Final Project

Anthony Gallante

Northwestern University School of Professional Studies
MSDS 422-DL

I. Executive summary

Hidden Markov Models (HMM) have been used by ecologists to infer animal behavior from spatiotemporal data on several occasions. In this project, I acknowledge the difficulty that arises when time-series covariates are included in the Hidden Markov process and given the ability to influence inferred hidden state transitions in the Viterbi algorithm. Using the published paper "An application of upscaled optimal foraging theory using hidden Markov modelling: year-round behavioral variation in a large arctic herbivore" as a point of comparison, I aim to reproduce the conclusions of Beumer et al. through an alternative models and incorporate movement prediction capabilities using a Recurrent Neural Network (RNN). I believe that this approach may be an acceptable modern approach to behavioral inference due to an RNN's ability to handle sequential data.

II. Problem statement and research objectives

Hidden Markov Models (HMM) are an unsupervised machine learning model which use Markov Chain properties to describe a sequence of observed datapoints through the presence of, and transitions between, unknown hidden states. The sequence of datapoints, usually discrete time-series data, are treated as emissions from these states as the system transitions from one hidden state to another. When the system transitions to a new state, a single observation from a finite set of possibilities is created with an associated probability. While higher order HMMs are possible, a traditional HMM is bound by the assumption that the next state transition is *only* determined by the system's current state and its associated transition probabilities; prior emissions and past state transition sequences are not considered. While HMMs can be used to answer several questions, the Viterbi algorithm can produce the most-likely sequence of hidden state transitions that describes a set of observed datapoints. These hidden states are not defined by the user and must be interpreted after the HMM is trained.

Although it may be convenient to treat it as such, it is my opinion that animal behavior, even when reduced to a small set of hidden states, does not reflect a true Markov chain process. Beumer et al. partially address this problem by considering the effects of several spatial and temporal covariates. Using their method, the fundamental Markov property that assumes a constant transition matrix can be updated to support a dynamic transition matrix that changes with covariate values at the current time-step.

Recurrent Neural Networks (RNN) are capable of all the same analysis as HMMs, with the added pattern recognition capabilities of neural networks and the clarity that come along with supervised learning techniques. Perhaps most favorably in this context, covariates and emissions do not have to be treated separately—muskoxen data can be fed into a model as a time-series feature set.

The objective of this assignment is to attempt to recreate the findings of Beumer et al. using a Recurrent Neural Network with observable and covariate variables included in the same feature space and supervised learning techniques. I will be using the dataset of northeastern Greenland female muskoxen used in the study.

The objectives for this project have been captured in the following hypotheses, which have been provided in the midpoint review assignment. As my knowledge on the topic grew, I started to realize that some of these hypotheses were poorly formed and are rather irrelevant. I will, however, expand on them nevertheless in section *VI*. The purpose of this assignment is to learn—and there is no better way to learn than to address mistakes head-on.

- **Hypothesis 1:** *A Recurrent Neural Network can be used to recreate the results from the Beumer et al. paper, which used Hidden Markov Models.*
- **Hypothesis 2:** *In order to replicate the results from the Beumer et al. paper using a Recurrent Neural Network, the step length and turning angle features will need to be weighed differently from the other features.*

- **Hypothesis 3:** *A Recurrent Neural Network will yield more consistent results between the summer and winter datasets than a Hidden Markov Model because states are defined prior to the model fitting process.*
- **Hypothesis 4:** *Shapley values for the Recurrent Neural Network, indicating feature importance, will be consistent with important features from the Beumer et al. Hidden Markov Model.*

III. Exploratory data analysis

I first made an initial effort to understand the three states used by Beumer et al. in their paper. I will refer to states 1, 2 and 3 as "resting", "foraging", and "relocating" respectively, and have attempted to match the colors of my exploratory analysis to match those used in the study.

The first assumption we must make is that Beumer et al. intended to study the behavioral states returned by the HMM. These three states are defined by step length, the average distance the muskox travelled in one hour, and turning angle, the average change in a muskox's bearing over the same hour. State characterizations are summarized below, while scatterplots depicting each datapoint's inferred hidden state are shown in Figure 1.

- State 1, or resting, is characterized by short step lengths and mild undirected movement.
- State 2, or foraging, is characterized by moderate step length and high undirected movement.
- State 3, or relocating, is characterized by moderate to long step lengths with directed movement.

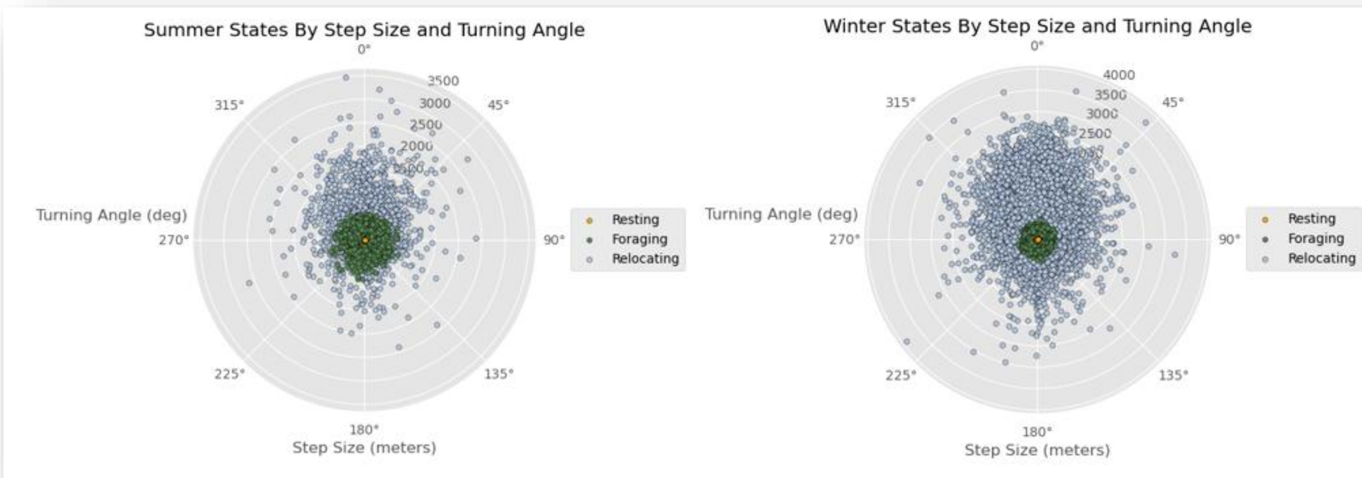

Figure 1: Inferred Hidden States of Female Muskoxen in the Summer and Winter Months

We immediately notice that while states 1, 2, and 3 are somewhat similar between the summer and winter months, the step length thresholds that differentiate between states 2 and 3 are considerably different. Because of the unsupervised nature of HMM state inference, each seasonal model had the ability to fine tune their states: it is only after the model is trained that one can interpret a state as "foraging." For example, the median step length for the summer HMM is 96.17m while the median step length in the winter HMM is 50.42m. For this reason, we cannot directly compare state 1 of the summer model with state 1 of the winter model. In the case of states 1 and 3, the differences between the summer and winter HMMs are less extreme but are still apparent.

By labelling the dataset with step length and turning angle behaviors prior to the fitting process, we can eliminate the need to interpret states by ensuring that a datapoint is classified based on a more rigorous definition of resting, foraging, and relocation. This will be one of the primary objectives of the project.

In Figure 2 below, I attempt to identify obvious trends in the female muskoxen behavior by plotting the location of each datapoint and displaying them side-by-side, separated by their hidden state. As one would expect, the datapoints classified in the resting and foraging states tend to form clusters, while datapoints connecting these clusters are almost entirely made up datapoints in the relocating state.

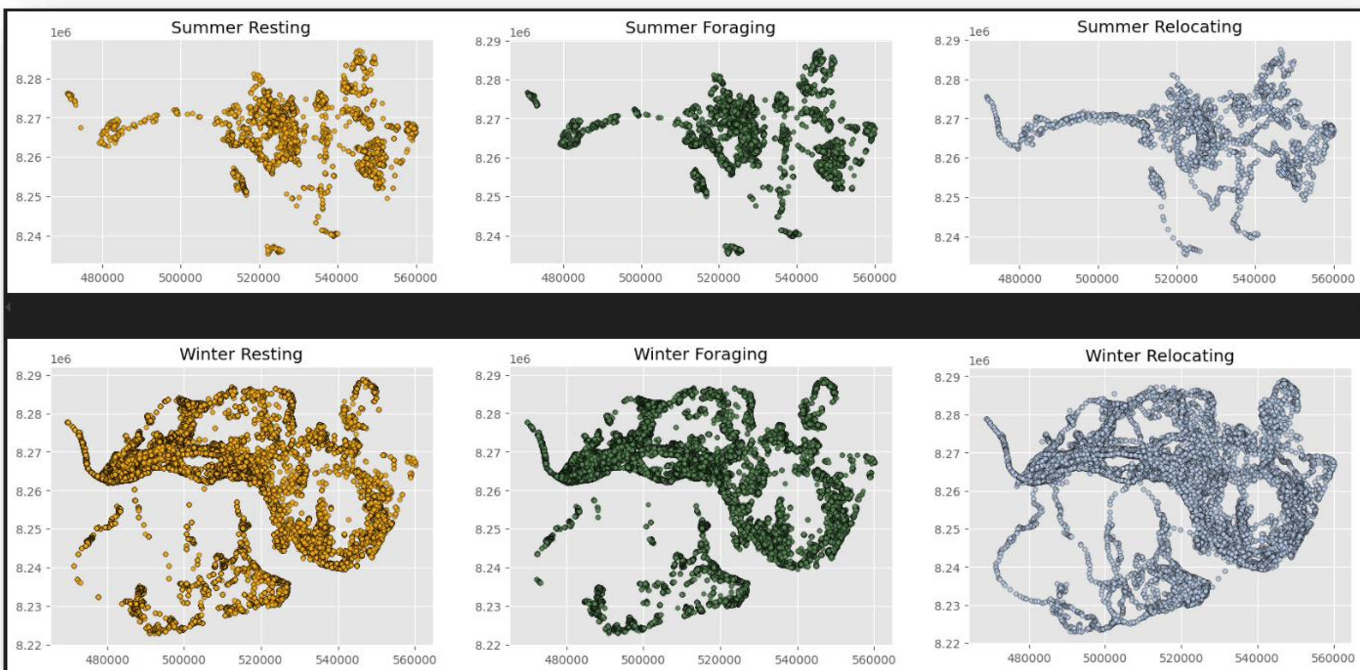

Figure 2: UTM Northing (y-axis) and UTM Easting (x-axis) locations of each hidden state inference.

In Figure 3, I move one step further, breaking up the images from Figure 2 by “burst,” a segmentation technique used by the researchers to separate sequences of data and maintain consistently spaced time measurements. Each color represents a different burst, which we can use to get an idea of how the muskoxen rest, forage, and relocate on smaller time scales.

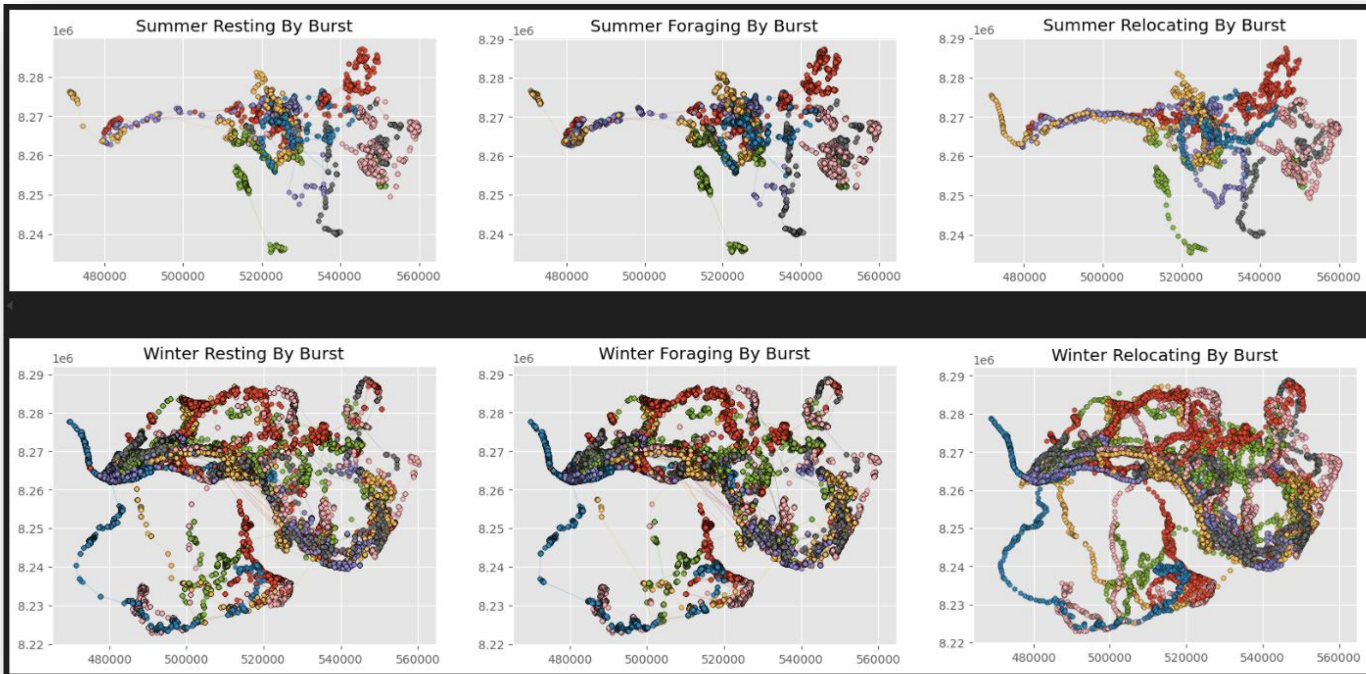

Figure 3: UTM Northing (y-axis) and UTM Easting (x-axis) locations of each state, colored by burst.

I decided that keeping track of important covariates would also be important. Because a Hidden Markov Model treats covariate features differently from observed features, it might be beneficial to know which covariates affected the model the most. Recurrent Neural Networks do not typically treat features this way; it may be worth investigating applying higher initial weights to the step length and turning angle variables to recreate this technique. Without access to Beumer et al.'s HMM parameters or coefficients to the equations that govern transition matrix behavior, I decided to create a proxy model using a random forest classifier to predict state transitions using covariates alone. I then used the SHAP Python module to calculate Shapley values of the features in that random forest classifier model.

While I take the results from this method with a grain of salt, we see that time of day may influence muskoxen behavior more than the landcover type, for example. This is at least intuitive and gives us an idea that there may not be a single covariate that is order of magnitudes more important than others.

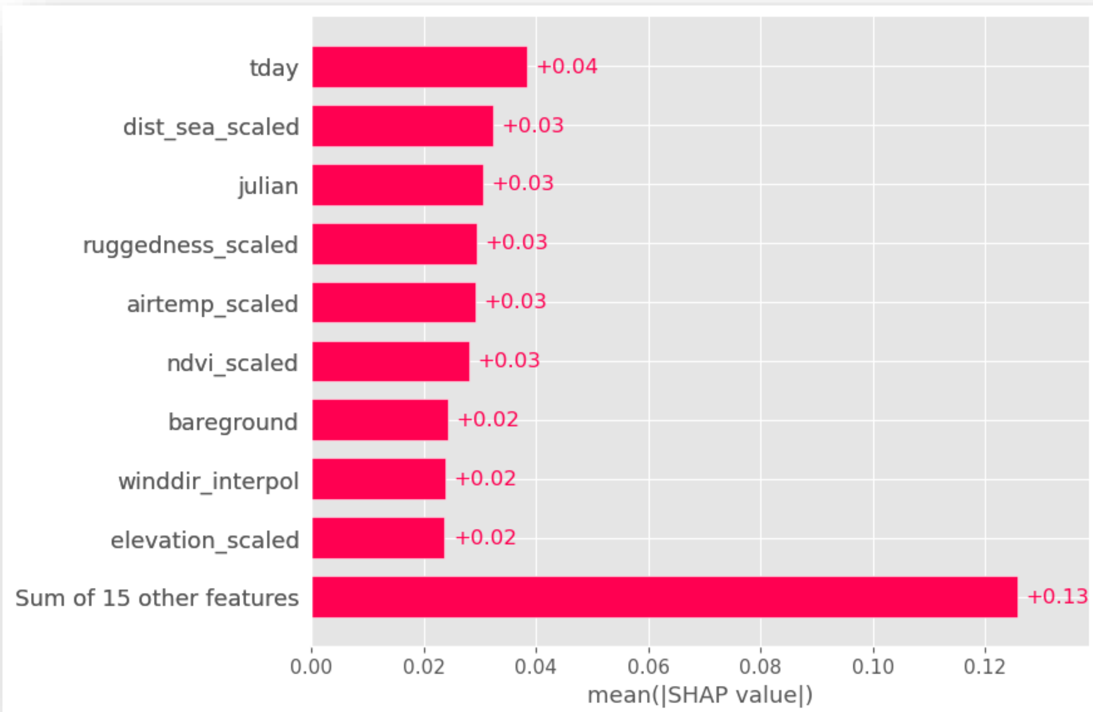

Figure 4: Covariate importance Shapley values as determined by a random forest proxy model.

IV. DATA PREPARATION AND FEATURE ENGINEERING

Because replication is the primary objective of this project, very little original data preparation and feature engineering is likely to be necessary. The dataset used in the original study is publicly available at the author's Zenodo repository.

An effort to label the data manually was made to better understand the results of the HMM used by Beumer et al. Through this effort, I found that hidden muskoxen states were defined almost exclusively through step size, the Euclidean distance travelled each hour (measured in meters). Additionally, the unsupervised nature of HMMs introduced different definitions of the foraging and relocating states, making them difficult to compare directly. Instead, I trained an artificial neural network to predict a state based on a given definition, which could be used across both seasons. To match the researchers definitions of resting, foraging, and relocating, I only needed to use the “step” data series, which could be labelled simply by applying the logic shown in Table 1; however, I took this chance to generalize the process if only a small portion of datapoints could be labelled manually.

Table 1: State definitions for cross-season comparisons

State	Definition
Resting	Step Size < 25 meters
Foraging	25 < Step Size < 100
Relocating	Step Size > 100 meters

V. Methodology and Tools Used

Due to the small files and simplicity of the models involved, all computation was completed locally on my personal computer hardware using an Anaconda distribution of Python version 3.9.18 through the Visual Studio Code (VS Code) software. All code was typed by hand, however a number of generative AI products were referenced for troubleshooting and general guidance, including Anthropic’s Claude and Microsoft’s VS Code Copilot.

While many open source and publicly available libraries were used, TensorFlow and Scikit-Learn were critical for model development. TensorFlow was the primary library used for neural network development—both the feed-forward neural network used for classification and the recurrent neural network and long-short-term-memory (LSTM) structures—while Scikit-Learn was used for traditional machine learning models.

A total of five machine learning models were made during this project: a random forest, used to identify potentially important covariate features from the Beumer et al.'s hidden Markov model, a feed-forward artificial neural network used to classify manually labelled datapoints, and three recurrent neural networks using LSTM units were used for track prediction.

Multiple RNNs were created for movement prediction purposes through a keras_tuner random hyperparameter search. Because the models were predicting locations, mean squared error was used as the loss function, which was minimized during training. A summary of the best performing model structure and parameters can be found below in Table 2.

Table 1: Best RNN hyperparameter search values

Layer	Parameter	Value
Input Layer	Number of Units	28
LSTM	Number of Units	288
Dropout	Dropout Rate	0.2
Dense	Number of Units	82
Dropout	Dropout Rate	0.2
Dense	Number of Units	140
Reshape	Shape	(5, 28)

VI. Findings and Conclusions

In this section, I address each of the hypotheses formed in section II.

Hypothesis 1: *A Recurrent Neural Network can be used to recreate the results from the Beumer et al. paper, which used Hidden Markov Models.*

While this statement is technically correct, it introduces my personal misunderstanding of the purpose of recurrent neural networks. They do handle time series data, but not in the same way as a hidden Markov model. Since the midpoint project delivery, I have found that it might be more valuable to implement an example of an RNN in an appropriate context: behavior prediction. This is certainly possible with HMMs, though it does not seem to be an objective of the researchers in the original paper. Thus, my first deviation from my prior hypotheses. Figure 5 below shows examples of predicted muskoxen movement paths using the prior 5 elements of the track as predictors.

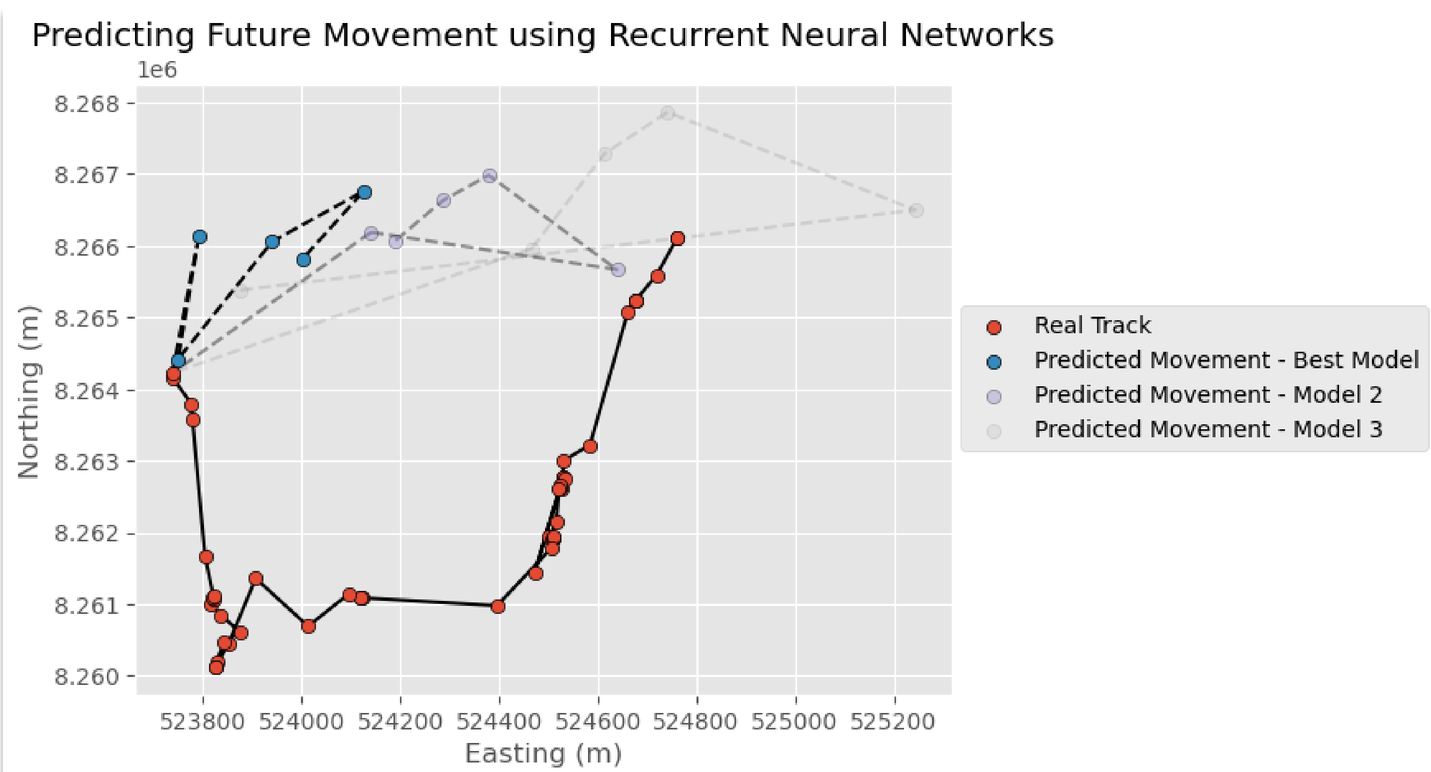

Figure 5: Predicted muskoxen movement pattern using an RNN

Hypothesis 2: *In order to replicate the results from the Beumer et al. paper using a Recurrent Neural Network, the step length and turning angle features will need to be weighed differently from the other features.*

My second hypothesis turned out to be incorrect. Manual weights do not need to be assigned to more important features of a neural network, however, I did not use a recurrent neural network to predict or replicate Beumer et al.'s paper. Instead, I used a 7-layer artificial neural network of nearly alternating densely connected layers and dropout layers to prevent overfitting. During the exploratory data analysis phase, it became clear that while both muskoxen step-size and turning-angles were used in the Beumer et al.'s study to assign hidden states to the dataset, step-size appeared to be much more important to the interpretation of the states. Using the summer muskoxen dataset as an example, we see that the resting, foraging, and relocating states can be identified as nearly non-overlapping states determined by step-size (Figure 6).

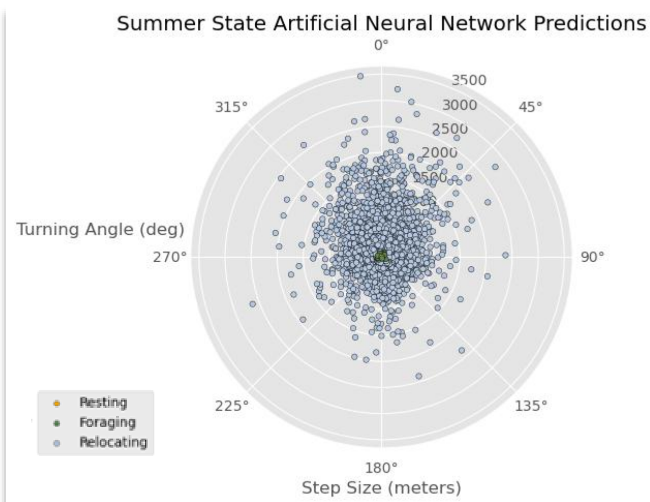

Figure 6: State predictions using an artificial neural network

Hypothesis 3: *A Recurrent Neural Network will yield more consistent results between the summer and winter datasets than a Hidden Markov Model because states are defined prior to the model fitting process.*

Hypothesis 3 is not applicable, due to the conscious decision to deviate from the original hypotheses. The artificial neural network used to address Hypothesis 2 was trained on a small set of data labelled according to Table 1.

Hypothesis 4: *Shapley values for the Recurrent Neural Network, indicating feature importance, will be consistent with important features from the Beumer et al. Hidden Markov Model.*

Hypothesis 4, like Hypothesis 3, is also no longer appropriate. The artificial neural network used to classify features does not consider the covariate features.

VII. Lessons Learned and Recommendations

This was an exciting exercise. I was able to learn from the many mistakes I made during this project. Model selection is much less arbitrary than I had originally believed, though I never thought that it was not important. I would be interested in applying transfer learning methods to this scenario. Primarily, I learned that recurrent neural networks excel at sequence prediction while HMMs are better suited for unsupervised state detection. Perhaps a hybrid approach could be applied to future ecological movement studies.

Applying pre-trained models to a geospatial temporal dataset will likely elicit exciting insights in a fraction of the time it took to perform parameter searches or fit a hidden Markov model. I am also interested in a form of transfer learning in which a model is trained on a higher fidelity dataset and used to identify movement features (turning, slowing down, speeding up, etc.) in a dataset with a lower sampling rate.

VIII. References

Beumer, L.T., Pohle, J., Schmidt, N.M. *et al.* An application of upscaled optimal foraging theory using hidden Markov modelling: year-round behavioural variation in a large arctic herbivore. *Mov Ecol* **8**, 25 (2020). https://doi.org/10.1186/s40462-020-00213-x

Claude 3.7 Sonnet (Anthropic, 2025)

Microsoft Copilot (Microsoft, 2023)

Appendix

All code and data used in this project will be made publicly available on my personal GitHub repository at: https://github.com/AnthonyGallante/MSDS_422_Project.