be much of a discernible pattern. State proportions appear to be relatively constant and exhibited at proportions similar to overall frequencies in the data apart from the first and last hour periods of activity.

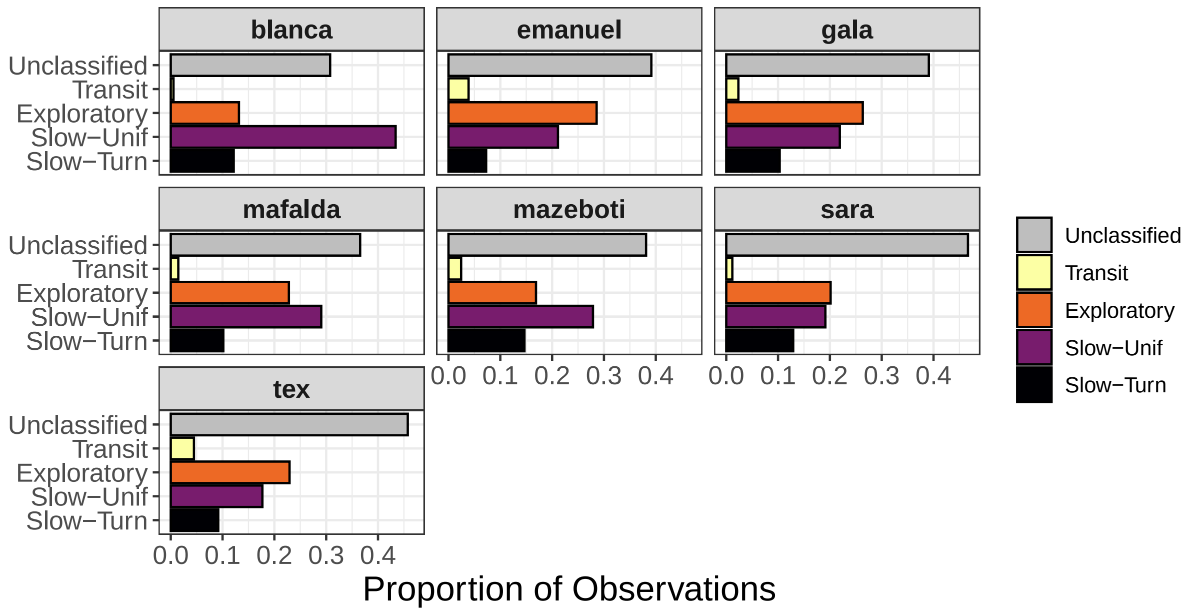


Figure 12: Summary of behavioral states by ID, season, and hour of day.

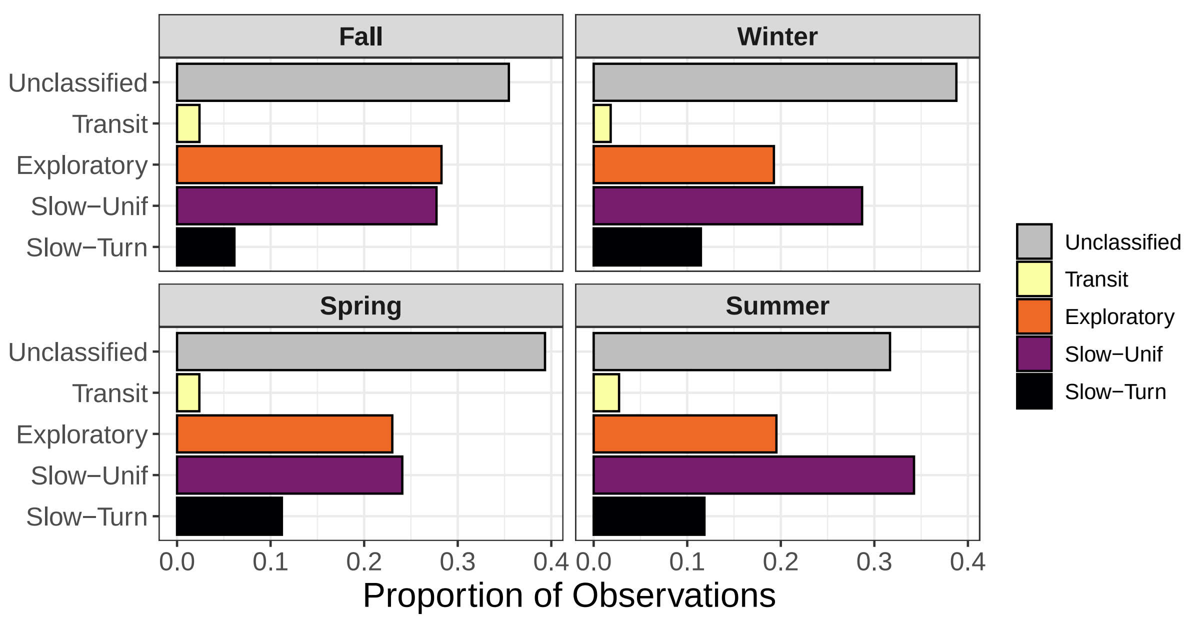


Figure 13: Summary of behavioral states by ID, season, and hour of day.

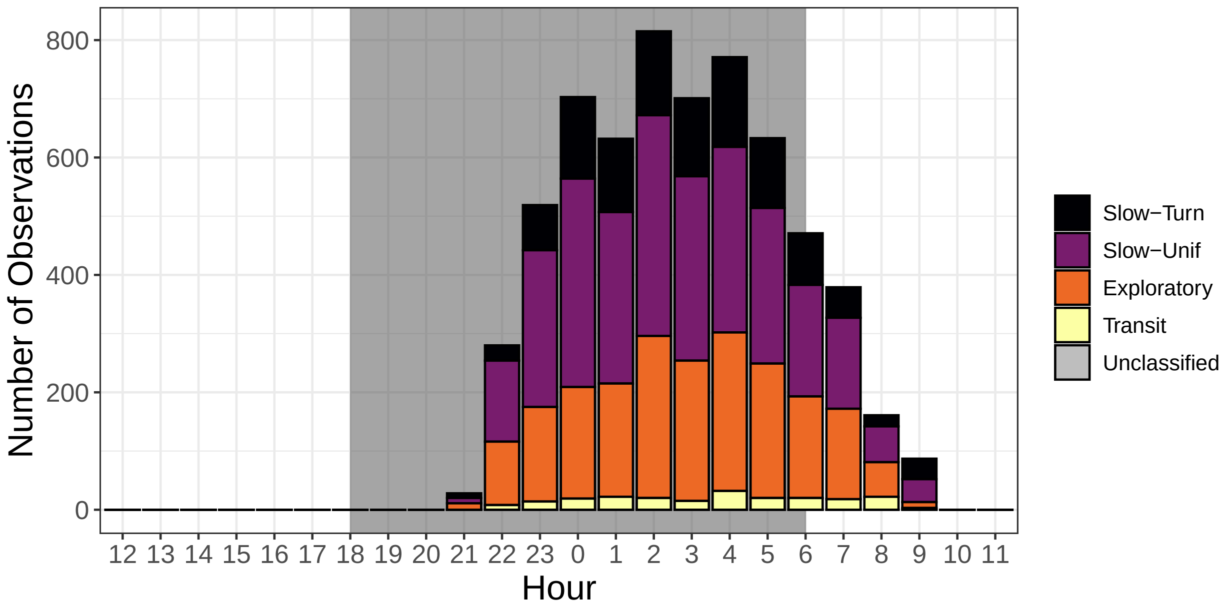


Figure 14: Summary of behavioral states by ID, season, and hour of day.

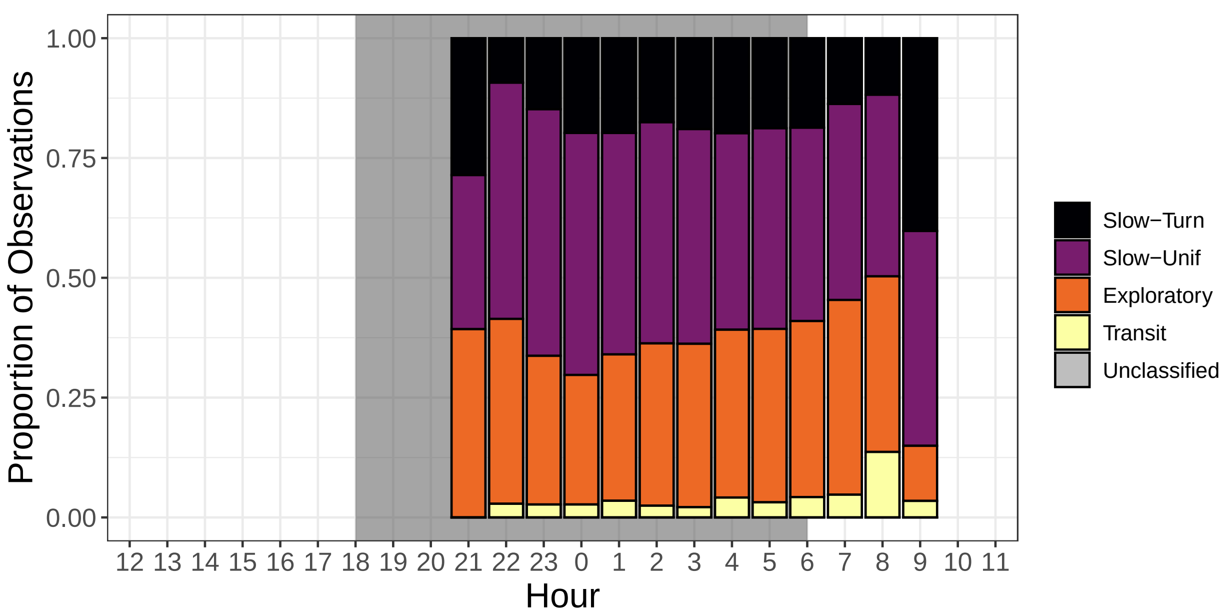


Figure 15: Summary of behavioral states by ID, season, and hour of day.

Additionally, I summarized the proportion of behavioral states associated with each land class. Distributions of continuous variables were analyzed for each state as well. There did not appear to be any major differences in the proportion of behavioral states by land class, but Slow-Uniform appeared to be the most prevalent across all land classes. There generally did not appear to be any clear patterns in the distributions of the remaining environmental covariates across behavioral states. However, the Transit state did appear to be associated with slightly greater values of vegetation productivity (NDVI, Greenness), water presence/moisture (NDWI, Wetness), and a higher elevation.

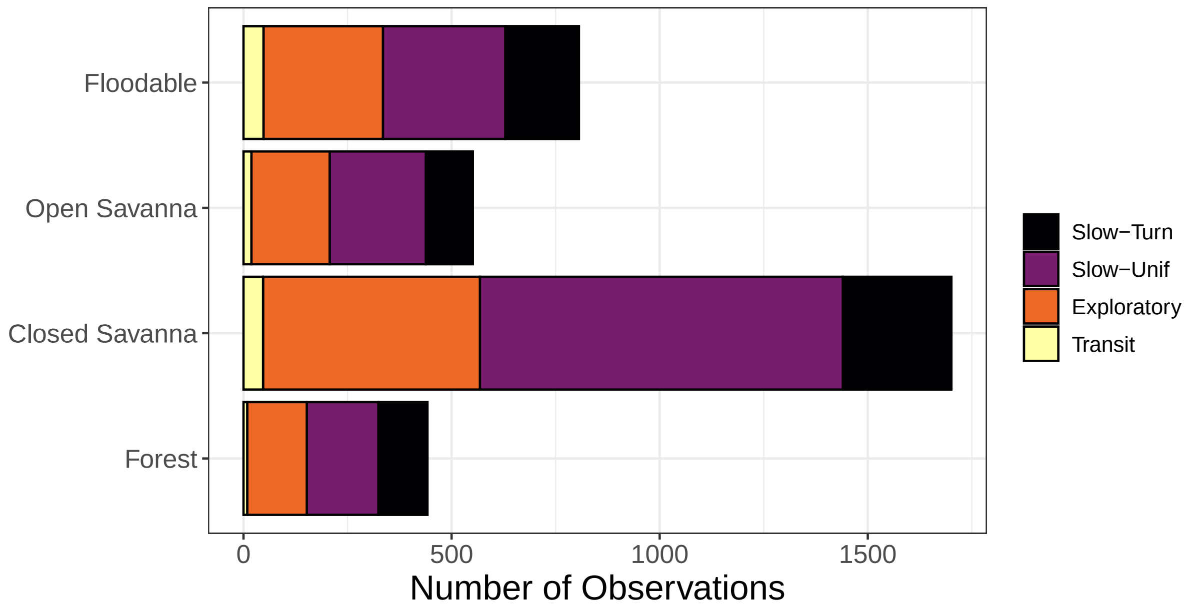


Figure 16: Summaries of environmental covariates per behavioral state.

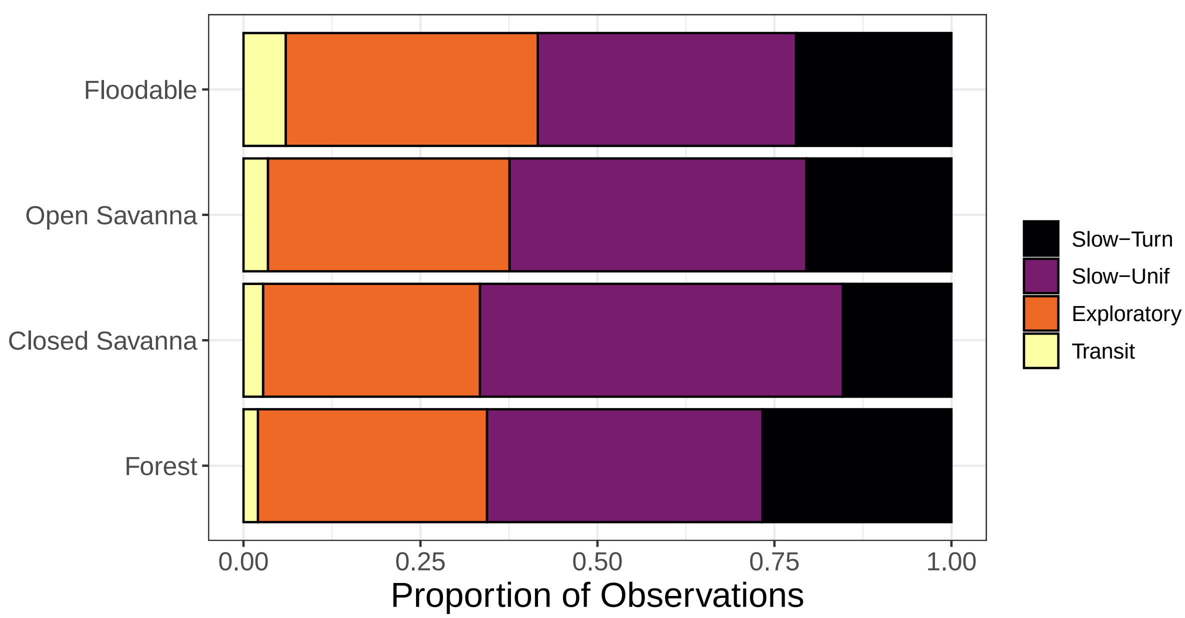


Figure 17: Summaries of environmental covariates per behavioral state.

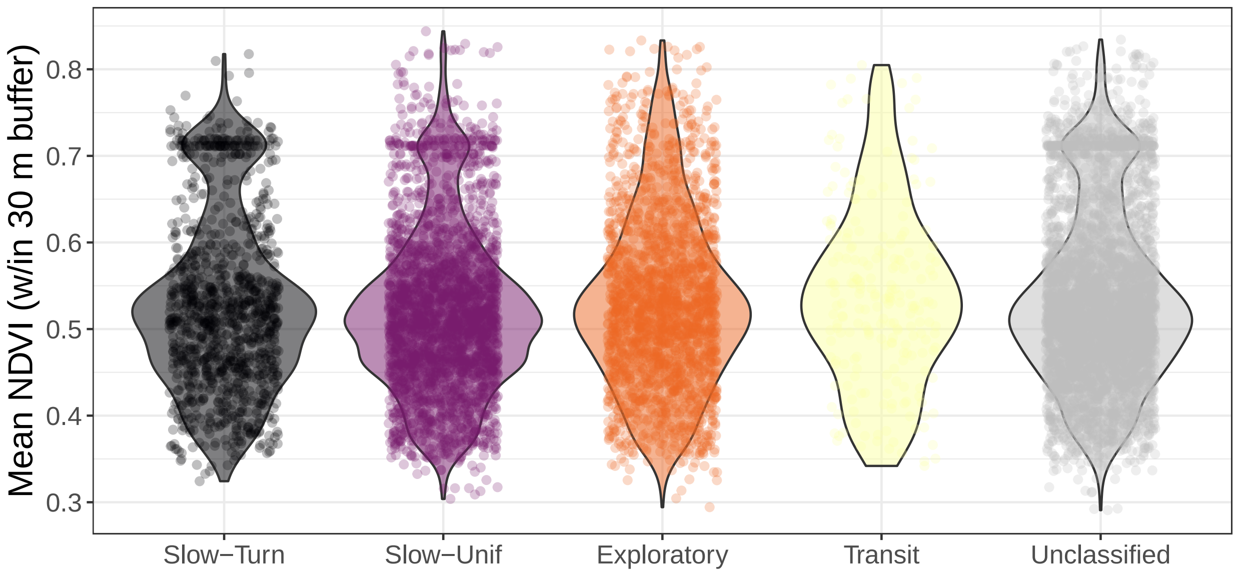


Figure 18: Summaries of environmental covariates per behavioral state.

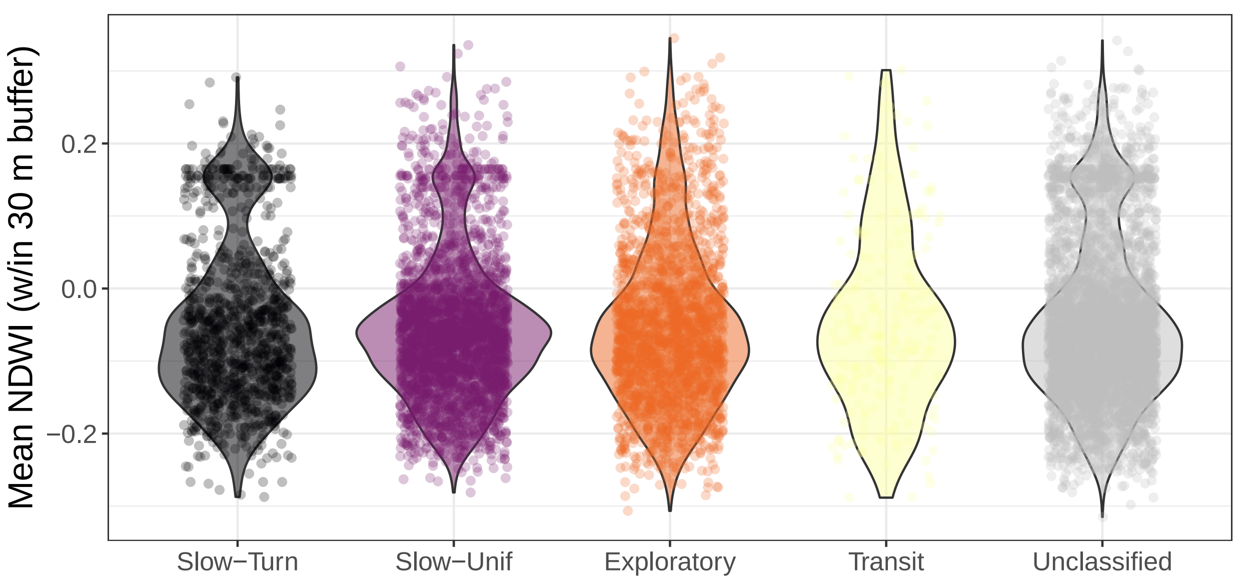


Figure 19: Summaries of environmental covariates per behavioral state.

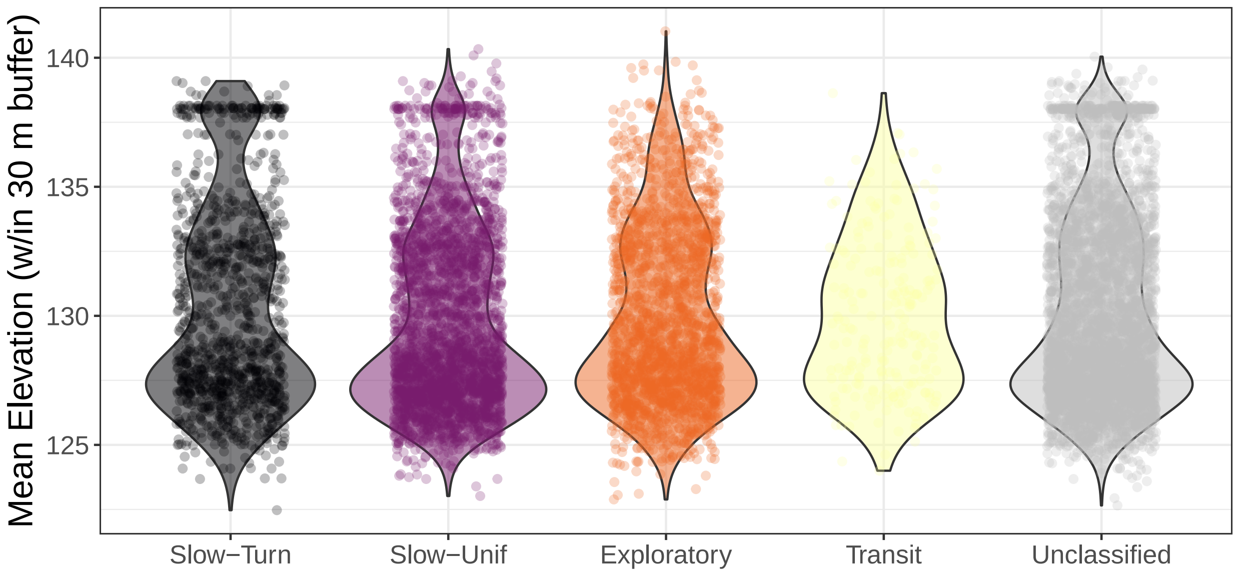


Figure 20: Summaries of environmental covariates per behavioral state.

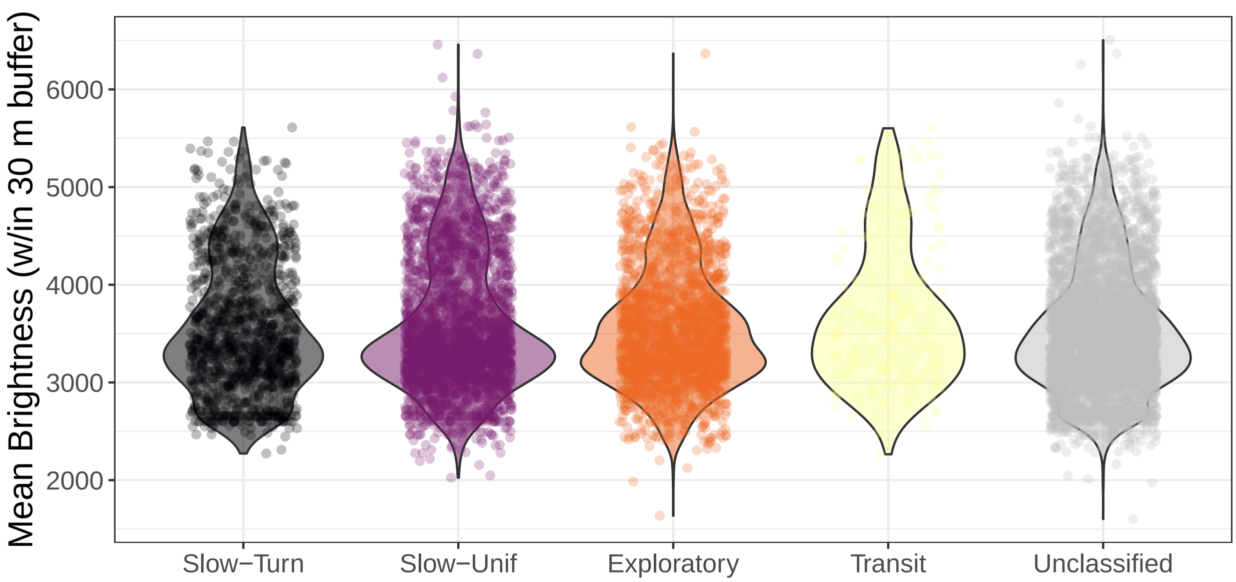


Figure 21: Summaries of environmental covariates per behavioral state.

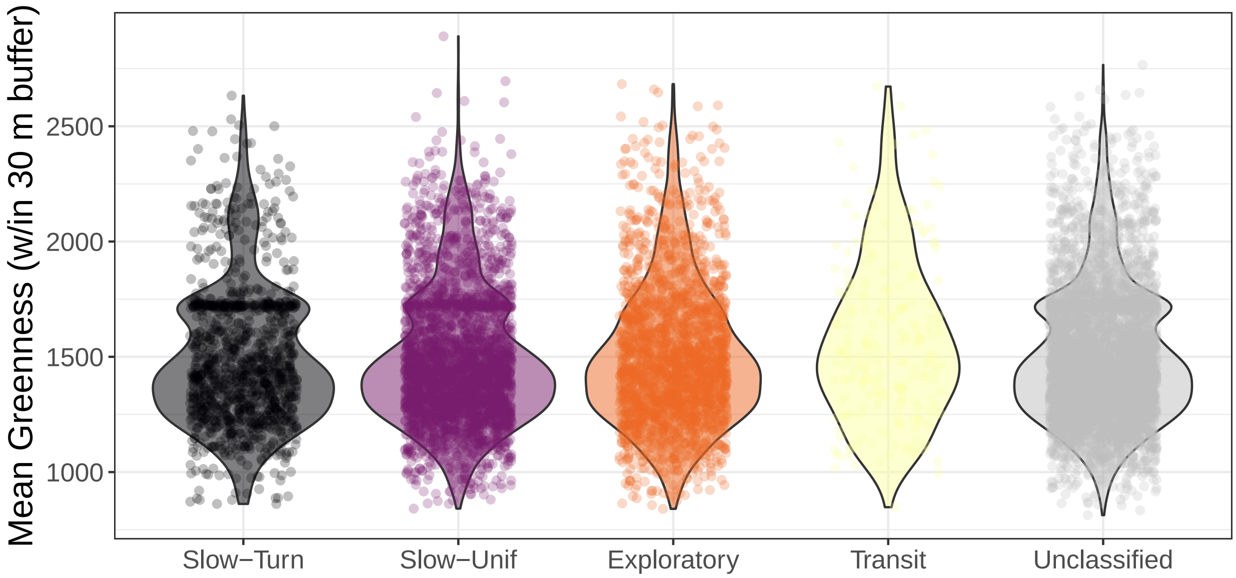


Figure 22: Summaries of environmental covariates per behavioral state.

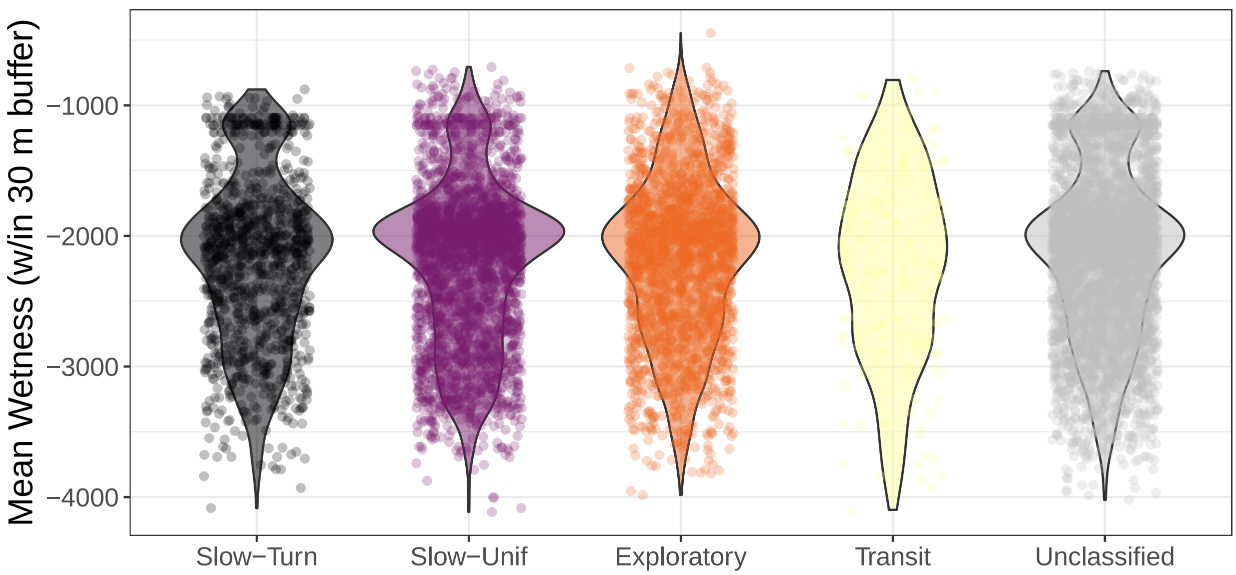


Figure 23: Summaries of environmental covariates per behavioral state.

We can also take a look at the 50 and 95% kernel density isopleths for each state to get a sense of what types of land cover might be used the most during different behaviors within a spatial context. Based on a simple visual inspection of this plot, it appears that a greater amount of Floodable habitat falls within the 95% isopleth for the Transit state while the remaining states primarily overlap with Closed and Open Savanna. This corresponds with the previous figures that were shown.

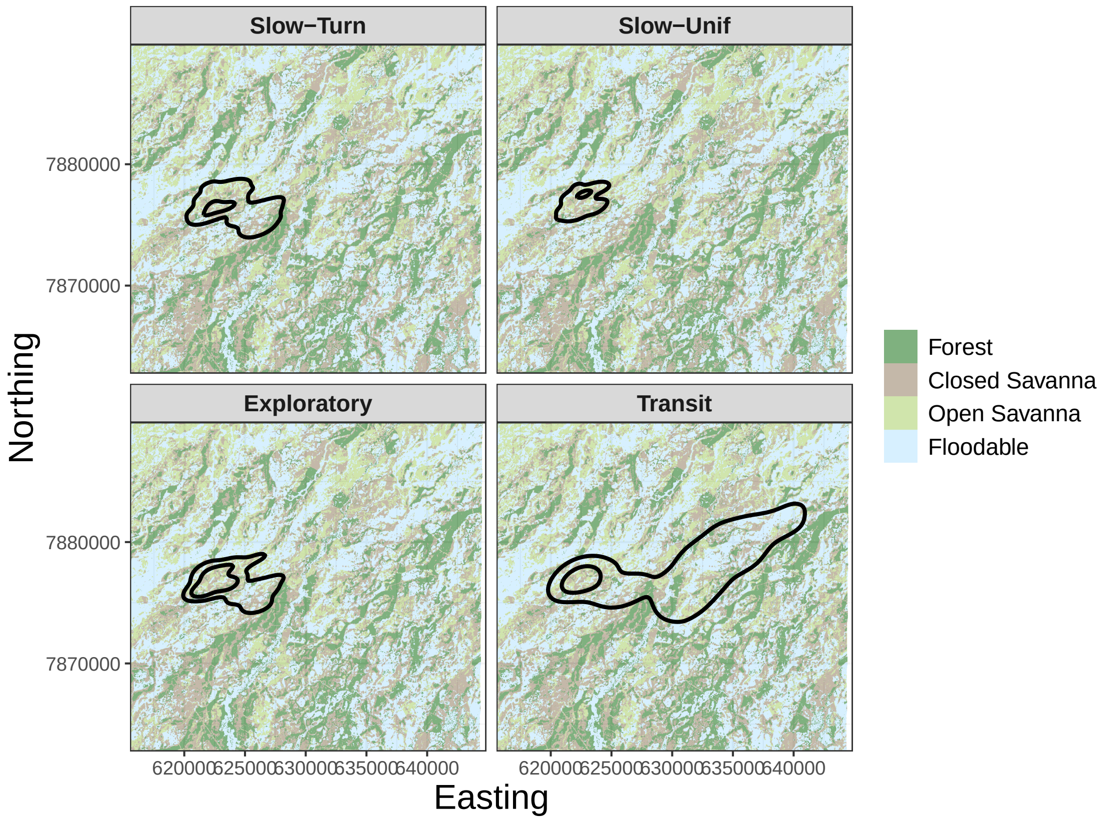


Figure 24: The 50 and 95 percent isopleths are shown from a kernel density estimate of locations for each state across all IDs.

Comparisons of state-habitat relationships

Multiple pairwise logistic regressions were performed to make pairwise comparisons of environmental covariates on the probability of being in one state or another. This resulted in a total of six comparisons. As mentioned previously, only greenness and wetness were included as covariates in addition to animal ID.

Slow-Turn vs Slow-Uniform

In this comparison, Slow-Uniform behavior was more likely to occur in habitats with greater vegetation productivity (greenness), but the Slow-Turn state was more likely to occur at wetter locations. The 95% confidence interval for the wetness coefficient only barely included the value of 1.

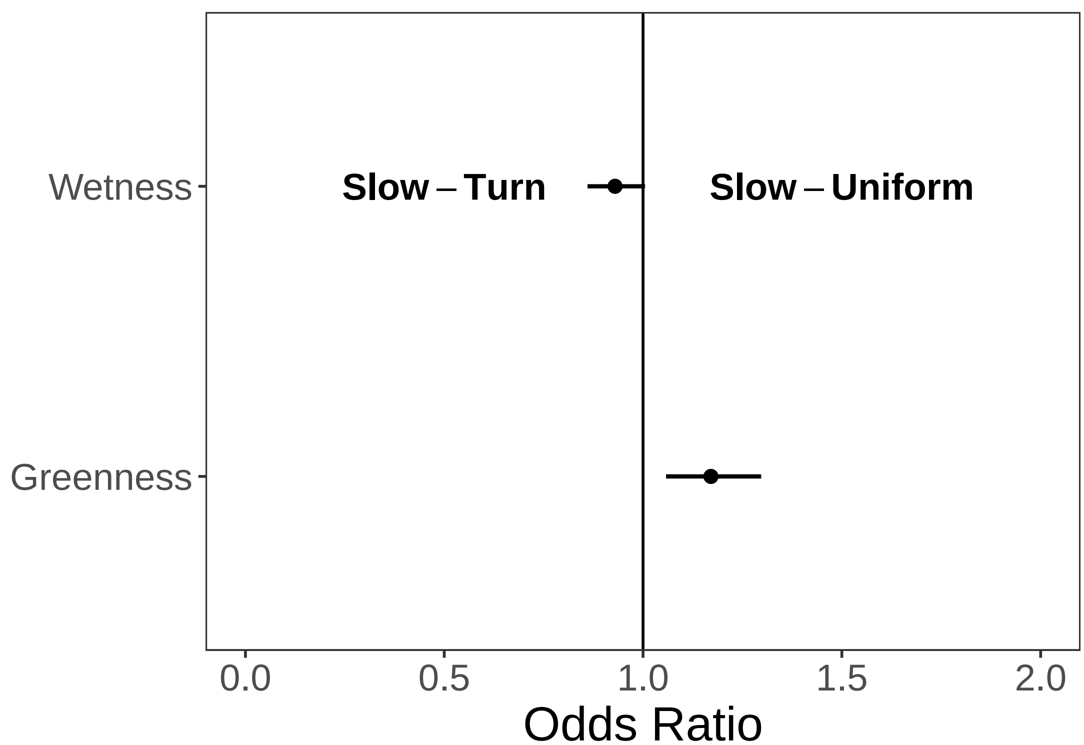


Figure 25: Odds ratios for each of the environmental covariates where values greater than 1 favor a Slow-Uniform state.

Slow-Turn vs Exploratory

Somewhat similar to the previous comparison, increasing values in greenness were associated with a greater probability of exhibiting an Exploratory state, while wetter areas did not significantly impact the probability of exhibiting either state.

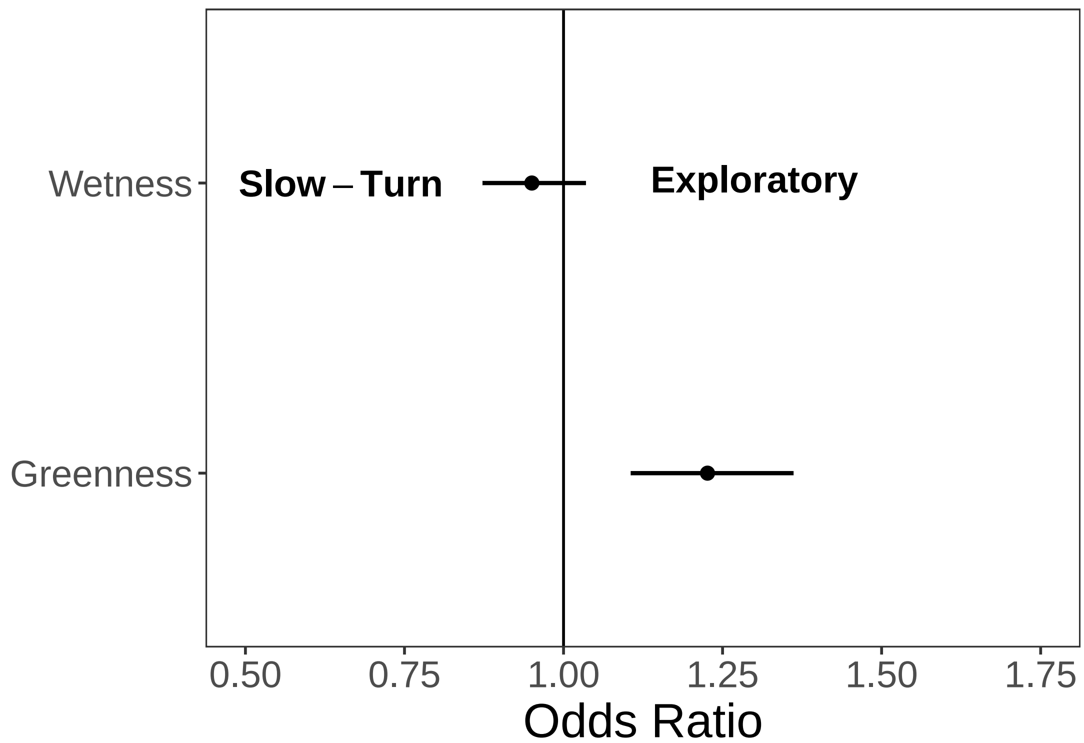


Figure 26: Odds ratios for each of the environmental covariates where values greater than 1 favor an Exploratory state.

Slow-Turn vs Transit

Nearly identical to the previous comparison, Transit was more likely to occur with greater levels of vegetation productivity whereas wetness did not appear to have a significant effect on the probability of either state.

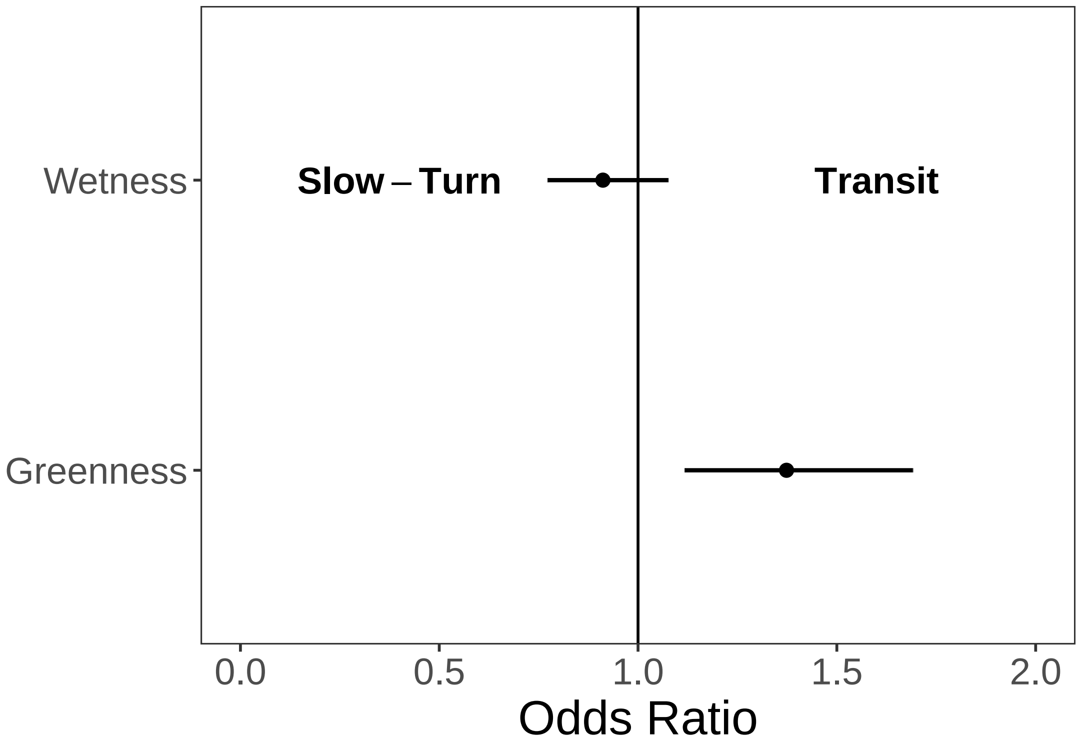


Figure 27: Odds ratios for each of the environmental covariates where values greater than 1 favor a Transit state.

Slow-Uniform vs Exploratory

In this comparison, neither of the covariates significantly impacted the probability of exhibiting either state. This relationship differs largely from that observed when comparing against the Slow-Turn state.

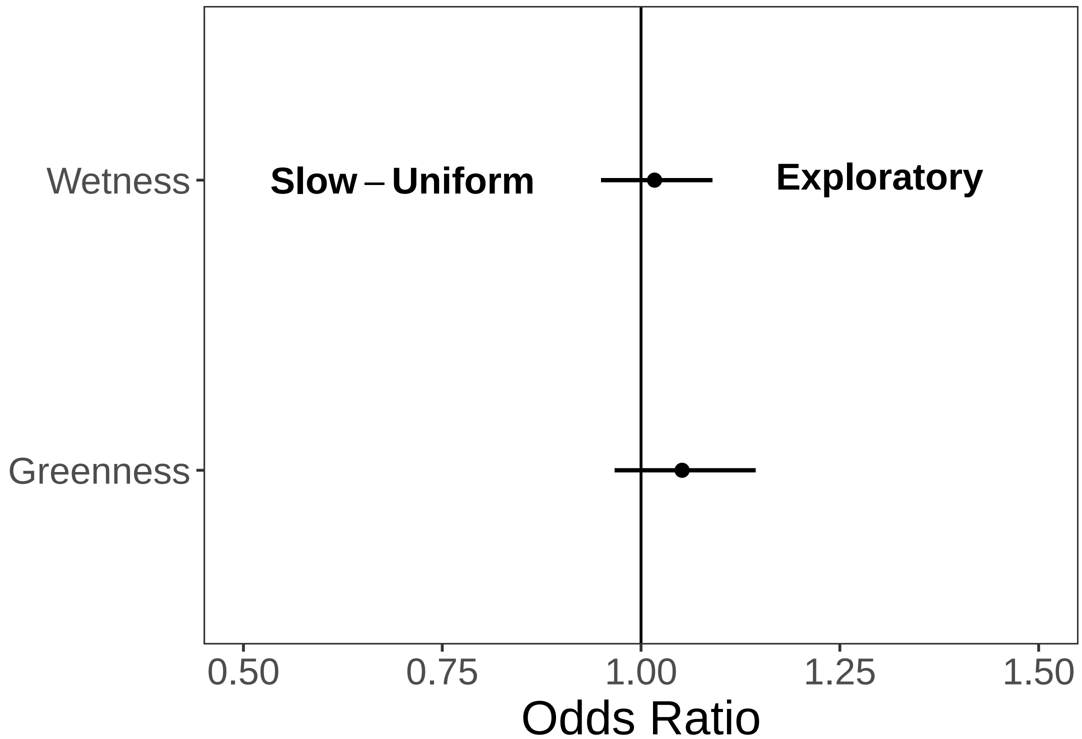


Figure 28: Odds ratios for each of the environmental covariates where values greater than 1 favor an Exploratory state.

Slow-Uniform vs Transit

Nearly identical to the previous comparison, neither of the covariates significantly impacted the probability of exhibiting either the Slow-Uniform or Transit state.

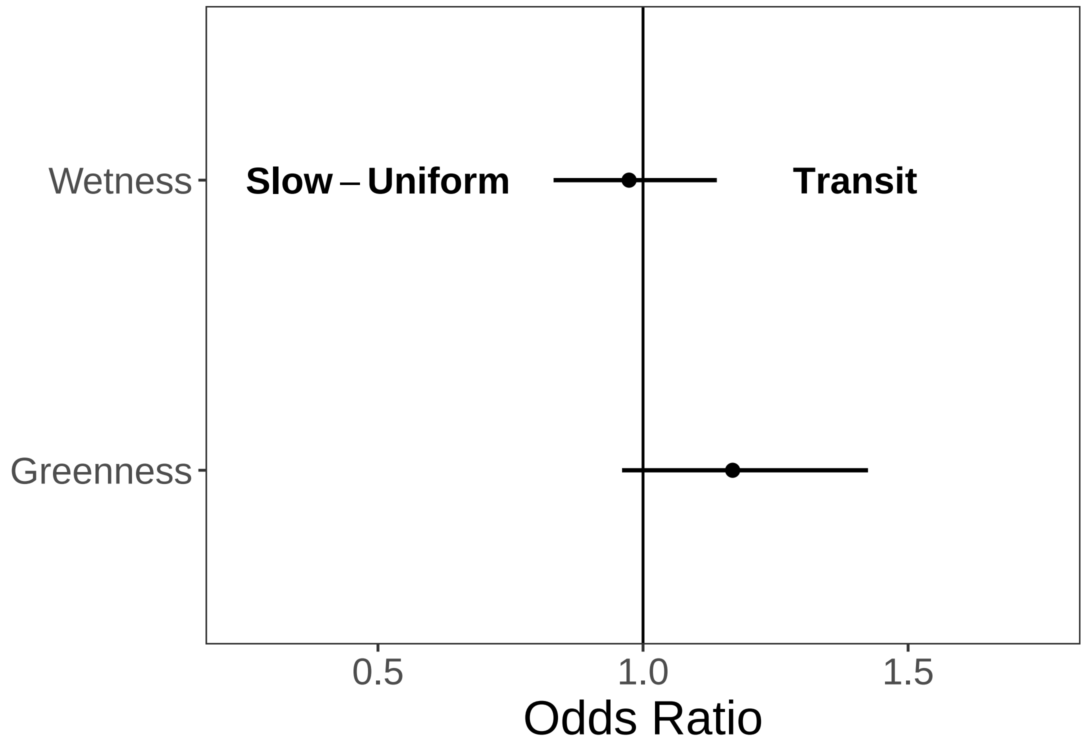


Figure 29: Odds ratios for each of the environmental covariates where values greater than 1 favor a Transit state.

Exploratory vs Transit

Similar to the previous two comparisons, neither of the covariates significantly impacted the probability of exhibiting either state.

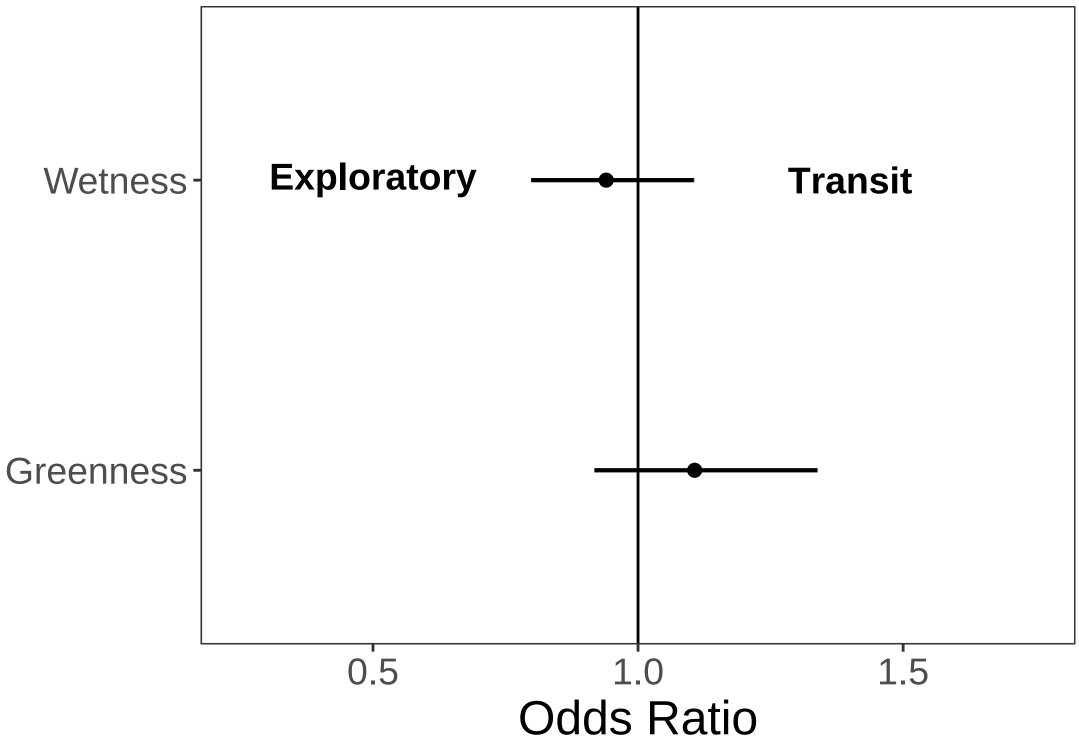


Figure 30: Odds ratios for each of the environmental covariates where values greater than 1 favor a Transit state.

Discussion

Based on our analysis that used a non-parametric Bayesian mixture model to estimate the number of optimal states and their distributions for each data stream (i.e., activity counts, speed, turning angles), four *active* states appeared optimal for these data. Two states had relatively low activity counts and low speed, but differed in their distributions of turning angles. The state with turning angles near $-\pi/\pi$ radians was labeled "Slow-Turn" while the other had a uniform distribution of turning angles and was labeled "Slow-Uniform". These state names primarily serve as a place-holder for more descriptive labels. The third state had a higher level of activity counts and higher speeds with turning angles slightly concentrated around 0 radians, although only slightly. This state was labeled "Exploratory" and may represent an intermediate activity state while armadillos are searching for food, shelter, or conspecifics. The fourth state displayed the greatest activity counts, highest speeds, and turning angles highly concentrated near 0 radians. These states were then assigned to the 62% of observations where certainty of belonging to a single state was high ($\geq 75\%$). When there was greater uncertainty as to which state an observation belonged, it was labeled as "Unclassified". These state estimates were used in the subsequent analyses that compared the effects of elevation, greenness, and wetness on exhibiting one state over another.

When evaluating general trends in the proportions of behavioral states, there was little variation in the frequency of the states that were exhibited across individuals, seasons, and by hour of the day. In general, the "Exploratory" and "Slow-Uniform" states were most common, followed by "Slow-Turn" and then "Transit". This seems to make sense given that the "Exploratory" state may be related to searching for sites to forage, whereas "Transit" would typically occur when moving quickly over long distances between patches of habitat. It is currently unclear what behaviors are represented by the slower behavioral states. Additionally, relationships with environmental covariates did not appear to be easily discernible.

When assessing the effects of environmental covariates on the probability of exhibiting one state over another, there was typically a consistent pattern when comparing the Slow-Turn state against the others. In each of those three logistic regressions, the Slow-Turn state was favored when habitat was wetter, whereas the faster state was favored with higher values of greenness. The comparisons among the other three states (Slow-Uniform, Exploratory, Transit) generally did not show any significant differences.

While there do appear to be potential ecological patterns that emerge from comparisons among states, it is currently difficult to discriminate among states for the given set of data and analyses. It is possible that not enough behavioral variation was captured by these specific tracks since there were not distinct values of environmental covariates that characterized each of the different states. However, further analyses with a larger sample size may be able to discriminate these behaviors further. Additionally, these states may be used for additional analyses, such as to assess habitat selection for specific behavioral states.