Modeling Stochastic Processes in Ecology

Doctoral Qualifying Exam by Fabiana Ferracina

May 5, 2021

Abstract

Advances in data collection technology and computational hardware have led to increased interest and ability to predict ecological processes, thus creating the need for sophisticated models that can account for uncertainties that exist in both the data collection and the estimation of parameters. The concern of statistical ecology is with the development of such models in order to understand nature's stochastic processes despite uncertainties. The goal of understanding ecological processes such as species distribution, animal behavior, the spread of forest fires and climate events is not only important for humanity's continued survival and well-being, but approaches developed in the pursuit of this goal can lead to novel statistical discoveries with applications beyond the field of ecology.

From general linear models, to Bayesian approaches, to stochastic processes such as hidden Markov models, statisticians and quantitative ecologists have provided powerful and effective tools to study many important ecological problems. The focus of this paper will be on models that study data generated by point processes. Species distribution from occurrence and detection datasets can be modeled as an inhomogeneous Poisson process, while the movements of an animals through space and time can be modeled as a Markov process. These stochastic processes both depend on parameters we wish to estimate, leading to further need of statistical thinking and creativity in order to find efficient and intuitive parameter estimation approaches.

1 Introduction

"We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis,

> it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes."
>
> – Pierre Simon Laplace, A Philosophical Essay on Probabilities

Sadly, even if such intellect could exist in our universe, as pointed out by Laplace in the same essay, it would remain infinitely removed from the human mind. However, the study of stochastic processes, which aims at trying to figure out the state of those items through time, has not only been close to the human mind but also very much close to the human heart. Be it from need or curiosity, humans throughout their history have pondered what the state of a natural process will look like in the future. Natural processes such as the distribution of animals through space and time, weather events or the spread of fire follow non-deterministic paths, however their dependence on time make such processes predictable to those with access to time series data and statistical expertise (Esling and Agon, 2012).

Advances in data collection technology and the continued increase in computational power over time, although helpful, are not sufficient to answer questions of complex ecological systems. Models that are too simple can ignore individual and environmental variability, species interactions and transient dynamics (i.e. short term and/or nonequilibrium dynamics), rendering predictions unrealistic (Green et al., 2005). On the other hand, models that attempt to capture as much of such complexities as mathematically possible become computationally intractable (Pascual, 2005).

Gimenez et al. (2014) in their recent paper, performed an analysis of the current research trends in statistical ecology and found a shift from sampling design issues to predictive studies which aim at explaining the process underlying the ecological patterns under study. They found the top questions being asked are of species distribution, quantifying biodiversity, population dynamics, animal movements and the interpretation of citizen science data.

These topics are connected to each other both ecologically (e.g. biodiversity affects population dynamics; both species distribution models and citizen science data deal with imperfect detection issues; etc) and statistically (e.g. distribution and movement are stochastic; for each topic hierarchical frameworks can reveal important inherent structures and separate different sources of error; etc), however specialization is necessary to develop appropriate models and understanding.

The goal of this paper is to explore natural processes from a stochastic lens on an ecological eye, and vice-versa. In section 2 the focus will be on the ecological topic of species distribution modeling to construct a hierarchical framework where the resulting statistics can be informed by process and data, while in section 3 the focus will be on the statistical topic of hidden Markov models to investigate different applications in ecology: the spread of forest fires and animal movement. Both of these sections address different parameter estimation techniques and the importance of hierarchy.

We will develop and apply tools from Statistics to answer questions such as: how are oceanic whitetip sharks distributed throughout their range? and how can we model imperfect detection of a species?, do forest fires spread like neuronal spikes? are the behaviors of young bald eagles distinguishable from those of older ones? Although one may find that the intellect of an all-knowing demon is not needed to answer these questions, they can still be very challenging to attack when one only has access to points in space, image pixels and locations through time, respectively.

2 Species Distribution Modeling

Modeling species distribution is a challenging problem, as empirical data is generally limited to species occurrence - the presence or absence of a member of a species at a given time and place of observation. These data are limited by observation error, which is often significant as both predator and prey are driven towards camouflage by evolutionary pressure (Stevens and Merilaita, 2011), and species members may simply not be present at the observation place and time but instead be nearby. Thus, we need to be more perceptive and creative when developing such models than simply using the data acquired. We need incorporate the biology and process which led to the state of our sample when we collected the data.

When members of species move through space and time, they do so for a specific purpose which can be challenging to study and identify. Some shark species prefer warmer temperatures (Strasburg, 1958), some bird species migrate to breeding grounds (Somveille et al., 2015), and either may return to a place we call their home (Hueter et al., 2005, Greenwood, 1980). Regardless of (although informed by) what temperatures they prefer, where they go to breed and whether or not they exhibit philopatric behaviors, a species can be thought of existing as points in space at a given time. Their presence in a particular space and

time follows a distribution informed by the process that got them there and the covariates that determined the space's suitability. An observer then encounters a biased subset of this population's distribution pattern, records and aggregates the data.

Traditionally, depending on how the data has been aggregated (presence-only, presence-absence or count), the investigator applies the appropriate well-tested and popular model for their data (Aarts et al., 2012, Yackulic et al., 2013). However, from the covariates that informed the process to the subset of population patterns observed, there is a hierarchical structure that is universal to all species distribution models. Understanding this hierarchy can help us better estimate the parameters that lead a species to occupy or not occupy a particular space in a particular time by accounting for the errors involved in each order of the hierarchical structure. Thus, studying species distribution models as a hierarchical framework can help us better understand a species distribution.

2.1 Hierarchical Species Distribution Framework

Statistically, the pattern created by points (i.e. the species) in space and time can be described by an inhomogeneous (or state-dependent) Poisson point process (Aarts et al., 2012, Hefley et al., 2013, Renner and Warton, 2013). The log-likelihood for this process is

$$\log L(\lambda; \mathbf{U}) = \sum_{i=1}^{n} \log \lambda(\mathbf{s}_i, t_i) - \int_{\mathcal{A}} \int_{0}^{T} \lambda(\mathbf{s}, t) dt d\mathbf{s} - \log(n!),$$

where n is the number of points or individuals, and the matrix $\mathbf{U}$ is a $n \times 3$ matrix with rows containing the 2D coordinates $\mathbf{s}_i$ (such as longitude and latitude) and the time t_i, $i = 1, \ldots, n$. $\mathcal{A}$ is the two dimensional area of study during the $[0, T]$ time interval. The function $\lambda(\mathbf{s}, t)$, called the intensity function, describes the expected abundance of points in a space-time volume.

For a particular location and time of interest we obtain the integrated intensity function $\bar{\lambda}$. It can be shown that the number of points u within the space-time volume of interest follows Poisson($\bar{\lambda}$) (Hefley and Hooten, 2016, Dorazio, 2014, Fithian and Hastie, 2013). Therefore, the presence or absence of a species in the space-time volume of interest is determined by the probability of $u > 0$. We can model this probability with the distribution Bernoulli$\left(1 - e^{-\bar{\lambda}}\right)$, since $P(u > 0) = 1 - e^{-\bar{\lambda}}$. This makes sense, since as the intensity goes to zero (or infinity),

the probability of presence goes to zero (or one).

Consideration for what sort of data (presence-only versus presence-absence) we are able to obtain is reflected in the intensity function $\lambda(\mathbf{s}, t)$. For presence-only data, the intensity function derived from the data will be biased due to imperfect detection and spacial autocorrelation (this is also true for presence-absence data, with the difference that we can model the detection probabilities with some level of confidence when we have access to absence data). Thus we adapt sampling bias into the Poisson process model by considering that we are modeling only subset of all individuals in the space-time volume, that is unobserved individuals are removed or "thinned" - hence the model's name *thinned inhomogeneous Poisson point process model*, or thinned IPP (Fithian et al., 2015). Thinned IPP models can be implemented using various approaches and tools, in fact Renner and Warton (2013) showed the equivalence between Maxent, a popular model for presence-only data, and a thinned IPP model.

Given absence data, then the parameters of interest of the ("unthinned") IPP can be obtained from hierarchical logistic regression models that jointly describe and predict the probability of occupancy and detection of animals or plants (Kéry and Schaub, 2011). Such models can be used to separate the ecological and observation components that make up observed data. The separation of these two different components allows us to describe the data as the result of two different by linked processes: the process dictating where the species occurs and the process dictating how the species is detected when present (Guillera-Arroita et al., 2014).

In case study I, the distribution of oceanic whitetip shark was modeled using presence-only data and various approaches, such as generalized linear models, Maxent models and Bayesian parameter inference models via a Markov Chain Monte Carlo algorithm. In case study II, Dorazio and Rodriguez (2012)'s hierarchical Bayesian approach to a site occupancy model is adapted and applied to trout data[1].

2.2 Case Study I: Oceanic Whitetip Shark Distribution

2.2.1 Introduction

As recently as 1968, the *Carcharhinus longimanus* or the oceanic whitetip shark proved its abundance by being a large percentage of fisheries total shark catch. Only recently, due to

[1] work with Christopher Custer, a Statistics Ph.D. student at WSU at the time

decrease in their catch percentage, this species has received attention in terms regulatory and management protections. Despite the recent surge in research for this species, there is still little understanding of their distribution, movements and global abundance due to their small numbers in the available datasets (Young and Carlson, 2020). There is strong evidence to believe that the oceanic whitetip population has suffered significant decline in the recent decades due to overfishing (Rigby et al., 2020, Young and Carlson, 2020, Bonfil et al., 2008), and climate change will likely exacerbate their decline. Continued data collection and study for this species is essential to determine information to help the development and implementation of conservation policy. It is also important to model and study the process which created the data observed, so we may better understand the factors that are likely affecting this species distribution.

2.2.2 Models

There are several approaches we can take when modeling species distribution, and as the hierarchical framework described earlier insinuated (and Renner and Warton (2013) showed), these models are all equivalent. In this case study we will compare the results of three different species distribution modeling approaches: generalized linear model and Maxent (maximum likelihood approaches), as well as Markov chain Monte Carlo (Bayesian approach). For comparison purpose, the same presence data and environmental covariates are used across the models, as well as the same linear relationship between the response and covariates (with the exception of the Bayesian approach which revealed that some covariates should not be modeled linearly).

The occurrence of a particular species in a particular 2D space cell i is modeled as a Bernoulli variable with parameter ϕ_i, i.e. $y_i \sim \text{Bernoulli}(\phi_i)$. The parameter ϕ_i is determined by a linear model of M predictors X_i, for cell i, via a logit link function, i.e.

$$\log\left(\frac{\phi_i}{1-\phi_i}\right) = \alpha + \sum_{j=1}^{M} \beta_j X_{ij}.$$

In the Bayesian approach, the parameters $\alpha, \beta_{j=1,\ldots,M}$ are modeled via a multivariate Normal jumping distribution $\text{MVN}(\mu, \Sigma_{M\times M})$, where μ is initially sampled from $N(0,1)$ and the standard deviation $\Sigma^{1/2}$ is fixed at $0.1 \cdot \text{diag}(1, \text{ncol} = M)$.

2.2.3 Application to Oceanic Whitetip Shark Data

Given the global range of the oceanic whitetip shark (fig. 1), we model the entire global marine environment. Oceanic climatic variables were obtained from Bio-Oracle (Tyberghein et al., 2012, Assis et al., 2018), and oceanic whitetip occurrence records were obtained from OBIS (2020). The data was processed using R version 4.0.2. The occurrence records were thinned to one record per 30 km and background point were randomly sampled from the currently known range for the oceanic whitetip shark in order to account for sampling bias.

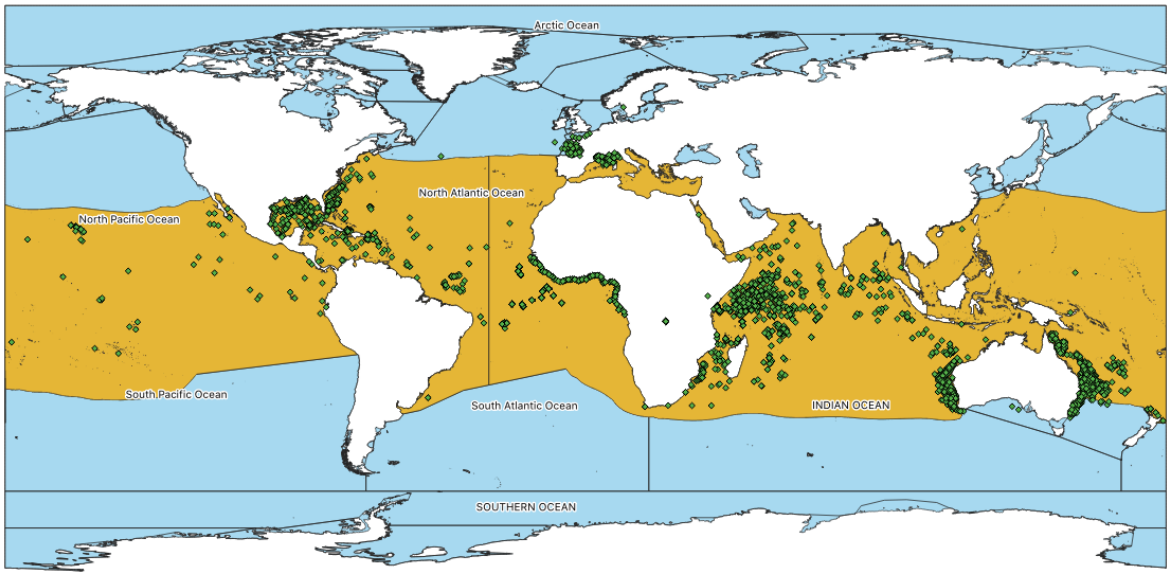

Figure 1: *Carcharhinus longimanus* range as per IUCN and distribution based on presence-only data obtained from OBIS.

The covariates used in all models pertain to sea surface conditions, due to the epipelagic nature of the oceanic whitetip shark, they are: mean temperature, temperature range, current velocity range, mean salinity, and salinity range. Variables were selected based on the availability of data layers for the present climate, as well as model selection with GLM and climate-envelopes.

The GLM results showed all covariates having a significant positive effect on the log odds of occupancy. The resulting suitability scores and AUC from this model are depicted in figures 2 and 3, respectively. The results from the Maxent model with linear features were very similar to those of the GLM model, as we can see from the suitability scores in figure 4 and the AUC in figure 5.

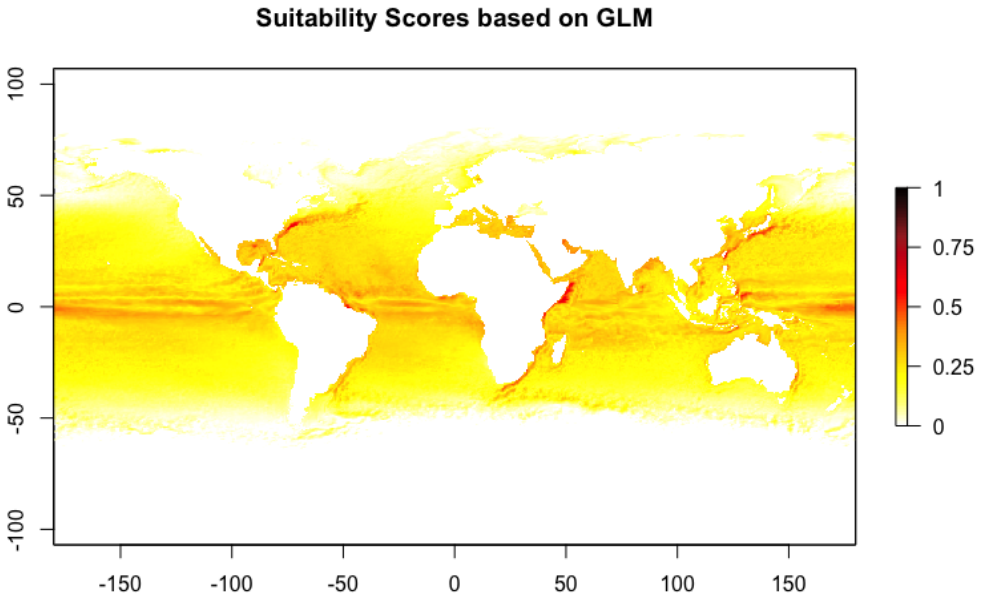

Figure 2: Suitability scores obtained from the generalized linear model with five covariates: mean temperature at sea surface, temperature range at sea surface, mean salinity at sea surface, salinity range at sea surface, and current velocity range at sea surface.

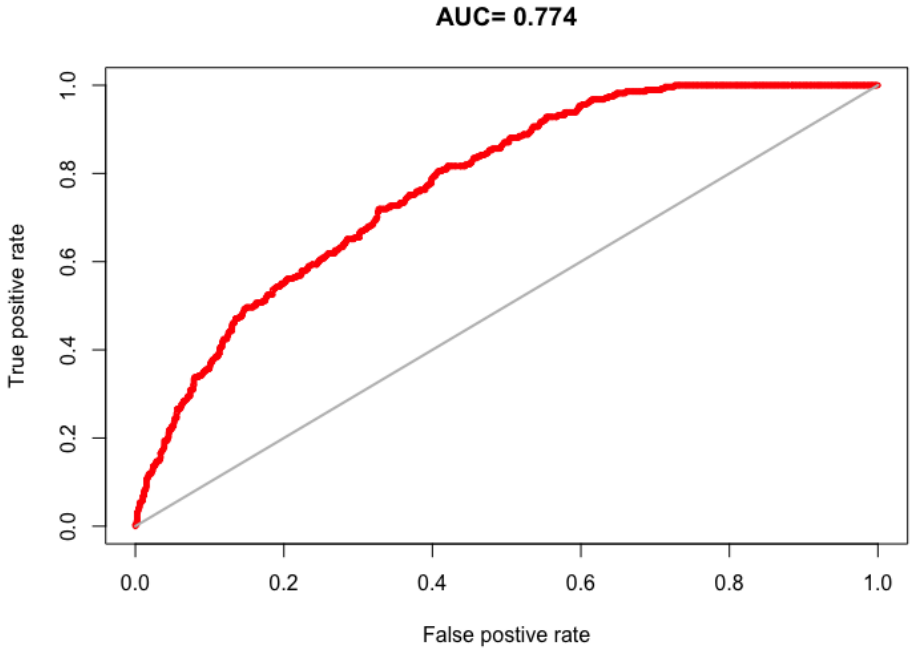

Figure 3: AUC from the generalized linear model with five covariates: mean temperature at sea surface, temperature range at sea surface, mean salinity at sea surface, salinity range at sea surface, and current velocity range at sea surface.

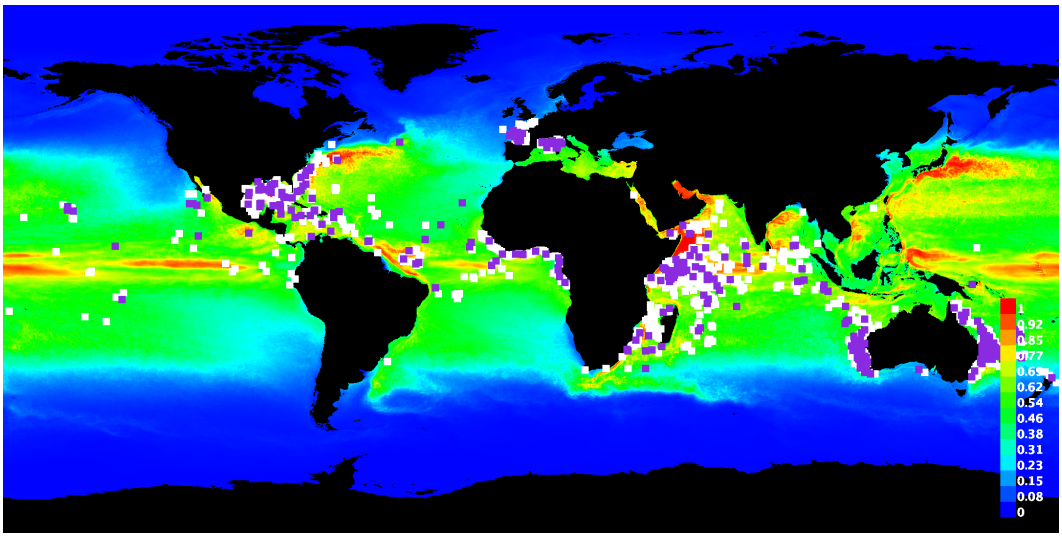

Figure 4: MaxEnt applied to the map of suitability scores.

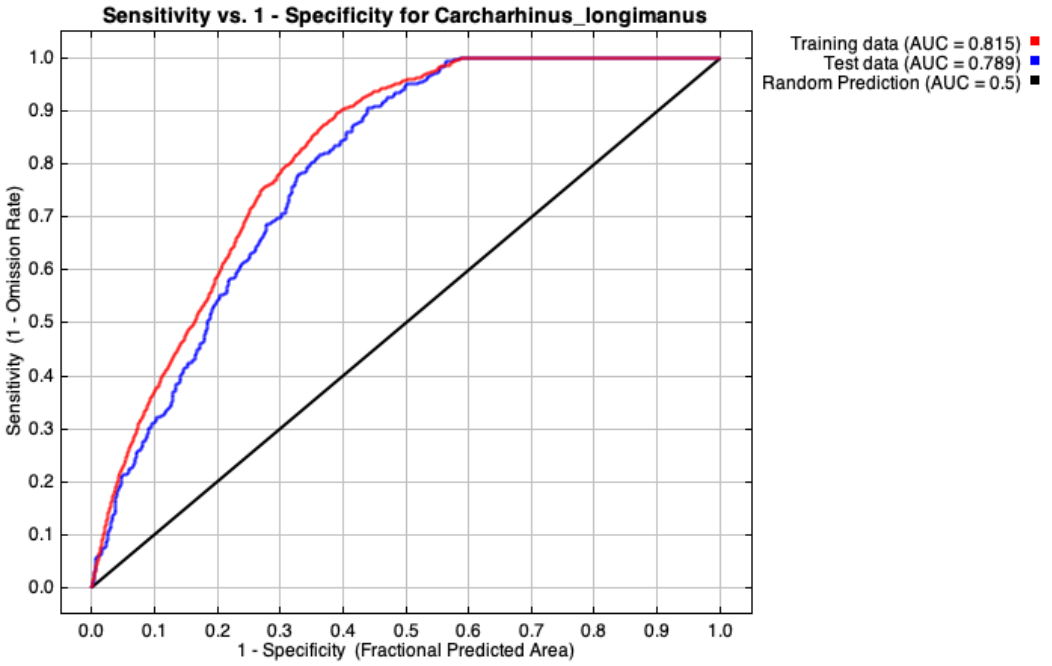

Figure 5: The AUC-ROC for the MaxEnt calculation, showing a reasonable performance and not too much overfitting.

The Bayesian approach was more difficult to implement and requires further work to obtain better results. The suitability scores from this model are depicted in figure 6, which shows similar results from the GLM and Maxent models above (the graininess is due to the model being applied to only the presence and random background points, instead of all longitude/latitude pairs - this was done for computational efficiency). Salinity covariates were dropped and range covariates modeled quadratically in order to achieve convergence (the linear model containing all covariates did not converge, perhaps due to choice of jumping distribution).

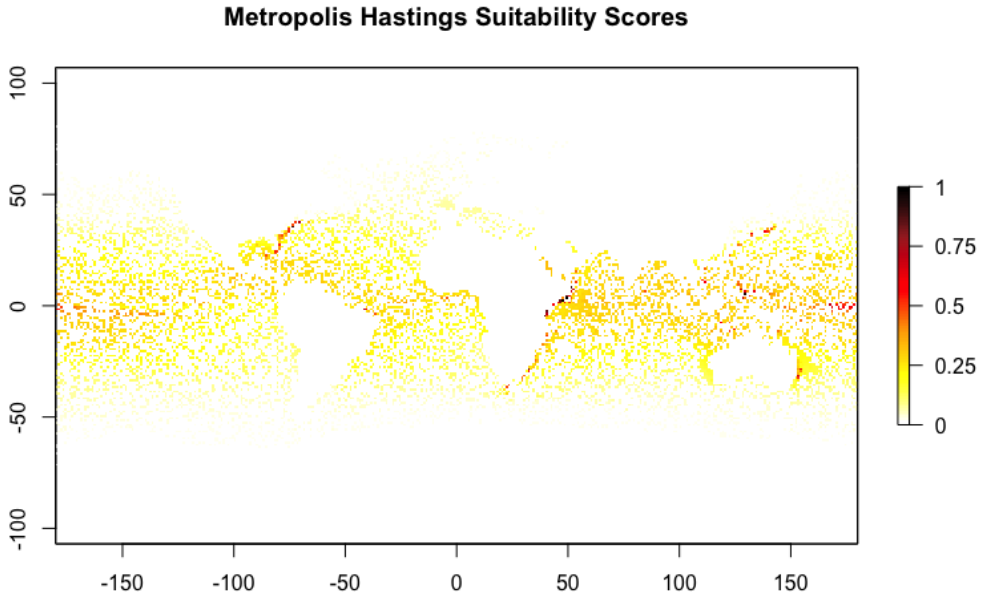

Figure 6: Suitability scores obtained via a Metropolis-Hasting algorithm.

2.3 Case Study II: A Site Occupancy Model for Trout

2.3.1 Introduction

In this section, we investigate a Bayesian approach for addressing the problem of species occurrence modeling in the face of imperfect detection by reproducing and adapting the site-occupancy model proposed by Dorazio and Rodriguez (2012). The aim will be to illustrate the profound usefulness of statistical computing tools such as Gibbs sampling, while simultaneously reviewing the usability and modularity of state-of-the-art ecological models and tools.

2.3.2 Model

As stated in Dorazio and Rodriguez (2012), site-occupancy data is usually collected during a number of independent surveys greater than one, denoted J, in n representative sites. The data collection consists of surveyors recording the detection of a particular species. Let y_{ij} be a binary random variable such that $y_{ij} = 1$ indicates a successful detection of the species of interest in site i during survey j. As the authors in the aforementioned paper, we will also assume J is constant across all n sites. Thus, the matrix Y of detection data is $n \times J$ dimensional. However, given that detection is imperfect we will need another binary random variable z_i for each site i, such that $z_i = 1$ pertains to the occurrence of the species in site i. Thus, the vector $\mathbf{z}$ of presence data is $n \times 1$ dimensional.

Since the z_i's and y_{ij}'s are binary random variables they follow Bernoulli distributions, with parameters pertaining to the probability of species occurrence, ψ_i, and the probability of species detection given occurrence, p_{ij}, respectively:

$$z_i|\psi_i \sim \text{Bernoulli}(\psi_i)$$
$$y_{ij}|z_i, p_{ij} \sim \text{Bernoulli}(z_i p_{ij})$$

As the authors in the original study, we will treat informative covariates of species occurrence obtained at each of the n sites, as a $n \times r$ matrix of r regressors, denoted $\mathbf{X}$, however instead of using a probit-regression as Dorazio and Rodriguez (2012) do in their work, we will use a logistic regression.

As noted by Clark and Altwegg (2019), using the logistic link function is more algebraically tractable due the existence of an analytic expression for its PDF. They also go on to point out the intuitiveness of interpreting the coefficient effects of the estimated model as an odds ratio, something that cannot be done with the probit link function.

Thus ψ_i, the probability of species presence in location i can be formulated as follows:

$$\log\left(\frac{\psi_i}{1-\psi_i}\right) = \hat{\beta}_0 + \mathbf{x}_i^T \hat{\beta}_{r\times 1} \Rightarrow \psi_i = \frac{1}{1+e^{-(\hat{\beta}_0 + \mathbf{x}_i^T \hat{\beta}_{r\times 1})}} \sim \text{LogisticDF}(\hat{\beta}_0 + \mathbf{x}_i^T \hat{\beta}_{r\times 1}),$$

where LogisticDF denotes a "standard" logistic distribution of location 0 and scale 1. Making $\mathbf{X}$ as a $n \times (r+1)$ dimensional matrix by including a vector of ones to correspond to the intercepts we obtain: $\psi_i \sim \text{LogisticDF}(\mathbf{x}_i^T \hat{\beta})$.

Similarly, we can have informative covariates of species detection given species occurrence in a matrix denoted W. In this case, if we have q covariates in each of the n sites for each of the J surveys, then W is $n \times J \times (q+1)$ dimensional, using the similar trick of adding a vector of ones to each $n \times J$ array of covariates in order to better capture the intercept in the notation. Thus, also modeling the p_{ij}'s as a logistic regression, we get that $p_{ij} \sim \text{LogisticDF}(\mathbf{w}_{ij}^T \hat{\alpha})$.

The joint posterior density then, is proportional to some constant multiple of the prior distributions for $\hat{\alpha}$ and $\hat{\beta}$, assumed to be independent, times the joints for $z_i|\psi_i$ and $y_{ij}|z_i, p_{ij}$:

$$\begin{aligned}
\pi(\hat{\beta}, \hat{\alpha}, \mathbf{z}|\mathrm{Y}) &\propto \pi(\hat{\beta})\pi(\hat{\alpha}) \cdot \prod_{i=1}^{n} \psi_i^{z_i}(1-\psi_i)^{1-z_i} \cdot \prod_{j=1}^{J} (z_i p_{ij})^{y_{ij}}(1-z_i p_{ij})^{1-y_{ij}} \\
&\propto \pi(\hat{\beta})\pi(\hat{\alpha}) \cdot \prod_{i=1}^{n} \left(\frac{1}{1+e^{-\mathbf{x}_i^T\hat{\beta}}}\right)^{z_i} \left(1 - \frac{1}{1+e^{-\mathbf{x}_i^T\hat{\beta}}}\right)^{1-z_i} \\
&\quad \cdot \prod_{j=1}^{J} \left(z_i \cdot \frac{1}{1+e^{-\mathbf{w}_{ij}^T\hat{\alpha}}}\right)^{y_{ij}} \left(1 - z_i \cdot \frac{1}{1+e^{-\mathbf{w}_{ij}^T\hat{\alpha}}}\right)^{1-y_{ij}}
\end{aligned}$$

The assumption of mutual independence between the $\hat{\beta}$ and $\hat{\alpha}$ parameters allow for easy derivation of the full conditional densities:

$$\begin{aligned}
k(\hat{\beta}|z_i, \mathbf{x}_i) &\propto \pi(\hat{\beta}) \cdot \prod_{i=1}^{n} \left(\frac{1}{1+e^{-\mathbf{x}_i^T\hat{\beta}}}\right)^{z_i} \left(1 - \frac{1}{1+e^{-\mathbf{x}_i^T\hat{\beta}}}\right)^{1-z_i} \\
k(\hat{\alpha}|y_{ij}, \mathbf{w}_{ij}) &\propto \pi(\hat{\alpha}) \cdot \prod_{i=1}^{n} \left(\prod_{j=1}^{J} \left(z_i \cdot \frac{1}{1+e^{-\mathbf{w}_{ij}^T\hat{\alpha}}}\right)^{y_{ij}} \left(1 - z_i \cdot \frac{1}{1+e^{-\mathbf{w}_{ij}^T\hat{\alpha}}}\right)^{1-y_{ij}} \right)
\end{aligned}$$

However, sampling from these distributions can be difficult, therefore we will use the same data-augmentation approach as in Dorazio and Rodriguez (2012), but here the latent variables introduced will be logistic random variables, that is, we will let $v_i \sim$ Logistic(location $= \mathbf{x}_i^T\hat{\beta}$, scale $= 1$) and assume $z_i = 1$ if $v_i > 0$ and $z_i = 0$ if $v_i \leq 0$. Similarly, we let $u_{ij} \sim$ Logistic(location $= \mathbf{w}_{ij}^T\hat{\alpha}$, scale $= 1$) and assume $y_{ij} = 1$ if $u_{ij} > 0$ and $z_i = 1$, or assume $y_{ij} = 0$ otherwise. Thus we can rewrite the original hierarchical model augmented to

include the latent variables as follows:

$$
\begin{aligned}
v_i &\sim \text{Logistic}(\text{location} = \mathbf{x}_i^T\hat{\beta}, \text{scale} = 1) \\
z_i|v_i &\sim \text{Bernoulli}(\mathbf{1}(v_i > 0)) \\
u_{ij} &\sim \text{Logistic}(\text{location} = \mathbf{w}_{ij}^T\hat{\alpha}, \text{scale} = 1) \\
y_{ij}|z_i, u_{ij} &\sim \text{Bernoulli}(z_i\mathbf{1}(u_{ij} > 0))
\end{aligned}
$$

The augmented joint posterior density is then:

$$
\begin{aligned}
\pi(\hat{\beta}, \hat{\alpha}, \mathbf{z}, \mathbf{v}, \mathbf{U}|\mathrm{Y}) \propto \pi(\hat{\beta})\pi(\hat{\alpha}) \cdot \prod_{i=1}^{n} \mathbf{1}(v_i > 0)^{z_i}\mathbf{1}(v_i \leq 0)^{1-z_i} \cdot \left(\frac{e^{-(v_i - \mathbf{x}_i^T\hat{\beta})}}{(1 + e^{-(v_i - \mathbf{x}_i^T\hat{\beta})})^2}\right) \\
\cdot \prod_{j=1}^{J} \left(z_i\mathbf{1}(u_{ij} > 0)\right)^{y_{ij}} \left(1 - z_i\mathbf{1}(u_{ij} > 0)\right)^{1-y_{ij}} \cdot \left(\frac{e^{-(u_{ij} - \mathbf{w}_{ij}^T\hat{\alpha})}}{(1 + e^{-(u_{ij} - \mathbf{w}_{ij}^T\hat{\alpha})})^2}\right)
\end{aligned}
$$

Sampling from the above augmented joint posterior is not directly possible. However, here is where the authors in our paper of interest apply their insight for an accurate and efficient estimate that relies on Gibbs sampling. We borrow and adapt their idea to our logistic model.

We used uniform priors for $\hat{\beta}$ and $\hat{\alpha}$ to indicate prior indifference regarding the magnitude of these parameters. The full conditionals for our adaptation of the Gibbs sampler are:

1. $v_i|\cdot \sim \begin{cases} \text{leftTruncatedLogistic}(\mathbf{x}_i^T\hat{\beta}, 1) & \text{if } z_i = 1 \\ \text{rightTruncatedLogistic}(\mathbf{x}_i^T\hat{\beta}, 1) & \text{if } z_i = 0 \end{cases}$
2. $\hat{\beta}|\cdot \sim \text{Normal}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{v}, (\mathbf{X}^T\mathbf{X})^{-1})$
3. $u_{ij}|\cdot \sim \begin{cases} \text{leftTruncatedLogistic}(\mathbf{w}_{ij}^T\hat{\alpha}, 1) & \text{if } y_{ij} = 1 \text{ and } z_i = 1 \\ \text{rightTruncatedLogistic}(\mathbf{w}_{ij}^T\hat{\alpha}, 1) & \text{if } y_{ij} = 0 \text{ and } z_i = 1 \end{cases}$
4. $\hat{\alpha}|\cdot \sim \text{Normal}((\hat{\mathbf{W}}^T\hat{\mathbf{W}})^{-1}\hat{\mathbf{W}}^T\hat{\mathbf{u}}, (\hat{\mathbf{W}}^T\hat{\mathbf{W}})^{-1})$
5. $z_i|\cdot \sim \begin{cases} \text{Bernoulli}(1) & \text{if } \mathbf{y}_i \neq \mathbf{0} \\ \text{Bernoulli}\left(\frac{1}{1+e^{-\mathbf{x}_i^T\hat{\beta}}}\right) & \text{if } \mathbf{y}_i = \mathbf{0} \end{cases}$

   where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iJ})^T$.

As in Dorazio and Rodriguez (2012)'s paper, $\hat{\mathbf{W}}$ is a $mJ \times (q+1)$ matrix, where mJ is the total number of observations per site $\times$ survey. Similarly, $\hat{\mathbf{u}}$ is a $mJ \times 1$ vector. Note that

to update $\hat{\alpha}$ we only use values of w_{ij} and u_{ij} such that $z_i = 1$, that is we only use the information from occupied sites.

2.3.3 Application to Trout Data

Brook trout are freshwater fish native to cold water streams of the eastern half of the United States, though they have been introduced for sport fishing across the country. In order to apply our model to a real-world dataset, we were able to obtain occupancy data from Oregon State University's "Datasets for hands-on portion of Occupancy" workshop (OSU, 2019). According to the site, these data were collected via electrofishing in the Upper Chattachochee River basin. The survey consisted of 77 streams which were sampled 3 times each. For each site, the elevation was collected. Relating this back to our data simulation, this is our X covariate and will be associated with the $\hat{\beta}$ parameters. Additionally, the elevation data was standardized for the modeling process to replicate the approach taken by Dorazio and Rodriguez (2012). For each sample, the mean cross-sectional sampling area was recorded. This is our W covariate associated with the $\hat{\alpha}$ parameters. The final piece of the dataset is the response variable, $\mathbf{y}$, which are of course binary datapoints represents species occurrence.

As in the model, the parameters $\hat{\beta}$ predict species occurrence at site i with elevation x_i while the parameters $\hat{\alpha}$ predict detection probability at a site i during survey j with a cross sectional sampling area of w_{ij}. Applying the Gibbs sampling algorithm to the trout data resulted in estimates for all four parameters as seen in figure 7. These models suggest a positive relationship with species occurrence and elevation along with a negative relationship between cross-sectional sampling size area and detection probability.

The negative estimate for α_1 was a bit surprising at first, as one would expect an increase in sampling area to increase probability of detection, but could be the result of two things: 1) sampling bias; perhaps the scientist covered more area when unable to locate a fish 2) habitat preference; larger sampling areas could be the result of wider, more accessible parts of the stream which may reflect less desirable trout habitat. Additionally, we can see in figure 7 that both of our slope estimates' (β_1 and α_1) 95% credible intervals do not contain 0. This means there is a 95% probability that our estimates are in fact greater than and less than, respectively, 0 thus suggesting a true relationship between our covariates and response.

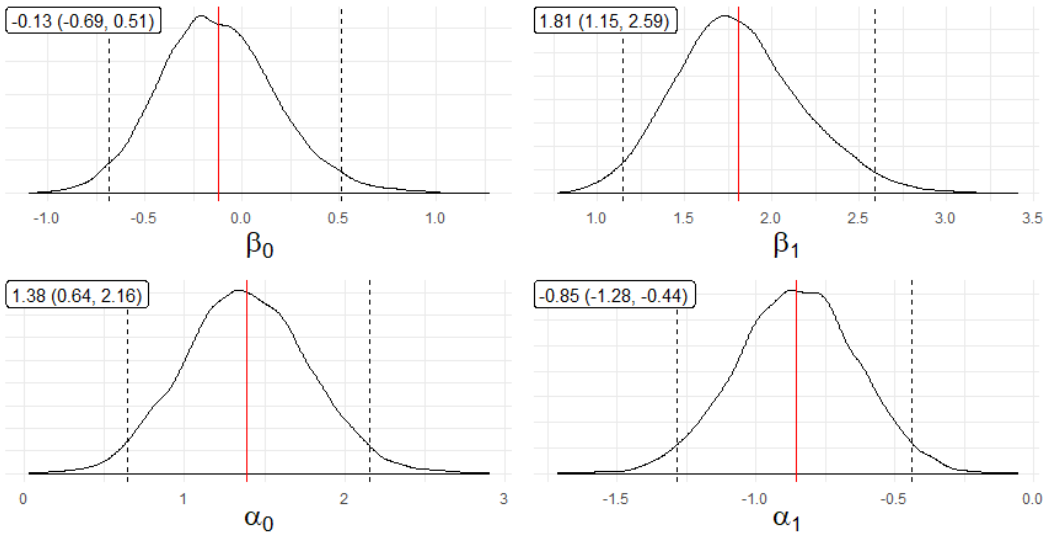

Figure 7: Posterior density curves of occupancy model parameter estimates with mean (red line) and 95% credible intervals (dashed lines).

3 Hidden Markov Models

Given a sequence of observations, usually over time, a hidden Markov model (figure 8) assumes that there is a hidden process behind the observations that we do not observe. Each observation at a given time (here assumed to be discrete) is emitted via some distribution given a hidden state, and the observations are independent of each other. The hidden process has the Markov property, where each state at a given time t depends only on the state at the previous time step $t-1$.

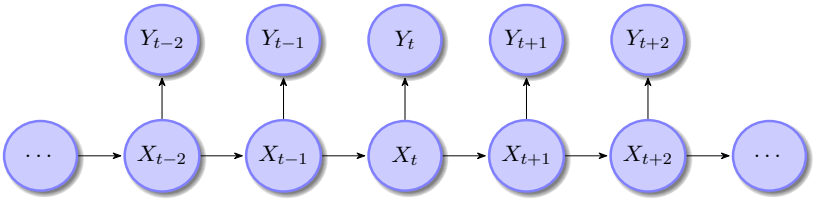

Figure 8: A hidden Markov model with hidden states X_t and observations Y_t for $t = 1, \ldots, T$.

The hidden Markov model framework allows one to build and computationally implement models that would otherwise be considered intractable, by separating the state and observation processes. A basic discrete-time finite state hidden Markov model (or HMM) can be completely defined by: 1) the number of states, N; 2) the transition probability matrix Γ, for the hidden states X, with entries $\gamma_{ij} = P(X_t = j | X_{t-1} = i)$ for $i, j \in \{1, \ldots, N\}$; 3) state-dependent distributions for the observations Y that are conditional on the state at time t, i.e $f(Y_t | X_t = i)$, where $f_i(Y_t)$ is a valid distribution function; and 4) an initial state distribution,

δ specifying the probabilities of being in each state when the observation sequence starts, i.e $P(X_1 = k)$ (McClintock et al., 2020).

HMMs have a broad set of applications in many fields and they can also be formulated in continuous time (McClintock et al., 2020), however our focus in this section will be on the ecological applications of discrete-time HMMs to time series data. Given the complete definition of the discrete-time HMM above, there are several parameters that we wish to estimate, such as those used for state dependent distributions. There are also parameters for the model determining the transition probabilities, which can either be assumed to be homogeneous and based on a stationary distribution or vary over time given covariates.

We can use two main approaches to obtain these parameters: 1) maximum likelihood or 2) Bayesian. In case study I, a network of HMMs is developed to classify presence of fire in a two dimensional forested area from satellite images. Model parameters are found via a maximum likelihood approach, specifically the Baum-Welch algorithm[2]. In case study II, a hidden Markov model is used in combination with the Bayesian parameter estimation approach developed in Leos-Barajas and Michelot (2018) to study different behavioral modes in bald eagles using GPS telemetry data.

3.1 Case Study I: Forest Fires and Neuronal Activity

3.1.1 Introduction

An analogy can be made between neuronal activity in the brain and fire spreading through a forest which revolves around the characterization of both neurons and forested lands existing in one of three states: active, either a neuron spiking or a forest burning; refractory, a neuron that just spiked or a forest that just burned; or quiescent, a neuron that is ready to fire again or a forest with new un-burned growth. The probability that a neuron (or forest) will transition between the three states is in turn influenced by the current state of its neighbors: neurons transmit spikes via synapses, and forest fires spread to nearby areas.

Many other factors can also affect these transition probabilities. For example, the spread of a forest fire may be encouraged by strong winds or dry weather, and neural activity can be modulated by arousal or sensory context (David, 2018). These factors may trigger any of the states at a particular time step depending on their probability of occurrence and the

[2]work done with Jacob Pennington, a Mathematics Ph.D. student at WSU

stochastic dynamics of signal spread. Ultimately, a realistic model of either system should incorporate these additional factors. However, here we focus on modeling the influence of neighboring neurons/forest plots.

Although the model developed and presented here can be used to classify the behaviors exhibited by both neuronal activity and forest fire spread, for this case study we only discuss the modeling of forest fire spread.

3.1.2 Model

We initially construct a simple discrete-time homogeneous Markov chain before expanding it to a hidden Markov model. Let X_n be a discrete random variable defined on the finite state space $\{a, r, q\}$ (for active, refractory, quiescent) such that the probability to transition from any state i to a new state j is $P(X_{n+1} = j | X_n = i) = p_{ji}$.

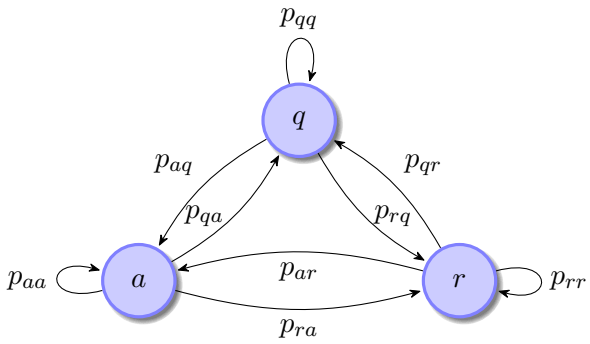

Figure 9: A DTMC for the three possible states of neuron/tree clusters. At each time step, the system transitions from its current state $i \in \{a, r, q\}$ to a new state j in the same set of states by travelling along an edge with a probability of p_{ji}.

Given the Markov chain described in 9, we have the following transition matrix P:

$$P = \begin{pmatrix} p_{aa} & p_{ar} & p_{aq} \\ p_{ra} & p_{rr} & p_{rq} \\ p_{qa} & p_{qr} & p_{qq} \end{pmatrix} \tag{1}$$

Although simple, the DTMC described above can serve as a foundation for more complex models, such as hidden Markov models (HMM), which is what we will explore next. As explained before, a HMM is a discrete-time finite state homogeneous Markov chain observed through a discrete-time channel characterized by a finite set of transition densities (such as Gaussian or Poisson) that are indexed by the states of the Markov chain. The underlying

process is not observable and the second process is a sequence of independent random variables that are conditional on the underlying Markov chain (Ephraim and Merhav, 2002).

Our hidden Markov model will take the Markov chain introduced above and place it behind a "curtain" so that we only observe noisy states that are functions of each of the original states. Thus observations $\{\mathcal{O}_n\}_{n=0}^{\infty}$ is a process (with a distribution yet to be determined) that is dependent on the respective underlying state for the particular time step of the original DTMC process, i.e. $\{X_n\}_{n=0}^{\infty}$. Realistically the observation sequence is time bounded by some time length T. Note that the number of unique symbols from the observation sequence need not be the same as the number of unique states occurring in the hidden Markov process, as noise and other phenomena might alter observed symbols pertaining to the same hidden state. Ideally, the distribution that connects observed to hidden states would capture these deviations with a high level of confidence.

By expanding (or rather hiding) the DTMC from figure 9 we obtain the following Markov model from figure 10, where for each state $i \in \{a, r, q\}$ (the state-space as before), we have $\mathcal{O}_i$ an observation of state i through a filter and/or with some noise.

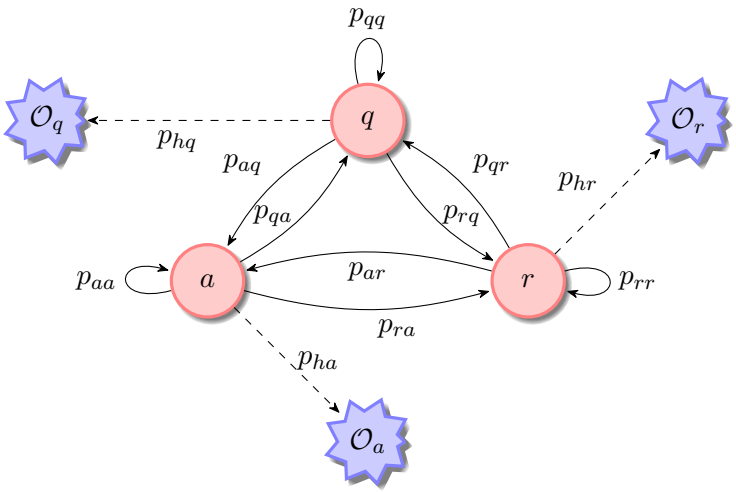

Figure 10: Hidden Markov model for the three possible states of neuron/tree clusters.

The complete specification of the HMM, along with the notation used in this case study, are as follows:

> T = the time length of the observation sequence. For our initial implementation we will observe 12 months of NASA's MODIS Images, however this can change based on the data and will be modeled here as T.

$N = |S| = 3$ be the number of states in the model.

$M = |V| =$ the number of observation symbols. As with T, this number will depend on the data. The implementation will use $M = 2$.

$S = \{a, r, q\}$ be the distinct states in the hidden DTMC.

$V = \{0, 1, \ldots, M-1\}$ be the distinct symbols that can be possibly observed. The implementation w ill only consider two possible symbols: 0 and 1.

$P =$ the $N \times N$ matrix of state transition probabilities defined in (1).

$B =$ the $N \times M$ matrix of observation probabilities, that is, the probability of observing symbol $k \in V$ given the hidden state $j \in S$.

$\pi =$ a 3×1 vector of initial state probabilities.

$\mathcal{O} = (\mathcal{O}_0, \mathcal{O}_1, \ldots, \mathcal{O}_{T-1})$ be the observation sequence. In the model implementation, this sequence is of length 12.

Therefore, a hidden Markov model λ is the 3-tuple (P, B, π). As in the DTMC and in order to be consistent with the transition matrix description provided earlier $P = \{p_{ji}\}$ is $N \times N$ and column stochastic, with $p_{ji} = P(\text{state } j \text{ at } t+1 \mid \text{state } i \text{ at } t)$. In the model implementation however, we find it easier to transpose the transition matrix making it effectively row stochastic. The matrix $B = \{b_j(k)\}$ is $N \times M$ and row stochastic, with $b_j(k) = P(\text{observing } k \text{ at } t \mid \text{state } j \text{ at } t)$. Both matrices, P and B have probabilities that are time homogeneous.

Thus we must solve the following problem: given an observation sequence $\{\mathcal{O}_n\}_{n=0}^{T-1}$, the number of states we wish to describe $N = 3$, and a number M of observed symbols (from our image data), we wish to find the model $\lambda = (P, B, \pi)$ that maximizes the likelihood of the given observation sequence. For this purpose, we will use the Baum-Welch reestimation formulas described in Rabiner and Juang (1986).

We incorporate spatial information by fitting HMMs to every node of a network, each representing a two dimensional plot of land. This network, which we call a hidden Markov network (HMN), has two layers. The hidden layer is a network of DTMCs, where each node is a plot of land with trees or other ecological objects that can be in one of three states with

respect to fire: active, refractory or quiescent. Thus each node follows the same DTMC rules specified above. In the network depicted in figure 11, we see an instance of the hidden layer. In this network, the nodes are trees and their time based changes, while the edges are their neighboring connections in space. In this particular instance of time, we have 9 nodes where 1 is refractory, 2 are active and the remaining 6 are quiescent. The center node C will be of particular interest for our model, as it is adjacent to 2 active nodes, a rule which will influence its DTMC transition probabilities matrix P.

In our HMN, we also have the observed layer, which is a network of 2D spatially connected observations such as pixels in an image. Thus each node follows the same HMM rules specified before. As we can see in figure 12, which is a single time step instance of a network of HMMs, each node can be one of two possible observation symbols: fire or no fire, and as before observations are tied to their respective underlying states via the row stochastic matrix B.

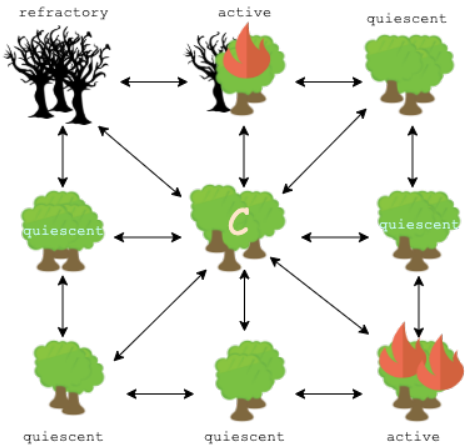

Figure 11: A network representing the 2D spatial connections of DTMCs.

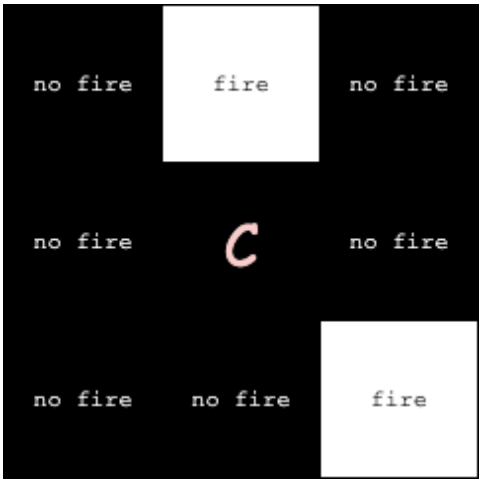

Figure 12: A network representing the 2D spatial connections of the observed symbols of a HMM whose underlying hidden states are depicted in 11.

Let us stop time and discuss an example of how the consideration of spatial neighbors affect the transition probabilities matrix P for a given node C. Say, for example, that at a particular point in time t_k, C has transition matrix

$$P = \begin{pmatrix} 0.5 & 0.0 & 0.2 \\ 0.5 & 0.9 & 0.0 \\ 0.0 & 0.1 & 0.8 \end{pmatrix},$$

recall that P is a column stochastic matrix, where the rows and columns represent the states: active, refractory and quiescent, in this order. Now suppose that during t_k, C is adjacent to two active (on fire) nodes. To obtain the transition matrix P' for time t_{k+1}, we consider two particular state transitions: active to active and quiescent to active. Since the spatial closeness of the plots of land mean that actively burning trees are more prone to continue burning if they are close to other actively burning trees, and similarly quiescent trees are more prone to enter an active burn state if they are close to actively burning trees, then we can update the transition matrix by multiplying these states by "warming factors" α (for state $a \to a$) and β (for state $q \to a$).

Continuing with the example transition matrix from above, if we consider warming factors $\alpha = 1.2$ and $\beta = 2$, then at time t_{k+1}, we obtain

$$P' = \begin{pmatrix} 0.59 & 0.0 & 0.5 \\ 0.41 & 0.9 & 0.0 \\ 0.0 & 0.1 & 0.5 \end{pmatrix},$$

by multiplying P_{aa} by α twice (once for each active neighbor) and multiplying P_{aq} by β twice, then re-normalizing the columns to sum to one. Thus P' is the transition matrix that will be used at the subsequent time step t_{k+1}.

3.1.3 Application to MODIS Image Data

We used NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) images in which each pixel reports whether a fire took place within a 1000m square plot of land at some time during a 24-hour day period, recorded in 2018 (these data represent approximately 1.5 million nodes in our network). Values indicating unburned land, missing data and water are also assigned. Although this labeling made the estimation of hidden states partially redundant - the labels on the images already tell us whether a plot is in the "active" state or not (but cannot distinguish between refractory versus quiescent), the images were ideal for model training and initial sanity checks. For example, if our model estimation process works as intended then a hidden "active" state should correspond to an observed "fire" state with very high probability.

The initial values assigned were:

$$\pi = \begin{pmatrix} 0.005 \\ 0.005 \\ 0.990 \end{pmatrix}, \; P = \begin{pmatrix} 0.50 & 0.01 & 0.09 \\ 0.25 & 0.90 & 0.01 \\ 0.25 & 0.09 & 0.90 \end{pmatrix}, \; B = \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \\ 0.9 & 0.1 \end{pmatrix}.$$

In order to implement the spatial influence strategy described above, we devised a cost function for determining the optimal combination of network influence parameters. We decided to use the mean squared error (MSE) between the fire density of the data and the fire density averaged over five simulations, where fire density was calculated by averaging fire labels over the spatial dimensions. For the ≈ 70 image dataset for the year 2010 of the Western Amazon, we obtained the following optimized parameters (averaged over all nodes in the network):

$$\pi = \begin{pmatrix} 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}, \; P = \begin{pmatrix} 0.39 & 0.00 & 0.11 \\ 0.38 & 0.89 & 0.03 \\ 0.23 & 0.11 & 0.86 \end{pmatrix}, \; B = \begin{pmatrix} 0.0 & 1.0 \\ 0.0 & 1.0 \\ 1.0 & 0.0 \end{pmatrix}.$$

Some simulation snapshots (each taken from the same time step) are shown in figure 13 for four example combinations of network parameters that demonstrate the effect of each parameter on the network's behavior, as well as two snapshots from the data during the onset (top) and peak (bottom) of the fire outbreak. Increasing α creates more "clusters" of ongoing fires (bottom row), while increasing β causes the fire to cover a broader area (middle column). These traits appear to resemble the peak and onset, respectively, of the fire outbreak in the data.

While the $\alpha = 1.0, \beta = 2.0$ combination (top left) achieved the lowest MSE over the whole year, we can see that it does not closely resemble the activity in the data during the fire outbreak. This discrepancy is also highlighted in figure 14 which shows that the simulation's fire density is essentially flat throughout the year (as was the case with other parameter values). In short, the model cannot emulate the temporal dynamics of the fire outbreak phenomenon, so the "best" parameter combination according to our chosen cost function is the one whose flat-line fire density most closely matches the annual average.

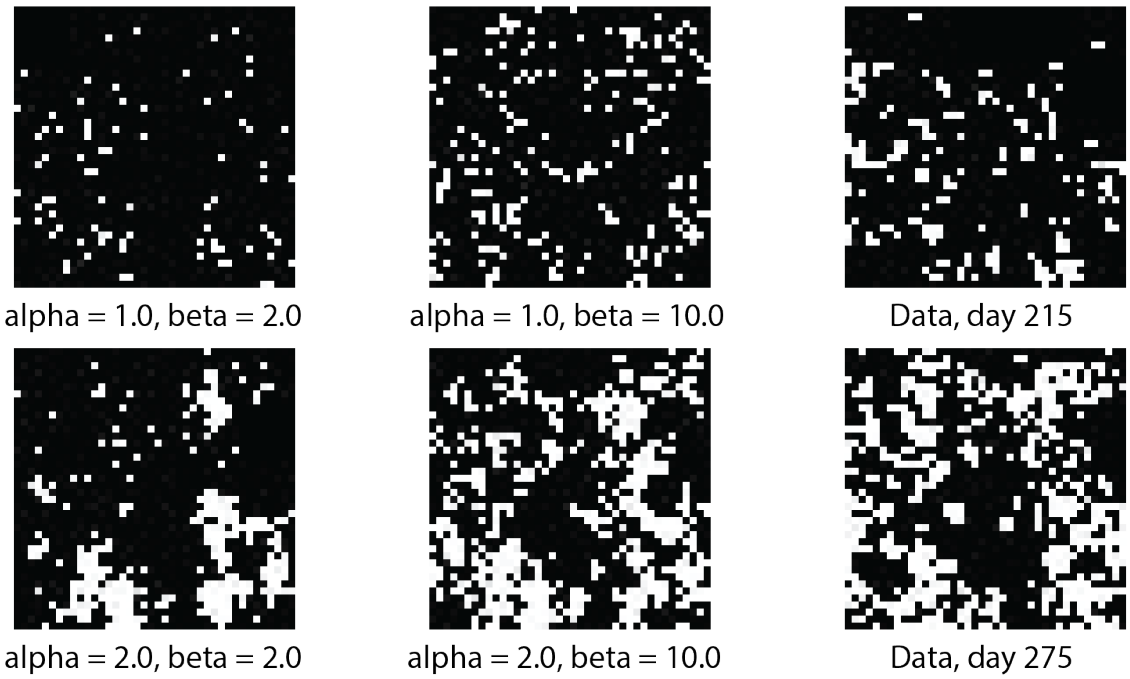

Figure 13: Example time steps from four network simulations compared to two time steps taken from the data. White points indicate pixels labeled as "on fire." For animated examples, see accompanying data files.

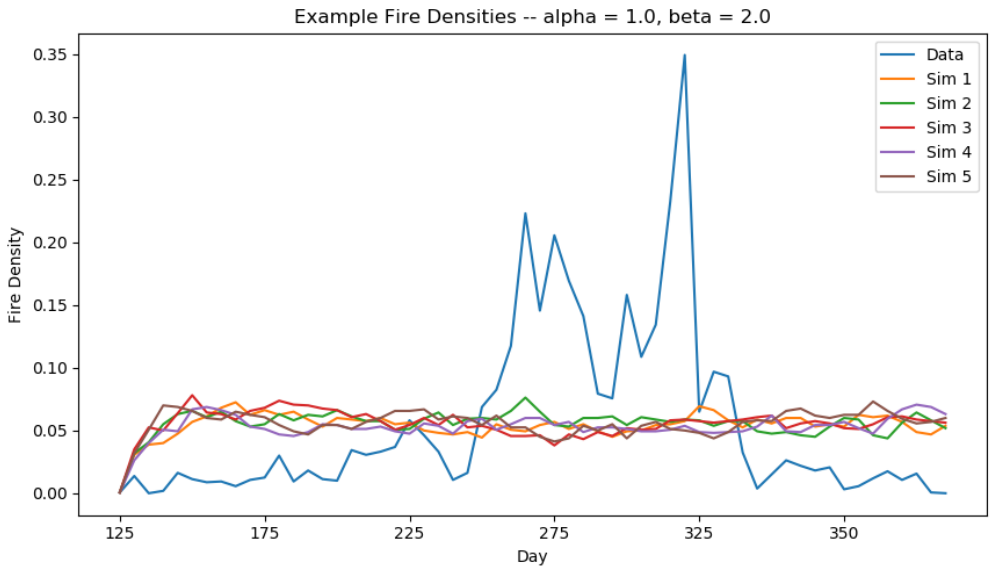

Figure 14: Fire densities

3.2 Case Study II: Bald Eagle Behavioral States

3.2.1 Introduction

When animals move through space and time, they do so to express a behavior. Animal behavior is usually not observed by humans, but the movements that are tied to their behavior can be. Modern tagging technology allows scientists to capture rich datasets of spatial-temporal observations describing movements of tagged animals. The relationship between observed movements and unobserved behaviors can be captured by a hidden Markov model (HMM), such as the state-space model used by Pirotta et al. (2018). In their model, golden eagle movement data obtained from GPS tracking (which included elevation) was used to describe five different flight modes representing the hidden states.

Beyond the general identification of distinct patterns in an animal's movement and how they relate to behavior, we may wish to identify age-dependent patterns in order to better describe the animal's life history. For instance, Sergio et al. (2014) show that the migratory behavior of black kites is gradually developed throughout an individual's life and that birds of different ages used different traveling tactics to cope and exploit drift and tail winds respectively.

Haliaeetus leucocephalus, or bald eagles - the subject of this study - live for about 20 years in the wild, and express different behaviors throughout their lifespan. The goal is to answer the scientific question: given bald eagle movement data, can we detect how the eagles' behaviors differ based on their age? If we can find a relationship between movement and age, then we can use this information to predict the life stage of eagles in the wild.

Understanding how bald eagles behave during a particular life stage can help wildlife managers make recommendations and decisions regarding conservation. Although bald eagles are considered "least concern" from a conservation standpoint[3], injuries and fatalities from collisions with human structures, such as aircrafts, wind turbines and power lines are ongoing issues wildlife managers must address (Erickson et al., 2005).

In this study, a Bayesian approach is used to estimate the parameters of an HMM that distinguishes the different observable patterns from bald eagle movements which can be tied to their behavioral modes. Using telemetry data collected from bald eagles by the U.S. Geological

[3]They were once endangered, however populations in the United States recovered substantially since strict protection laws for bald eagles were passed in 1966 and 1972

Survey (USGS), a hidden Markov model is constructed to differentiate movement patterns by age, and thus facilitate modeling the unobserved ontogenetic behavioral patterns.

3.2.2 Model

When considering animal movements based on geographic coordinates there are two sets of observation data we can model, the step lengths (in meters) and the turning angles (in radians). After obtaining the histograms of the step lengths and angles available in the data set (fig. 15), by letting L_t be a random variable for the step length at time t, and ϕ_t be a random variable of the turning angles at time t, as in Leos-Barajas and Michelot (2018), we pick the following distributions:

$$L_t|X_t = j \sim \text{Gamma}(\alpha_j, \beta_j)$$
$$\phi_t|X_t = j \sim \text{vonMises}(\mu_j, \kappa_j)$$

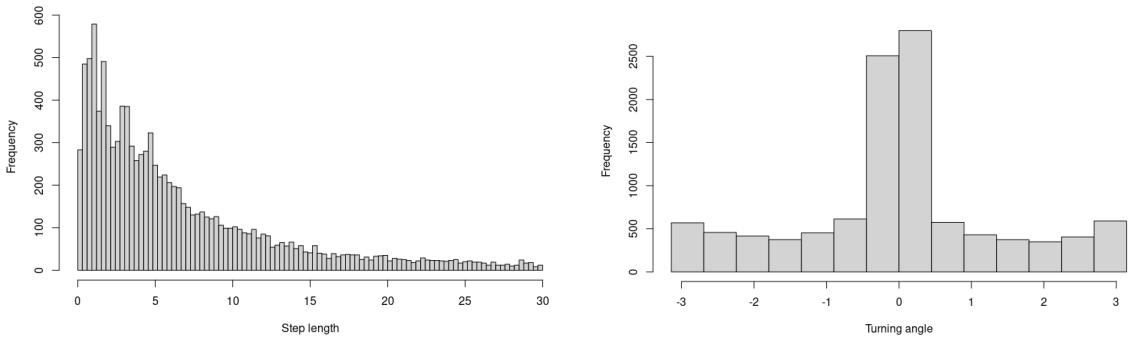

Figure 15: Histograms for step length and turning angle observations.

As for the transition probabilities, here they are modeled as a time-dependent logistic function of the covariates. The covariates used in the current implementation of this model are variables that are part of the same telemetry data, such as speed, altitude and heading. For the transition probability $\gamma_{ij}^{(t)}$ and covariates $\omega_l^{(t)}$ at time t, $l = 1, \ldots, p$, we have that

$$\gamma_{ij}^{(t)} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^{N} \exp(\eta_{ik})},$$
$$\text{where } \eta_{ij} = \begin{cases} \beta_0^{(ij)} + \sum_{l=1}^{p} \beta_l^{(ij)} \omega_l^t & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}.$$

With the aforementioned state-dependent distributions and the transition probabilities, we can write the marginal likelihood function

$$\mathcal{L}_m = \sum_{x_1=1}^{N} \cdots \sum_{x_T=1}^{N} \delta_{x_1}^{(1)} \prod_{t=2}^{T} \gamma_{x_{t-1},x_t} \prod_{t=1}^{T} f_{x_t}(y_t),$$

which can be formulated via the forward algorithm to obtain the posterior of the state-dependent distributions.

3.2.3 Application to Bald Eagle Data

The data considered are telemetry observations from ten bald eagles living in the Great Plains of the United States, who have been tagged with GPS receivers. These eagles were tagged shortly after their birth in 2016 and were followed until 2020. These data were provided by Tricia Miller from Conservation Science Global, Inc. and Todd Katzner from the USGS for the purpose of better understand young bald eagle behavior and age of dispersal.

The dataset contains 3,313,548 position readings with the following measured variables: animal id, date, latitude, longitude, speed, heading, altitude, battery power, differences in time, distance and altitude from previous position fix, and binned fix rate.

The hidden Markov model constructed focuses on one bird measured from May of 2016 until August of the same year (a three month period, after which the bird died). The model used only latitude and longitude[4], speed, and altitude. The provided dataset had irregular time steps and contained coordinates with a high horizontal and vertical dilution of precision (HDOP and VDOP). The dataset was narrowed by restricting the data to an HDOP and VDOP of less than 10 (small enough so precision is considered good).

HMMs require regular time steps, thus data imputation was required. Data was imputed with the R package `momentuHMM` (McClintock and Michelot, 2018) using a regular time step of 10 minutes. Once the data was adequately cleaned and prepared, step lengths were further narrowed to less than 30 meters, which accounted for 80% of the prepared data.

After running two chains of 1000 iterations the algorithm obtained reasonable estimates and effective sample sizes (1500 on average, total data size is 13561 observations) with sufficient convergence ($\hat{R}$ values were all very close to 1). Plotting the state-dependent distri-

[4]Transformed to UTM - Universal Transverse Mercator - coordinates in order to work with step lengths in meters instead of kilometers

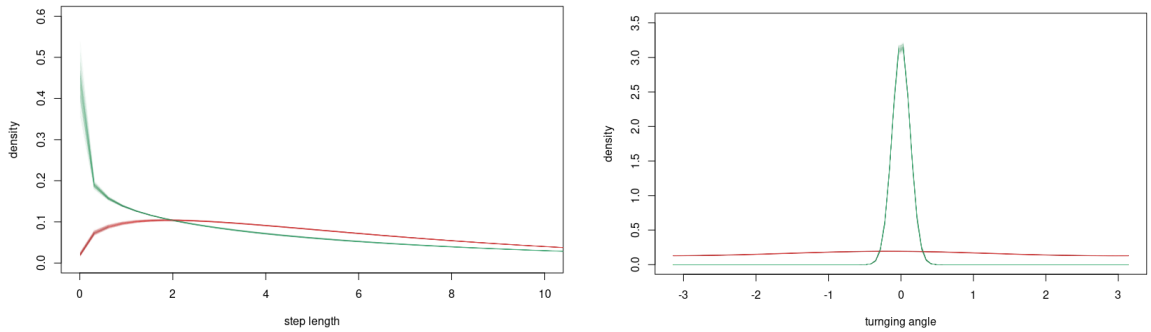

Figure 16: State dependent distributions for step length and turning angle observations.

butions using the parameter estimates from our model for both the step lengths and turning angles, we see two distinct distributions for each variable which should map to two different behavioral modes expressed by this bird (figure 16). The green curves for both step lengths and turning angles describe observations that have a higher density of smaller values, while the red curves are more spread out with similar densities for both small and large values.

Although we would have to implement state decoding to reach a conclusion on the connection between the movement observations and the animal's behavioral modes, we can make a reasonable guess that the green curves describe resting times or more sedentary times, and the red curves describe more active times. Further consideration leads us to conclude that more sedentary times may indicate that the data informing the green distributions come from earlier in this bird's life, since nestlings are quite sedentary, while the red distributions are informed by data from a later time when this bird is starting to explore its surroundings. These guesses can be validated by looking at the time series of the step lengths for the first 3-4 days and comparing to the sequence for a month later (figure 17).

We can check the adequacy of our model's fit by generating 1000 replicates from the posterior predictive distribution $f(\tilde{\mathbf{y}}|\mathbf{y}) = f(\tilde{\mathbf{y}}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathbf{y})$, where $\boldsymbol{\theta} = (\alpha_1, \beta_1, \mu_1, \kappa_1, \alpha_2, \beta_2, \mu_2, \kappa_2)$. Then we use the method `ppc_dens_overlay` from the R package `bayesplot` Gabry and Mahr (2017) to overlay the replicates and empirical distributions on the same plot (figure 18).

The relationship between the state transitions and the chosen covariates here did not reveal any interesting patterns. This is because the variables used as covariates are also time series data measured simultaneously with the positional data.

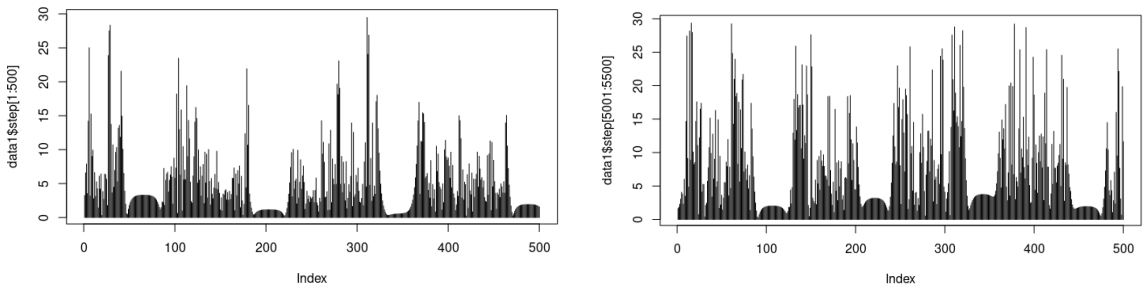

Figure 17: Sequence of step lengths for every 10 minutes starting at the beginning of the bird's life and going up to 5000 minutes or about 3.5 days (left). Sequence of step lengths for every 10 minutes starting about one month after the beginning of the bird's life and going up to 55000 minutes or about 3.5 days after the 35th day (right).

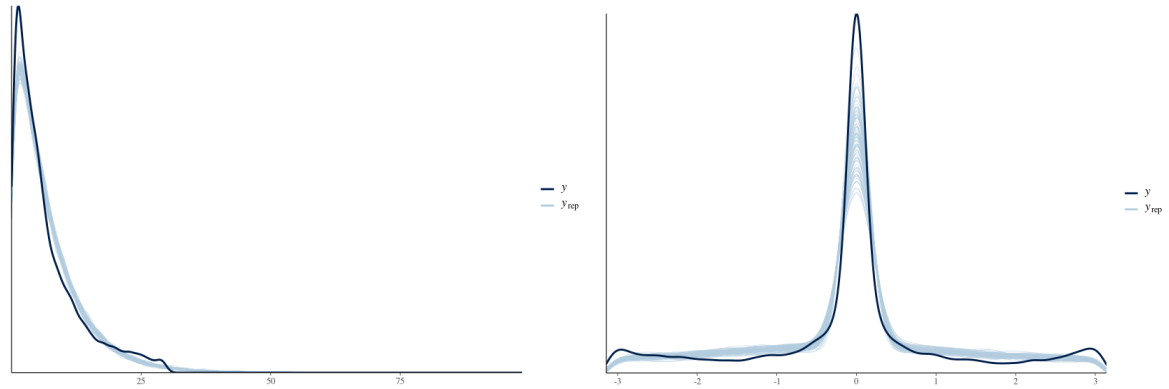

Figure 18: Posterior predictive checks y_{rep} overlayed with the empirical distribution y for the step length and turning angle distributions.

4 Next Steps and Conclusion

The ideas described above represent a very small subset of the possible questions one can ask from the field of statistical ecology. There will always be more questions to ask, and even with the same question, there will always be more to do. Later work will focus on continued investigation of the bald eagle data and the ecological questions they raise.

Overall, while the model described in 3.2 yielded reasonable results, it is still quite simple and only working for a subset of the data that was sanitized of extreme values. There is still work to be done in order to formulate a model that is general, can handle extreme values, can describe more than two states, and can incorporate covariates that are independent from the telemetry data, such as bioclimatic variables. Speed, heading and altitude were taken here to be covariates, however they would be better modeled as further state-dependent distributions. Future work should include:

- States that can account for the large range of step lengths observed in the data.
- Use of environmental and landscape variables as covariates.
- Better data cleaning, imputation, and tuning of initial parameters.
- Hyperparameters to model differences between birds.
- State decoding in order to determine the underlying state sequence that led to the observed data.

Ultimately, I plan to meaningfully reproduce Pirotta et al. (2018)'s model and further refine it into a hierarchical hidden Markov model (HHMM). HHMMs are a generalization of HMMs where hidden states are themselves HHMMs, that is states emit a sequence of symbols which depend on the "top" level hidden states (Fine et al., 1998). For example, given a sequence daily rainfall observations, we would like to determine the climate oscillation that emitted this sequence (fig. 19). However, there are intermediate layers of hidden states, called internal states, which separates the top level states from the states that directly emit the observations, called production states (Fine et al., 1998).

Accounting for this hierarchy is important as it allows us to learn about the structure of the overall system underlying the observations. In the case of animal movement, the observations may be generated from different behavioral processes which occur at distinct time scales, thus

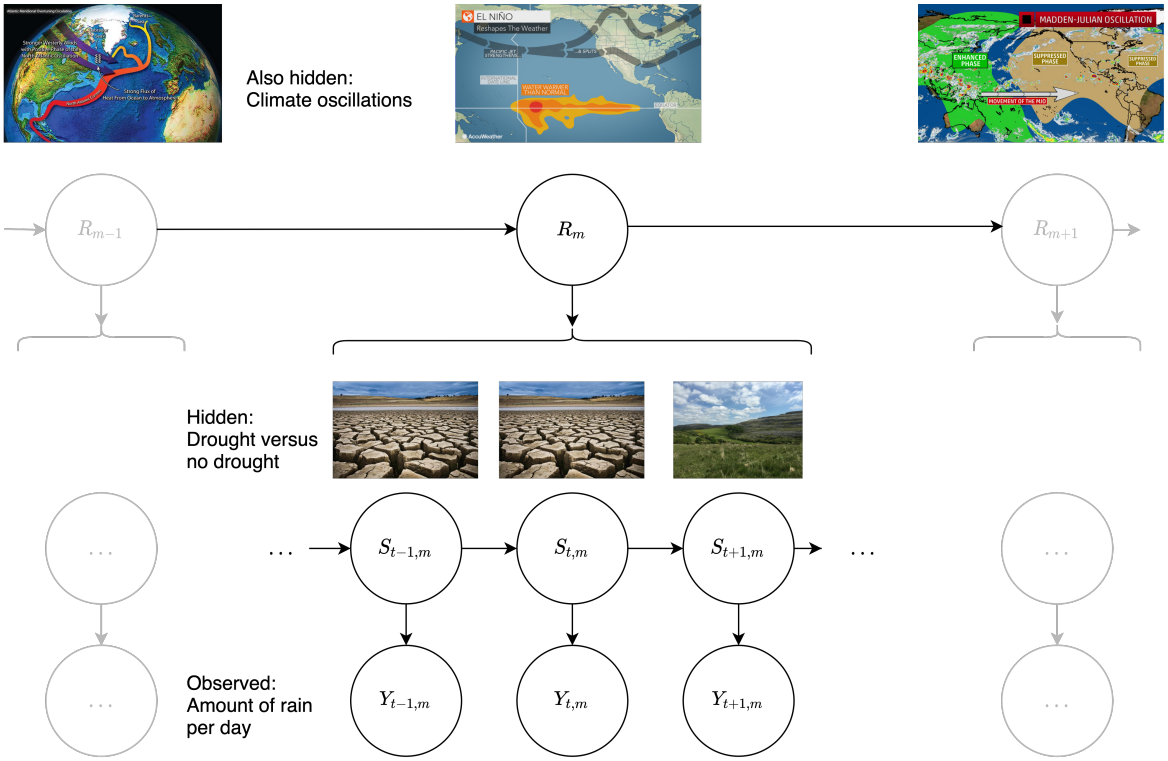

Figure 19: Hierarchical hidden Markov model for hidden climate oscillations at the top level, with inner nodes representing hidden weather states. Observed values are amount of rainfall per day.

understanding fine grained time behavioral modes can be used to understand coarse grained time behaviors (Leos-Barajas et al., 2017). For instance, movements can classify a sequence of flight modes, which in turn classify higher level behaviors or objectives such as fishing or migration.

The initial analysis presented here was a good start and sanity check of our intuition that there are differences in bald eagles movement and behavior through space and time (and therefore throughout their lifespan). Continued exploration of the data and the model described here should yield fruitful results that may help wildlife managers' conservation efforts to protect our national bird. Further, continued investigation of HMMs and their extensions not only can lead to advances in other fields where their application can reveal novel patterns, but also in the Statistical discipline itself as there currently are many computational challenges to address, such as selecting the appropriate number of hidden states (Pohle et al., 2017).

References

Aarts, G., Fieberg, J., and Matthiopoulos, J. (2012). Comparative interpretation of count, presence–absence and point methods for species distribution models. *Methods in Ecology and Evolution*, 3(1):177–187.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., and De Clerck, O. (2018). Bio-oracle v2. 0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, 27(3):277–284.

Bonfil, R., Clarke, S., and Nakano, H. (2008). The biology and ecology of the oceanic whitetip shark, carcharhinus longimanus. *Sharks of the Open Ocean: Biology, Fisheries and Conservation. Blackwell Publishing, Oxford, UK*, pages 128–139.

Clark, A. E. and Altwegg, R. (2019). Efficient bayesian analysis of occupancy models with logit link functions. *Ecology and evolution*, 9(2):756–768.

David, S. V. (2018). Incorporating behavioral and sensory context into spectro-temporal models of auditory encoding. *Hearing Research*, 360:107–123.

Dorazio, R. M. (2014). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography*, 23(12):1472–1484.

Dorazio, R. M. and Rodriguez, D. T. (2012). A gibbs sampler for bayesian analysis of site-occupancy data. *Methods in Ecology and Evolution*, 3(6):1093–1098.

Ephraim, Y. and Merhav, N. (2002). Hidden markov processes. *IEEE Transactions on information theory*, 48(6):1518–1569.

Erickson, W. P., Johnson, G. D., David Jr, P., et al. (2005). A summary and comparison of bird mortality from anthropogenic causes with an emphasis on collisions. In *In: Ralph, C. John; Rich, Terrell D., editors 2005. Bird Conservation Implementation and Integration in the Americas: Proceedings of the Third International Partners in Flight Conference. 2002 March 20-24; Asilomar, California, Volume 2 Gen. Tech. Rep. PSW-GTR-191. Albany, CA: US Dept. of Agriculture, Forest Service, Pacific Southwest Research Station: p. 1029-1042*, volume 191.

Esling, P. and Agon, C. (2012). Time-series data mining. *ACM Computing Surveys (CSUR)*, 45(1):1–34.

Fine, S., Singer, Y., and Tishby, N. (1998). The hierarchical hidden markov model: Analysis and applications. *Machine learning*, 32(1):41–62.

Fithian, W., Elith, J., Hastie, T., and Keith, D. A. (2015). Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, 6(4):424–438.

Fithian, W. and Hastie, T. (2013). Finite-sample equivalence in statistical models for presence-only data. *The annals of applied statistics*, 7(4):1917.

Gabry, J. and Mahr, T. (2017). bayesplot: Plotting for bayesian models. *R package version*, 1(0).

Gimenez, O., Buckland, S. T., Morgan, B. J., Bez, N., Bertrand, S., Choquet, R., Dray, S., Etienne, M.-P., Fewster, R., Gosselin, F., et al. (2014). Statistical ecology comes of age. *Biology letters*, 10(12):20140698.

Green, J. L., Hastings, A., Arzberger, P., Ayala, F. J., Cottingham, K. L., Cuddington, K., Davis, F., Dunne, J. A., Fortin, M.-J., Gerber, L., et al. (2005). Complexity in ecology and conservation: mathematical, statistical, and computational challenges. *BioScience*, 55(6):501–510.

Greenwood, P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal behaviour*, 28(4):1140–1162.

Guillera-Arroita, G., Lahoz-Monfort, J. J., MacKenzie, D. I., Wintle, B. A., and McCarthy, M. A. (2014). Ignoring imperfect detection in biological surveys is dangerous: A response to âĂŸfitting and interpreting occupancy models'. *PLOS ONE*, 9:1–14.

Hefley, T. J. and Hooten, M. B. (2016). Hierarchical species distribution models. *Current Landscape Ecology Reports*, 1(2):87–97.

Hefley, T. J., Tyre, A. J., Baasch, D. M., and Blankenship, E. E. (2013). Nondetection sampling bias in marked presence-only data. *Ecology and evolution*, 3(16):5225–5236.

Hueter, R., Heupel, M., Heist, E., and Keeney, D. (2005). Evidence of philopatry in sharks and implications for the management of shark fisheries. *Journal of northwest atlantic fishery Science*, 35:239–247.

Kéry, M. and Schaub, M. (2011). *Bayesian population analysis using WinBUGS: a hierarchical perspective*. Academic Press.

Leos-Barajas, V., Gangloff, E. J., Adam, T., Langrock, R., Van Beest, F. M., Nabe-Nielsen, J., and Morales, J. M. (2017). Multi-scale modeling of animal movement and general behavior data using hidden markov models with hierarchical structures. *Journal of Agricultural,*

Biological and Environmental Statistics, 22(3):232–248.

Leos-Barajas, V. and Michelot, T. (2018). An introduction to animal movement modeling with hidden markov models using stan for bayesian inference. *arXiv preprint arXiv:1806.10639*.

McClintock, B. T., Langrock, R., Gimenez, O., Cam, E., Borchers, D. L., Glennie, R., and Patterson, T. A. (2020). Uncovering ecological state dynamics with hidden markov models. *Ecology letters*, 23(12):1878–1903.

McClintock, B. T. and Michelot, T. (2018). momentuhmm: R package for generalized hidden markov models of animal movement. *Methods in Ecology and Evolution*, 9(6):1518–1530.

OBIS (2020). Ocean biodiversity information system. intergovernmental oceanographic commission of unesco. `www.iobis.org`.

OSU (2019). Datasets for hands-on portion of occupancy from osu. `http://people.oregonstate.edu/~peterjam/occupancy_workshop/hands_on.html`. Accessed: 2019-12-07.

Pascual, M. (2005). Computational ecology: from the complex to the simple and back. *PLoS Comput Biol*, 1(2):e18.

Pirotta, E., Katzner, T., Miller, T. A., Duerr, A. E., Braham, M. A., and New, L. (2018). State-space modelling of the flight behaviour of a soaring bird provides new insights to migratory strategies. *Functional ecology*, 32(9):2205–2215.

Pohle, J., Langrock, R., van Beest, F., and Schmidt, N. M. (2017). Selecting the number of states in hidden markov models-pitfalls, practical challenges and pragmatic solutions. *arXiv preprint arXiv:1701.08673*.

Rabiner, L. and Juang, B. (1986). An introduction to hidden markov models. *ieee assp magazine*, 3(1):4–16.

Renner, I. W. and Warton, D. I. (2013). Equivalence of maxent and poisson point process models for species distribution modeling in ecology. *Biometrics*, 69(1):274–281.

Rigby, C., Barreto, R., Carlson, J., Fernando, D., Fordham, S., Francis, M., Herman, K., Jabado, R., Liu, K., Marshall, A., et al. (2020). Carcharhinus longimanus. the iucn red list of threatened species 2019: e. t39374a2911619.

Sergio, F., Tanferna, A., De Stephanis, R., Jiménez, L. L., Blas, J., Tavecchia, G., Preatoni, D., and Hiraldo, F. (2014). Individual improvements and selective mortality shape lifelong migratory performance. *Nature*, 515(7527):410–413.

Somveille, M., Rodrigues, A. S., and Manica, A. (2015). Why do birds migrate? a macroecological perspective. *Global Ecology and Biogeography*, 24(6):664–674.

Stevens, M. and Merilaita, S. (2011). *Animal camouflage: mechanisms and function*. Cambridge University Press.

Strasburg, D. W. (1958). Distribution, abundance, and habits of pelagic sharks in the central pacific ocean. *Fisheries*, 1:2S.

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., and De Clerck, O. (2012). Bio-oracle: a global environmental dataset for marine species distribution modelling. *Global ecology and biogeography*, 21(2):272–281.

Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., and Veran, S. (2013). Presence-only modelling using maxent: when can we trust the inferences? *Methods in Ecology and Evolution*, 4(3):236–243.

Young, C. N. and Carlson, J. K. (2020). The biology and conservation status of the oceanic whitetip shark (carcharhinus longimanus) and future directions for recovery. *Reviews in Fish Biology and Fisheries*, pages 1–20.