

Chapter 2

Literature Review

2.1 Introduction to Population Biology

Population biology is the study of biological populations (Hastings, 2013). The goal of the study of population biology is to identify why a species populations fluctuate the way that they do. Central to this goal is the creation of theoretical models used to model the fluctuations of a species population (Royama, 2012). In the context of species conservation, accurate population models are vital. These models, if correctly engineered, can be used to analyse historical data, and extrapolated to future dates.

There are 4 primary factors that contribute to the change in population for most species. These include births and immigration, which increase a species population, and deaths and emigration, which decrease a species population. These 4 primary factors contribute directly to the intrinsic growth rate of the population. There are then further, secondary, factors that influence the four primary factors. These include population density, distribution, sex ratio and age structure. Furthermore, the environments carrying capacity also influences the 4 primary factors. All these factors, directly or indirectly, affect the population dynamics of the species exposed to them. The ways in which species interact with their ecosystem can be very complex and it's almost impossible to account for all the variables that influence a species population dynamic.

Nevertheless, population models are still useful, albeit not all encompassing. There are many examples of positive outcomes due to population models. One example is how population dynamics modelling played an important role in reversing the decline in the endangered Loggerhead Sea Turtle in the US (Crouse et al., 1987; Crowder et al., 1994). Initial conservation efforts, in the 1980s, had been focused on protecting the turtle nests and hatchlings. However, it was later discovered, using population models with a focus on species age, that the declining growth rate of these turtles were attributed to subadult and adult turtles at sea. These findings lead to the use of turtle excluder devices (TEDs) in shrimp trawls in turtle habitats and has since had a substantially positive impact on the growth rate of the Loggerhead turtles.

Population models have also been used to manage decisions regarding over-abundant species in an ecosystem. The American bullfrog is an alien species on Vancouver Island. These bullfrogs have been rapidly increasing and negatively affecting biodiversity in its ecosystem. A matrix model investigation conducted by Govindarajulu et al, (2005) found that the current strategy of removing tadpoles only lead to higher tadpole survival due to reduced competition in a density-dependent context. Instead, they reported that the culling

of tadpoles in autumn would provide the most effective approach in managing the over-abundant bullfrogs.

As is the case with all models, “all models are wrong, but some are useful”. The complexity of species interactions with each other and their environment make it impossible for any population modeller to account for all variables in their model. However, as proved above, models can still be very useful in species conservation when the correct variables are accounted for, and the correct model used.

2.2 State Space Models

State space models (SSMs) have become a popular tool used by ecologists to analyse ecological time-series. There is enough evidence to argue that SSMs should be the default tool in analysing time series for ecologists (Auger-Méthé et al., 2021). SSMs have been used in various ecological studies such as fisheries stock assessment (Aeberhard et al., n.d.), movement ecology (Patterson et al., 2008) and, most commonly, modelling population dynamics (Newman et al., 2014). These are only a few examples of where SSMs are used. Their application in ecology is vast.

SSMs are able to model temporal autocorrelation in such a way that it separates observation error from process error (Auger-Méthé et al., 2021). Herein lies the value of SSMs. SSMs are hierarchical models that essentially model two different time series. One being the state process (latent process, hidden process) which is the unobserved time series that represents the true underlying population dynamics. The other being the observation process that represents the observed counts of that state process. For example, in a situation where you are analysing the population dynamics of bird species, the actual bird population would be the state process and the bird counts would be the observation process. The bird counts would, of course, be subject to observation error. Therefore, separating the observation process from the state process provides a more accurate estimation of the true population. The beauty in this methodology is that variation attributed to population size, population growth rate, environmental stochasticity and any other forms of stochasticity can be estimated and programmed into the model (Kéry & Schaub, 2011). This allows the ecologist to program the SSM time series such that it accounts for known biological factors that affect the population dynamics. This includes birth rates, death rates, fecundity, immigration, emigration, etc. **Figure 1** shows a visual illustration of this separation between state and observation processes.

When using SSMs one must take a few assumptions into consideration:

1. SSMs can only produce unbiased estimates of population size if false negative (detection probability) and false positive (double counting) observations cancel out on average.
2. SSMs can only produce unbiased estimates if detection probability is less than 1 and has no discernible pattern over time.

3. Unbiased estimates cannot be produced if there are temporal patterns in detection probability or false positive rates.
(Kéry & Schaub, 2011)

Types of SSMs

There are many ways to develop an SSM. The key aspects to determine the structure of the model is based on your data and the requirements of the research. This section explores and scrutinizes some of the SSM models that can be used in the context of avian population dynamics.

Simple Exponential Model

The first model to consider is the simplest. SSMs, when applied to population dynamics, exhibits a Markovian process. This means the population size at time $t+1$ is dependent on the population size at time t . A simple model to describe the SSM is the exponential population dynamics model:

$$N_{t+1} = N_t \lambda_t \quad (1)$$

Where λ_t refers to the population growth rate at time t , T refers to the number of years in the time series. N_t refers to the population size at time t . Kery and Schaub (2011) use this model as the basis to understand SSMs.

The fundamental goal of population dynamics analysis is to calculate the population growth rate and determine the factors affecting it. This is a straightforward task when the true population is known each year. With a true population size the ecologist can determine factors such as the strength of density dependence and the influence of the environment on the population. However, knowing the true population of a species is rarely possible depending on the species being monitored. Instead, the ecologist only has raw counts which are highly susceptible to observation error. This observation error must be accounted for in the model. If not, the ecologist could misinterpret observation error for some form of biological variation.

SSMs deal with observation error as it separates the state process from the observation process. Equation 1 represents the state process:

$$N_{t+1} = N_t \lambda_t \quad (1)$$

$$\lambda_t \sim \text{Normal}(\bar{\lambda}, \sigma_{\lambda}^2) \quad (2)$$

In the case of the SSM, the population growth rate is represented as a Normal random process. The mean, $\bar{\lambda}$, is the long-term growth rate of the population. The variance, σ_{λ}^2 , represents the stochasticity of the growth rate caused by environmental variability. The first

year, N_1 , isn't defined in equation 1 but modelled as a prior distribution in a Bayesian framework or it can be fixed to the initial count.

The next set of equations needed is the observations. The observations are dependent on the state process and are defined as follows:

$$y_t = N_t + \varepsilon_t \quad (4)$$

$$\varepsilon_t \sim \text{Normal}(0, \sigma_y^2) \quad (5)$$

Here y_t represents the counts observed from the true population N_t . ε_t refers to the observation error with mean 0 and variance σ_y^2 (observation error). The model assumes that λ_t and ε_t are serially independent. That is, they are independent from each other and both are identically distributed.

These equations display the flexibility of the SSM. The state process can be modified to include processes such as survival rate, birth rate, immigration, etc. Or one could use a logarithmic model instead of an exponential one. The observation process can be modified to account for different types of observation errors such as binomial or poisson errors.

SSM for migrant and non-migrant avian species

The simple model, described above, is very useful as it implicitly accounts for various biological and environmental variation. Given the flexibility of the model, one can modify the model to account for known biological processes that affect population variation. This works well when modelling migratory avian species (Barshep, 2017; Camp et al., 2016; Wilson et al., 2011). Barshep (2017) uses SSMs on data contained in the CWAC dataset. The aim of the study was to model population changes of 29 migrant species and 25 non-migrant species then identify which life-history traits was most responsible for these fluctuations. Two different SSMs were used. Both were logarithmic models. Model one, for resident, "sedentary residents with limited or no seasonal movement between sites". Model 2, for migrants, "Palearctic migrants and intra-African migrants".

Model 1 is a univariate model that has summer/winter as a covariate. The observation process is modelled as followed:

$$\ln(y_t) = \mu_t + \gamma_t + \alpha_t, \alpha_t \sim N(0, \sigma_\alpha^2) \quad (6)$$

Where $\ln(y_t)$ is the logarithm of the count, μ_t is the true population, γ_t is the summer/winter covariate and α_t is the observation error.

The state process was modelled as a linear trend model as the resident species tend to remain in the same place all year round. The state process is represented as follows:

$$\ln(\text{relative population size at time } t): \mu_t = \mu_{t-1} + \beta_t - 1 + \omega_t, \omega_t \sim N(0, \sigma_\omega^2) \quad (7)$$

where β_t is the rate of change of the population over time. γ_t and β_t is modelled as follows:

$$\ln(\text{population change}): \beta_t = \beta_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (8)$$

$$\ln(\text{seasonal effect}): \gamma_t = -\gamma_{t-1} + \zeta_t, \zeta_t \sim N(0, \sigma_\zeta^2) \quad (9)$$

The variables α_t , w_t , ϵ_t and ζ_t are all serially independent. This concludes model one that is applied to resident species. See [figure 2](#) for a visual illustration of model 1.

Model 2 is applied to migrant species and considers their migratory movement. The migrant species in South African wetlands tend to breed in the Northern Hemisphere (during the southern winter) then migrate south for the southern summer. The young species tend to remain in the south during the subsequent southern winter period. Model 2 was thus formulated as a bivariate time series model with the summer population in the state process as the main focus. The proportion of species that stay over for the subsequent southern winter was modelled as the second covariate. [Figure 3](#) shows a visual description of this model. The mathematical representation of model 2 is as follows:

Observation process

$$\ln(\text{summer count}): s_t = \mu_t + \alpha_t, \alpha_t \sim N(0, \sigma_\alpha^2) \quad (10)$$

$$\ln(\text{winter count}): \omega_t = \mu_t + \lambda_t + e_t, e_t \sim N(0, \sigma_e^2) \quad (11)$$

where λ_t is the proportion of migrant species that stay over from summer to winter.

State process

$$\ln(\text{summer population change}): \mu_t = \mu_{t-1} + \beta_{t-1} + \omega_t, \omega_t \sim N(0, \sigma_w^2) \quad (12)$$

$$\ln(\text{population change}): \beta_t = \beta_{t-1} + \zeta_t, \zeta_t \sim N(0, \sigma_\zeta^2) \quad (13)$$

$$\ln(\text{winter to summer ratio}): \lambda_t = \lambda_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (14)$$

Here, again, α_t , e_t , w_t , ζ_t and ϵ_t are all serially independent.

These models may be highly simplified and could still explicitly account for factors such as birth rate, death rate, density dependence, etc. However, it is difficult to obtain this data and to account for environmental factors for each wetland. The SSM model is beneficial to use in this case as all the unknown ecological variability is captured in the error terms in the state process.

Inferential approaches to SSM

SSMs are flexible in that it allows for many types of biological processes to be coded into the model. SSMs are also flexible in terms of how it's fit. There are 2 fundamentally different ways to fit SSMs. That is a frequentist approach or a Bayesian approach (Auger-Méthé et al., 2021).

Frequentist approach

The frequentist, or classical, approach is common in most statistical analysis. The frequentist approach focuses on long run averages and p-values to determine the probability of obtaining a realised outcome, or something more extreme, given the null hypothesis (Bolker, 2008). If the outcome probability is significantly small you reject the null hypothesis, otherwise you cannot reject it. According to Bolker (2008), “the classical framework says you can never prove that something is true, you can only fail to prove it is false.”

The frequentist approach is widely used in statistics but has a few drawbacks. Firstly, the framework relies on hypothetical probabilities of outcomes based on the experiment. The p-value is also a contentious issue as it is often misused in ecological research (Yoccoz, 1991).

When fitting SSMs according to the frequentist framework, we essentially perform maximum likelihood estimation on the joint likelihood of the parameters, θ , and the states, z . The mathematical structure can be illustrated as follows:

$$L(\theta, z_{1:T} | y_{1:T}) = \prod_{t=1}^T g(y_t | z_t, \theta_0) f(z_t | z_{t-1}, \theta_p) \quad (15)$$

Where θ_0 and θ_p represent the observation process parameters and the state process parameters respectively. And T representing the length of the time series. To simplify the maximum likelihood process, one would first maximise the marginal likelihood and integrate out the hidden states (z). Thus, only having to maximise the marginal likelihood with respect to the parameters (θ). To estimate the hidden states, one would use the predicted parameters and calculate the conditional distribution of the hidden states given the observations (y) and the estimated parameters ($\hat{\theta}$). Calculating the maximum marginal likelihood with respect to the parameters is a computationally expensive problem. This is the common pitfall for SSMs in general, for both the frequentist and Bayesian approach.

There are two common methods one can use to evaluate the maximum marginal likelihood. Namely the Kalman Filter and the Laplace approximation method (which approximates the maximum marginal likelihood as opposed to evaluating it).

Bayesian approach

While the frequentist approach is based on hypothetical probabilities, the Bayesian approach is based on prior distributions. Where the frequentist method defines unknown model parameters as long-term averages, the Bayesian approach defines unknown model parameters as probability distributions. These distributions are based on expert knowledge.

The Bayesian approach resolves the conceptual issues brought about in the frequentist approach. Instead of hypothetical probabilities, the Bayesian approach focuses on the data at hand. This way the statistician can make direct statements about the probability of hypotheses or parameter values without the use of contentious p-values (Bolker, 2008).

Bayesian approaches can also encapsulate expert knowledge in the formulation of its priors. This could be both a good thing and a bad thing. These priors affect the outcome of the

model and has been used as a criticism of the Bayesian framework (Dennis, 1996). However, if the expert knowledge is sound, it would greatly benefit the model and will lead to favourable and more biologically sound outcomes. If incorrect, the ecologist may be adding erroneous biological or ecological beliefs into the model. Thus, producing an outcome that does not relate to reality.

Given the criticism of subjective priors (Lele & Dennis, 2009), the least controversial way for choosing priors is, therefore, to use non-informative priors as far as possible. These non-informative priors refer to normal or uniform priors. However, it is also believed that using only non-informative priors limits the benefits the Bayesian approach has to offer (Banner et al., 2020).

Besides the use of priors being a contentious topic, the Bayesian approach still holds a vast array of benefits. Bayesian models are conceptually easier to understand when applied to hierarchical models (Kéry & Schaub, 2011). In population dynamics, there is often a true population and observations based on that true population. These observations very seldom reflect the true population and are often prone to observation error. Given this hierarchical structure, the Bayesian approach is a more intuitive method to use.

When applying SSMs in a Bayesian context, the main goal is to determine the posterior distribution for the states and the parameters based on their prior distributions. The mathematical representation of this approach is as follows:

$$p(\theta, z_{1:T} | y_{1:T}, \lambda) = \frac{L(\theta, z_{1:T} | y_{1:T}) \pi(\theta | \lambda)}{\int \int L(\theta, z_{1:T} | y_{1:T}) \pi(\theta | \lambda) dz_{1:T} d\theta} \quad (16)$$

Where $L(\theta, z_{1:T} | y_{1:T})$ is the joint likelihood and $\pi(\theta | \lambda)$ is the prior distribution(s). The denominator of equation 16 represents the probability of the data for all possible values of the states and parameters. Equation 16 forms the basic structure of the bayes theorem:

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{y} \quad (17)$$

Where y refers to the observed data and θ refers to the state and the parameters.

The posterior distribution is the focus in the Bayesian approach. This is done by obtaining random draws from the right-hand side of equation 16. This is a computationally expensive task that is impossible to do by hand. There are various algorithms that approximate the posterior distribution. Namely, Metropolis Hastings samplers, Hamiltonian Monte Carlo and the Gibbs sampler, amongst many. These algorithms all fall under the broader Markov Chain Monte Carlo (MCMC) algorithm class. MCMC are methods that approximate the posterior distribution by sampling from a probabilistic distribution (b, 2017).

Another important aspect in the Bayesian approach is to apply convergence tests on the MCMC outcome, regardless of the sampler used. Testing convergence ensures that the sampler used, after approximating the probabilistic distribution many times, has converged to a common interval across all MCMC chains. This can often be viewed on a trace plot of

MCMC outcomes (Auger-Méthé et al., 2021). A more formal approach would be the Gelman-Rubin metric (Gelman & Rubin, 1992), which is a popular metric to measure convergence in an MCMC context.

Model Selection

A major drawback for SSMs is the issue of model selection. Model selection for SSM is still an area of active research (Jonsen et al., 2013). SSM model selection becomes difficult as the computation power needed to run an SSM is so large. Often ecologists end up using only one SSM model in the study because of limited computational power.

Model comparison methods will differ depending on whether the SSM is conducted in a Frequentist framework or a Bayesian framework. The most used model selection method for a frequentist SSM is the Akaike Information Criterion (AIC). The AIC is a popular model selection method in many statistical models. The AIC measures how well the model fits the data and penalizes models with a higher number of parameters. AIC, however, doesn't work well for SSMs, even in a frequentist framework, for various reasons (Auger-Méthé et al., 2021).

For SSMs in a Bayesian context, AIC is not well defined (Hooten & Hobbs, 2015). There are 3 notable model selection methods used for Bayesian SSMs. These are the WAIC, DIC and BIC. Many sources of literature have proven DIC and BIC to not be very good at all in Bayesian model selection (Auger-Méthé et al., 2021; Gelman et al., 2014; Hooten & Hobbs, 2015). Sources have found WAIC to be more effective than DIC and BIC, even though it still fails in many cases (Gelman et al., 2014). The method most of the literature agreed upon was a cross validation method. In this cross-validation method, the SSM would be fit with a training set of the data, while an unused subset of the data (the test set) is used to determine the effectiveness of the SSMs forecasting capacity. The SSM that forecasts values closest to the corresponding test set values is then considered the "best" model. This method works well but has its limitations when data is sparse. In the case of sparsity, splitting the data into a train and test set becomes difficult as the train set may not be sufficient to produce an accurate model. Furthermore, the cross-validation method for SSMs is computationally expensive.

Yoccoz (1991) and Cox et al. (1977), states that, at times, ecologists tend to overemphasize the use of statistical measures for model selection and fail to prioritise biological or ecological significance. Given the contentious issues regarding model selection for SSMs, it may be of benefit to choose the SSM that is based on sound ecological fundamentals rather than contentious model selection statistics. More time and effort should be focused on the modelling process and ensuring the correct ecological processes are modelled into the SSM.

2.3 Bioindicators

Fitting SSMs to population data helps the general population see how the species population has fluctuated over the years. But these models don't tell the viewer much

about the environment the species are in or the biodiversity of the ecosystem. This is where bioindicators come in. The second objective of this study is to develop bioindicators using the SSMS applied to the CWAC data. This section looks at bioindicators in use today and their purposes. Which range from indicating the species environment's health to population abundance aggregation.

2.3.1 Birds as bioindicators

An ecosystem's health is based on a variety of factors, many of which may be difficult to monitor. However, there are only few factors that can provide an overview of ecosystem health. Birds are perfect organisms to use in this regard as they reflect various environmental change directly (Mekonen, 2017). According to Gregory et al. (2003), Effective bioindicators are quantitative, easy to collect, simple to interpret, reflects causes of change in an ecosystem and indicates general patterns of taxa in the ecosystem. Bird count data fits this well. As bird counts are relatively easy to collect (and readily available), quantitative, simple to interpret and they reflect many different changes in an ecosystem.

Birds have been used to indicate the eutrophication levels in wetlands in the Mar Menor lagoon of south-eastern Spain (Amat & Green, 2010). It was discovered that the Great Crested Grebe abundance increased as the eutrophication levels in the lagoon increased. The increase in eutrophication was a result of nutrient run-off from nearby agricultural practices. Another example from Southern Spain shows the use of Red-Knobbed Coots to indicate the siltation rate in the water and soil in wetlands. The Red-Knobbed Coot experienced a stark decline in abundance in the 20th century in Southern Spain. This decrease in abundance is attributed to the increased siltation rates in the water and soil in the wetlands in Southern Spain. Caused by invasive agricultural practices. The increased rates of siltation negatively affected the quality of food for the Coots in their wetlands, and thus lead to their demise. Thus, the Red-Knobbed Coots population over time could be indicative of siltation rates in the wetlands in which they are based. These are only two examples of how effective waterbirds are as bioindicators. Other examples show how waterbirds can be indicators for changes in water level in a wetland (Burgess, 1969) and species abundance at lower trophic levels (Mekonen, 2017).

An important aspect of ecosystem health is biodiversity. Rich biodiversity is indicative of a healthy ecosystem. Each species has their part to play in the ecosystem with varying degrees of importance. Tracking biodiversity in an ecosystem can also indicate the presence of invasive species (Hooper et al., 2005). A biodiversity indicator is therefore essential in tracking the overall health of a wetland ecosystem. Thus, for the purposes of indicating wetland health, it's important to have bioindicators that are indicative of the waterbird's population over time and their biodiversity over time. These factors can be used by wetland conservation decision makers to determine various aspects of wetland health. It can also be used by waterbird conservationists to identify the stability of waterbird species in South Africa.

2.3.2 Biodiversity Indicators

There are multiple different ways one can formulate biodiversity indices. The first, and main task, however, is to identify what must be monitored and how (Yoccoz et al., 2001). In the South African wetland's context, one would want to monitor species that represent causal links. This means identifying and monitoring a species that reacts directly to specific environmental changes. In other words, it follows the pressure – state – response causal change (Niemeijer & de Groot, 2008). As previously discussed, species like the Great Crested Grebe (eutrophication level indicator) and Red-Knobbed Coots (siltation rate indicator) provide known causal links to environmental changes.

According to Magurran (2013), biodiversity is the “variety and abundance of species in a defined unit of study”. Buckland et al. (2005) adds one more aspect which is evenness. Evenness refers to the level of homogeneity of the ecosystem, regarding its species. For example, a wetland with 5 different types of waterbird species with each type containing 100 individuals each would be an exhibitor of perfect evenness. The research gives rise to two different types of indicators. Ones to monitor the variety of waterbird species in a wetland and one to monitor the abundance of waterbird species in a wetland.

2.3.2.1 Abundance Indicators

There are currently various indicators used, worldwide, that aggregates and measures species trends. These are used to provide an overview of the species health in an ecosystem. The main index in use, globally, is the Living Planet Index (LPI). There are other similar indices such as the Australian Threatened Species Index (TSX) and the Wild Bird Index (WBI). TSX uses a geometric mean of single species trends to form a composite index. The TSX also has a multispecies index that uses the LPI approach. The WBI also bases its index calculations on the LPI. The only difference being that, while the LPI is applied to any source of sound bird data, while the WBI uses bird data from formally designed surveys only.

Living Planet Index

The LPI is the most used index that monitors average vertebrate trends around the world. The LPI project was initiated in 1997 by the WWF and the World Conservation Monitoring Centre. The LPI measures vertebrate trends from as far back as 1970. The LPI uses time-series data from around the world and aims to measure the state of the world's population trends of vertebrate species. The Index covers the terrestrial, freshwater and marine habitats. The time series data is further broken up into its specific biomes and geographic realms or oceans.

There have been criticisms of the LPI as it tends to over represent birds and mammals and temperate realms (Collen et al., 2009). In response, the LPI was modified to include a proportionally weighted index (McRae et al., 2017). The weighted LPI tackles the problem of taxonomic and geographical bias. The weighted LPI is applied to the problem of bias that presents itself on a global level.

The LPI is calculated as follows:

$$\bar{d}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} d_{it} \quad (18)$$

Where \bar{d}_t is the average annual trends if there were multiple population time series of the same species. n_t refers to the number of populations of the species. d_{it} , the annual trend of a species, is modelled as follows:

$$d_t = \log_{10} \left(\frac{N_t}{N_{t-1}} \right) \quad (19)$$

Here, N_t refers to the population size of a species at time t . d_t refers to the trend of the species. After constructing species, group, regional or global trends, these are converted to an index as follows:

$$I_t = I_{t-1} * 10^{\bar{d}_t}, I_0 = 1 \quad (20)$$

The LPI is conducted using two different methods. One being the LPI-U (unweighted LPI) and the LPI-D (proportionally weighted LPI). In the LPI-U, the data is divided into 6 subsets. Based on region (tropical or temperate) and three systems (terrestrial, freshwater and marine). The index for each system is calculated by averaging the species trends within each system. The systems being (tropical terrestrial, tropical freshwater, temperate terrestrial, etc.). The index for each region (tropical or temperate) is then calculated by averaging the trends for each system. The tropical and temperate trends are then averaged to form a global trend which is converted into an index using equation 20.

The LPI-D approach is similar. Only it applies proportional weights to species when averaging the trends. For example, if birds represent 43.3% of terrestrial vertebrate species in the Palearctic realm, then this value is used in weighting the bird species when calculating the weighted average of the Palearctic realm trend for terrestrial species. The mathematical representation is as follows:

$$\bar{d}_{t, \text{realm.region}} = \frac{1}{N_T} \sum_{j=1}^{N_T} d_{jt} \cdot w_j \quad (21)$$

Where \bar{d}_t is the average trend of a given realm and region. N_T is the number of taxonomic groups within a realm, d_{jt} is the average trend of taxonomic group j at time t and w_j is the proportional weighting of that taxonomic group (0.43 according to the previous example).

Understanding waterbirds population movement over time, in South African wetlands can tell us a lot about the wetland. As discussed previously, some species populations may indicate certain wetland characteristics such as eutrophication levels and siltation rates. Monitoring the combined number of waterbird species can indicate when there are sharp declines or inclines in species abundance. These sharp turning points could indicate that the wetland is changing in some way and thus alert wetland managements to further assess the wetland and take necessary action.

2.3.2.2 Diversity Indicators

The diversity of species in an ecosystem is also a very important aspect of a healthy ecosystem. Richer species diversity, or higher levels of evenness, is indicative of a healthy ecosystem (Hooper et al., 2005). Diversity has been another contentious measure in ecology (Peet, 1974). The literature emphasizes that diversity is dependent on the research questions and the task at hand and there is no definite definition of what diversity is (Hurlbert, 1971).

In the context of this report, which focuses on South African wetlands and waterbirds, the aim is to monitor the evenness in the bird species in a wetland over time. The popular indices used to measure species evenness is the Shannon Entropy and the Simpson Index. There have been studies aimed at finding the best index for species diversity. The result being that one can't identify one all-encompassing index, but using a few together gives one a better overview of the ecosystem (Morris et al., 2014).

The closest framework there is to an all-encompassing diversity index is the Hill Numbers model (Hill, 1973). The Hill Numbers framework provides a range of measures of the effective number of species in a system. Each one differing in their propensity to add more, or less emphasis on the rarer species. The Hill Numbers general formula is as follows:

$$N_{\alpha} = (p_1^{\alpha} + p_2^{\alpha} + \dots + p_n^{\alpha})^{1/(1-\alpha)} \quad (22)$$

Where N_{α} is the diversity number of order alpha. These diversity numbers reflect an estimate of the effective number of species present in a sample. The p 's refer to the proportions of species in the sample and n represents the number of species in the sample. The first three diversity numbers are the most common diversity measures. N_0 is just the number of species in a sample. N_1 approximates to the Shannon Entropy (proof is supplied by Hill). The more general formula for the Shannon Entropy is as follows:

$$x = -\sum_{i=1}^S p_i \ln p_i \quad (23)$$

Where x is the index, p_i is the proportion of species i in the sample and S is the number of species in the sample.

The second diversity number, N_2 is the Simpson's Index. The more common general formula for the Simpson's Index is as follows:

$$x = \sum_{i=1}^S p_i^2 \quad (24)$$

According to Hill (1973), different indices measure different aspects of diversity. The Simpson's Index is more sensitive to the abundance of common species in an ecosystem. For this reason, it is also referred to as the measure of dominance concentration. Whereas the Shannon Entropy focuses more on species evenness in the ecosystem. The Shannon Index is a measure of entropy in an ecosystem (Jost, 2006).

According to Jost (2006), the problem with diversity indices is the way in which researchers use them. For example, the Shannon's index is not diversity measures in itself. The Shannon's Index, instead, is a measure of entropy in an ecosystem. Thus, the index should first be converted to a true diversity (effective number of species in an ecosystem) before any comments can be made regarding evenness in the ecosystem. Jost, uses the example of other fields such as physics, economics and information theory. It's common practice in all these fields to convert entropy values to the effective number of elements in a system instead of reporting on the entropy index value. One simply needs to exponentiate the Shannon Entropy value to determine the effective number of species in the ecosystem.

Further problems are addressed regarding diversity indices in ecology. Measuring diversity in a system has many differing semantic and technical aspects. Given the different aspects of diversity, and the plethora of possible indices to measure them, some have argued that diversity measures are a "nonconcept" (Hurlbert, 1971). However, given the variety of measures one can use, the issue is perhaps not the indices but how to choose which ones to use.

Whittaker (1965) argues that the best way to use diversity indices is to use various indices simultaneously, as each index can tell the user different aspects of the diversity in the system. In terms of the Shannon and Simpson index, the Shannon Entropy, when converted to effective number of species tells the user about the evenness in the ecosystem. The Simpson index, on the other hand, tells the user about the dominance concentration of the ecosystem and will identify if there are dominant species and could also indicate the presence of an alien species.

References

- Aeberhard, W. H., Cantoni, E., Field, C., Künsch, H. R., Mills Flemming, J., & Xu, X. (n.d.). Robust estimation for discrete-time state space models. *Scandinavian Journal of Statistics*, n/a(n/a). <https://doi.org/10.1111/sjos.12482>
- Amat, J. A., & Green, A. J. (2010). Waterbirds as Bioindicators of Environmental Conditions. In C. Hurford, M. Schneider, & I. Cowx (Eds.), *Conservation Monitoring in Freshwater Habitats: A Practical Guide and Case Studies* (pp. 45–52). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9278-7_5
- Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R., King, A. A., Leos-Barajas, V., Flemming, J. M., Nielsen, A., Petris, G., & Thomas, L. (2021). A guide to state-space modeling of ecological time series. *ArXiv:2002.02001 [q-Bio, Stat]*. <http://arxiv.org/abs/2002.02001>
- b. (2017, December 24). *A Zero-Math Introduction to Markov Chain Monte Carlo Methods*. Medium. <https://towardsdatascience.com/a-zero-math-introduction-to-markov-chain-monte-carlo-methods-dcba889e0c50>
- Banner, K. M., Irvine, K. M., & Rodhouse, T. J. (2020). The use of Bayesian priors in Ecology: The good, the bad and the not great. *Methods in Ecology and Evolution*, 11(8), 882–889. <https://doi.org/10.1111/2041-210X.13407>
- Barshep, Y. (2017). Identifying ecological and life-history drivers of population dynamics of wetland birds in South Africa. *Global Ecology and Conservation*, 12.
- Bolker, B. M. (2008). Ecological Models and Data in R. In *Ecological Models and Data in R*. Princeton University Press. <https://doi.org/10.1515/9781400840908>
- Bozeman Science. (2015, October 1). *Population Ecology*. <https://www.youtube.com/watch?v=PQ-CQ3CQE3g>
- Buckland, S. T., Magurran, A. E., Green, R. E., & Fewster, R. M. (2005). Monitoring change in biodiversity through composite indices. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1454), 243–254. <https://doi.org/10.1098/rstb.2004.1589>
- Burgess, H. H. (1969). Habitat Management on a Mid-Continent Waterfowl Refuge. *The Journal of Wildlife Management*, 33(4), 843–847. <https://doi.org/10.2307/3799315>
- Camp, R. J., Brinck, K. W., Gorresen, P. M., & Paxton, E. H. (2016). Evaluating abundance and trends in a Hawaiian avian community using state-space analysis. *Bird Conservation International*, 26(2), 225–242. <https://doi.org/10.1017/S0959270915000088>
- Collen, B., Loh, J., Whitmee, S., McRAE, L., Amin, R., & Baillie, J. E. M. (2009). Monitoring Change in Vertebrate Abundance: The Living Planet Index. *Conservation Biology*, 23(2), 317–327. <https://doi.org/10.1111/j.1523-1739.2008.01117.x>

Cox, D. R., Spjøtvoll, E., Johansen, S., van Zwet, W. R., Bithell, J. F., Barndorff-Nielsen, O., & Keuls, M. (1977). The Role of Significance Tests [with Discussion and Reply]. *Scandinavian Journal of Statistics*, 4(2), 49–70.

Crouse, D. T., Crowder, L. B., & Caswell, H. (1987). A Stage-Based Population Model for Loggerhead Sea Turtles and Implications for Conservation. *Ecology*, 68(5), 1412–1423. <https://doi.org/10.2307/1939225>

Crowder, L. B., Crouse, D. T., Heppell, S. S., & Martin, T. H. (1994). Predicting the Impact of Turtle Excluder Devices on Loggerhead Sea Turtle Populations. *Ecological Applications*, 4(3), 437–445. <https://doi.org/10.2307/1941948>

Dennis, B. (1996). Discussion: Should Ecologists Become Bayesians? *Ecological Applications*, 6(4), 1095–1103. <https://doi.org/10.2307/2269594>

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6), 997–1016. <https://doi.org/10.1007/s11222-013-9416-2>

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>

Govindarajulu, P., Altwegg, R., & Anholt, B. R. (2005). Matrix Model Investigation of Invasive Species Control: Bullfrogs on Vancouver Island. *Ecological Applications*, 15(6), 2161–2170. <https://doi.org/10.1890/05-0486>

Gregory, R. D., Noble, D., Field, R., Marchant, J., Raven, M., & Gibbons, D. W. (2003). Using birds as indicators of biodiversity. *ORNIS HUNGARICA*, 14.

Hastings, A. (2013). *Population Biology: Concepts and Models*. Springer Science & Business Media.

Hill, M. O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, 54(2), 427–432. <https://doi.org/10.2307/1934352>

Hooper, D. U., Chapin, F. S., Ewel, J. J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J. H., Lodge, D. M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A. J., Vandermeer, J., & Wardle, D. A. (2005). Effects of Biodiversity on Ecosystem Functioning: A Consensus of Current Knowledge. *Ecological Monographs*, 75(1), 3–35. <https://doi.org/10.1890/04-0922>

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, 85(1), 3–28. <https://doi.org/10.1890/14-0661.1>

Hurlbert, S. H. (1971). The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology*, 52(4), 577–586. <https://doi.org/10.2307/1934145>

Jonsen, I. D., Basson, M., Bestley, S., Bravington, M. V., Patterson, T. A., Pedersen, M. W., Thomson, R., Thygesen, U. H., & Wotherspoon, S. J. (2013). State-space models for bio-loggers: A methodological road map. *Deep Sea Research Part II: Topical Studies in Oceanography*, 88–89, 34–46. <https://doi.org/10.1016/j.dsr2.2012.07.008>

- Jost, L. (2006). *Entropy and diversity*. *Oikos*, 113(2), 363–375.
<https://doi.org/10.1111/j.2006.0030-1299.14714.x>
- Kéry, M., & Schaub, M. (2011). *Bayesian Population Analysis using WinBUGS* - . 444.
- Lele, S. R., & Dennis, B. (2009). Bayesian Methods for Hierarchical Models: Are Ecologists Making a Faustian Bargain? *Ecological Applications*, 19(3), 581–584.
- Magurran, A. E. (2013). *Measuring Biological Diversity*. John Wiley & Sons.
- McRae, L., Deinet, S., & Freeman, R. (2017). The Diversity-Weighted Living Planet Index: Controlling for Taxonomic Bias in a Global Biodiversity Indicator. *PLOS ONE*, 12(1), e0169156. <https://doi.org/10.1371/journal.pone.0169156>
- Mekonen, S. (2017). *Birds as Biodiversity and Environmental Indicator*. 7.
<https://core.ac.uk/reader/234657570>
- Morris, E. K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T. S., Meiners, T., Müller, C., Obermaier, E., Prati, D., Socher, S. A., Sonnemann, I., Wäschke, N., Wubet, T., Wurst, S., & Rillig, M. C. (2014). Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution*, 4(18), 3514–3524. <https://doi.org/10.1002/ece3.1155>
- Newman, K. B., Buckland, S. T., Morgan, B. J. T., King, R., Borchers, D. L., Cole, D. J., Besbeas, P., Gimenez, O., & Thomas, L. (2014). State-Space Models. In K. B. Newman, S. T. Buckland, B. J. T. Morgan, R. King, D. L. Borchers, D. J. Cole, P. Besbeas, O. Gimenez, & L. Thomas (Eds.), *Modelling Population Dynamics: Model Formulation, Fitting and Assessment using State-Space Methods* (pp. 39–50). Springer. https://doi.org/10.1007/978-1-4939-0977-3_3
- Niemeijer, D., & de Groot, R. S. (2008). Framing environmental indicators: Moving from causal chains to causal networks. *Environment, Development and Sustainability*, 10(1), 89–106. <https://doi.org/10.1007/s10668-006-9040-9>
- Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., & Matthiopoulos, J. (2008). State-space models of individual animal movement. *Trends in Ecology & Evolution*, 23(2), 87–94. <https://doi.org/10.1016/j.tree.2007.10.009>
- Peet, R. K. (1974). The Measurement of Species Diversity. *Annual Review of Ecology and Systematics*, 5(1), 285–307. <https://doi.org/10.1146/annurev.es.05.110174.001441>
- Read “Using Science to Improve the BLM Wild Horse and Burro Program: A Way Forward” at NAP.edu. (n.d.). <https://doi.org/10.17226/13511>
- Royama, T. (2012). *Analytical Population Dynamics*. Springer Science & Business Media.
- Whittaker, R. H. (1965). Dominance and Diversity in Land Plant Communities. *Science*, 147(3655), 250–260. <https://doi.org/10.1126/science.147.3655.250>
- Wilson, H. B., Kendall, B. E., Fuller, R. A., Milton, D. A., & Possingham, H. P. (2011). Analyzing Variability and the Rate of Decline of Migratory Shorebirds in Moreton Bay, Australia. *Conservation Biology*, 25(4), 758–766. <https://doi.org/10.1111/j.1523-1739.2011.01670.x>

Yoccoz, N. G. (1991). Use, Overuse, and Misuse of Significance Tests in Evolutionary Biology and Ecology. *Bulletin of the Ecological Society of America*, 72(2), 106–111.

Yoccoz, N. G., Nichols, J. D., & Boulmier, T. (2001). Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, 16(8), 446–453. [https://doi.org/10.1016/S0169-5347\(01\)02205-4](https://doi.org/10.1016/S0169-5347(01)02205-4)