Using state-space time series analysis on wetland bird species to formulate effective bioindicators in the Barberspan wetland

Gareth R. Edwards, Res Altwegg, Birgit Erni

Summary

A summary will go here

Introduction

With the ongoing concern about biodiversity loss, monitoring biodiversity at various levels of biological organisation is important (Pereira et al. 2013). At the organismal level, biodiversity is generally measured by three components: number of species, total number of individuals (abundance), and how the individuals are distributed across species (evenness). We lose biodiversity if abundance decreases, if species go extinct, if communities become dominated by a few common species, or a combination of these processes. To measure changes in biodiversity, we need indicators that are sensitive to changes in these components (S. T. Buckland et al. 2005; S. Buckland, Yuan, and Marcon 2017).

One type of index that is sensitive to these three aspects of biodiversity change is based on aggregated species trends (Stephen T. Buckland et al. 2011). Let n_i be the number of individuals of species i and $n = \sum_{i=1}^{S} n_i$ the total number of individuals of S species in the community, so that $i = 1, \ldots, S$. Evenness measures how uniform the species proportions p_i are, where $p_i = n_i/n$. Now consider surveying a single location and counting the individuals of a group of species over time. The counts $n_{i,j}$ are the number of individuals encountered of species i in year j. Because population growth is a multiplicative process, changes in n_i over time are best measured as ratios $n_{i,j+1}/n_{i,j}$. Trends in the abundance of individual species can then be aggregated using the geometric mean (Stephen T. Buckland et al. 2011) or equivalently, the arithmetic mean of the changes in log abundances $G_j = exp\left(\frac{1}{S}\sum_{i=1}^{S} log\frac{n_{i,j+1}}{n_{i,j}}\right)$. The well-known Living Planet Index, for example, is based on the geometric mean (Loh et al. 2005; McRae, Deinet, and Freeman 2017) even though it measures change against a baseline, rather than from one year to the next.

The geometric mean weights all species trends equally and is therefore sensitive to small clusters of species with extreme trends (Leung et al. 2020). It is also sensitive to trends in rare species, which are

often difficult to estimate well (S. Buckland, Yuan, and Marcon 2017). An alternative approach for measuring change in biodiversity is to calculate a biodiversity index for each year and then examine changes in this index over time. Many measures of diversity exist but the ones that can be interpreted as the effective number of species seem the most intuitive ones (Jost 2006). One family of such diversity measures is the Hill numbers (Hill 1973):

$$^q D = \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)}$$

By varying the free parameter q, the Hill numbers are species richness for $q = 0$, the exponentiated Shannon entropy for $q = 1$ and the inverse of Simpson's index for $q = 2$.

When estimating trends in biodiversity, it is important to take into account the observation process that has given rise to the data. In most situations, it is impossible to census wild populations completely. We would like to know the number of individuals of species i, i.e. $n_{i,t}$, that is present in a population in year t. Instead, we obtain a count $c_{i,t}$ that is related to $n_{i,t}$ as $c_{i,t} = n_{i,t} \times p_{i,t}$ where $p_{i,t}$ is a detection rate. The observation process has three effects on the observed counts. 1) since $p_{i,t}$ usually varies because of variable conditions like weather, visibility, observer skills, characteristics of a species, etc., the counts are more variable than the actual population sizes (Link and Nichols 1994). 2) The $p_{i,t}$ vary because some individuals escape detection and others might be double counted. Often, non-detection is a bigger issue than double counting and so the $p_{i,t}$ tend to be < 1. As a result, the raw counts tend to underestimate the true population sizes. 3) Sometimes, surveys are missed altogether, leading to $p_{i,t} = 0$ for all species in a particular year t.

Here, we explore trends in waterbird populations at a particular wetland, Barberspan, in South Africa. Waterbirds are an important component of biodiversity in wetlands and also often serve as indicators for the condition of their environment (Amat and Green 2010; Gregory and Strien 2010; Sekercioglu 2006), such as the level of eutrophication of wetlands (Amat and Green 2010). Waterbids are also notoriously difficult to count as they tend to flock and can uncooperatively fly around during surveys. The waterbirds at Barberspan have been counted twice per year, in mid winter and mid-summer, since 1993. However, as is often the case in long-term monitoring programmes, counts were not always carried out an the time series therefore contain gaps. We therefore estimate population sizes and their trends using state-space models to reduce the effect of the observation process (Auger-Méthé et al. 2021). We then calculate indices for biodiversity change by aggregating population trends and using biodiversity indicators.

Methods

Study Area   The bird counts used in this study come from the Barberspan wetland which is the pilot site for this study. The Barberspan wetland is situated in North West province, South Africa. It is a protected wetland site known as Barberspan Bird Sanctuary that covers a 3200ha area. The wetland is centered on a shallow lake which covers approximately 257ha to 2000ha depending on rain fall (Remisiewicz and Avni 2011).

RAMSAR reports that the Barberspan wetland regularly supports approximately 20000 waterbirds with 365 different bird species (Rothmann J 2015). Barberspan is one of the first RAMSAR sites in South Africa and the only RAMSAR site in North West province.

Barberspan wetland is home to thousands of Little Stints and Curlew Sandpipers. It's also a popular location for intra-African nomadic waders such as Kittlitz's Plover and Crowned Lapwing. The bird species with the highest abundance in Barperspan wetland is the Red-Knobbed Coot, Egyptian Goose, Yellow Billed Duck, Southern Pochard and the South African Shelduck (Remisiewicz and Avni 2011). Appendix A contains further information regarding the waterbirds found in Barberspan wetland that was used in this study. The Barberspan wetland is rich in bird data and thus is an ideal wetland to use as a pilot site for this study.

Exploratory Data Analysis   The data used in this paper is from the Coordinated Waterbird Counts (CWAC) initiative. The counts are conducted twice a year, once in January/February (austral summer) and once in June/July (austral winter). The counting procedure is standardised and follows a rigorous set of instructions that counters are to follow. The counting protocol is such that all counts take place at approximately the same time each day, when the sun is either behind or beside the counter. More details about the counting process can be found at the following link (https://cwac.birdmap.africa/).

The CWAC dataset contains bird counts from 1993 to 2013 for the Barberspan wetland. Missing counts were inserted into the dataset as missing values that will subsequently be predicted by the state-space time series model. Only birds with 15 or more non-zero counts were used in the analysis as the variance parameters for the state-space time series model didn't converge when applied to bird counts with fewer than 15 non-zero counts.

State-Space Time Series Model   A state-space time series model (SSM) was fitted to the bird count data. State-space models have become increasingly popular tools in modelling population dynamics (Auger-Méthé et al. 2021). The key benefit of a state-space model is its ability to separate

observation error and latent error. By removing the observation error from the latent error, one is able to predict population trends more accurately (Kéry and Schaub 2011). SSMs are an ideal tool to use for time series analysis for the CWAC dataset, as bird counts are often prone to high levels of observation error and follow a markovian process, where the value at time t is dependent on the value at time t-1 (see figure 1 for illustration of markovian chain).
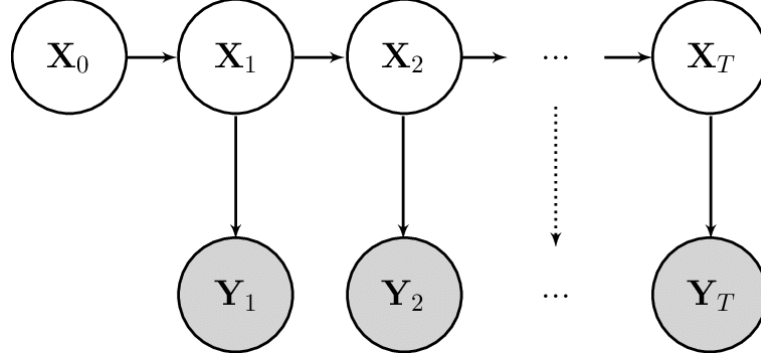

Figure 1: A visual representation of a state space model where X_t represents the unobserved population level changes. SSMs use the observed population counts, Y_t, to estimate the population level changes.

In this study the CWAC bird counts are modeled using an existing state-space time series model that was created by Barshep et al. (2017) and applied to bird count data in the CWAC dataset. Barshep et al. (2017) developed a bivariate time series model that treats summer and winter counts as separate variables, with the focus being on the summer population. The winter counts are modeled as a proportion of the summer counts as some birds from Barberspan migrate north during the southern winter.

SSMs model the state process and observation process separately such that state error and observation error can be separated for more accurate modelling of the count data. The mathematical illustration of the bivariate model is displayed below:

observation process

$$ln(summer\ count) : s_t = \mu_{t-1} + \alpha_t, \quad \alpha_t \sim N(0, \sigma_\alpha^2) \tag{1}$$

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

state process

$$ln(summer\ population\ level) : \mu_t = \mu_{t-1} + \beta_{t-1} + w_t, \quad w_t \sim N(0, \sigma_w^2) \tag{3}$$

$$ln(population\ change) : \beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim N(0, \sigma_\zeta^2) \tag{4}$$

$$ln(summer\ to\ winter\ ratio) : \lambda_t = \lambda_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{5}$$

where, s_t is the summer count at time t and w_t is the winter count at time t. μ_t is the latent summer population, β_t is the rate of change at time t and λ is the winter to summer ratio. The winter to summer ratio refers to the difference in the bird population from summer to winter. The palearctic migrants usually migrate north for winter, and leave their young behind, thus leaving a smaller proportion of birds in the South African wetlands in winter compared to summer. α_t, e_t, w_t, ζ_t and ϵ_t are all normally distributed, serially independent and independent of each other with variance σ_α^2, σ_e^2, σ_w^2, σ_ζ^2 and σ_ϵ^2 respectively.

The time series model presented above follows a Markovian process as the value at time t is dependent on the value at time t-1. The aim of this analysis is to sample from the probability distributions that represent the summer and winter counts. The current distribution for these summer and winter counts are unknown, but by applying Monte Carlo simulations from the summer and winter count distributions we can generate values that are from an approximation of the true count distribution (the stationery distribution). This process is known as a Monte Carlo Markov Chain (MCMC) approach (Brooks et al. 2011). The simulated values are known as the posterior distribution and is later used to calculate bio indices with credible intervals.

In this analysis the MCMC algorithm used is the Gibbs Sampler, which is applied using the JAGS package in R (Plummer 2003; R Core Team 2023; Kellner 2024). The JAGS software implements Bayesian inference based on Gibbs sampling by using a MCMC approach (Coro 2017). The hyperparameters available in the JAGS program are number of chains (n.chains), number of iterations (n.iter) and burn-in length (n.burnin). The number of chains refers to the number of markov chains created. These are the number of chains that run in parallel which converge to form the stationery distribution. A trace plot can also be plotted to show if these chains converge or not for a given summer count value. If all chains converge around the same approximation then we can be satisfied that the model has reached convergence. The number of chains was set to three in this analysis. The hyperparameters for number of iterations and "burn-in" for the Gibbs Sampler was 10 000 and 5000 respectively. The

number of iterations refers to the number of times the Gibbs Sampler samples from the stationery distribution that represents the summer or winter counts. The "burn-in" hyperparameter refers to the first few iterations that are to be discarded given that these first few iterations are usually less probable outcomes, but as the model runs for a longer period of iterations, the outcomes tend toward a higher probability region and thus are more likely outcomes (Kéry and Schaub 2011). This generates an output of 15 000 count values for each bird for each year with the first 5000 iterations being the "burn-in" iterations.

Biodiversity indices The 10 000 approximated count values for each bird count per year was used in calculating the modified LPI, exponentiated Shannon index and the Simpson index. The benefit of using the whole posterior sample of bird counts is so that a credible interval could be calculated for each index.

A popular population abundance index is the Living Planet Index (LPI) that is designed to track changes in population sizes over time. The LPI is used as an aggregated abundance index. It calculates the percentage change from a reference year (1970) to all subsequent years. This is done by averaging the rate of change (from reference year to current year) across all species of a species group (McRae, Deinet, and Freeman 2017).

A modified LPI approach was used to form an abundance index in this study. The modified LPI used calculates percentage change from each year to the next for each bird type in Barberspan and then calculates the average for that year. This is unlike the standard LPI approach that calculates the rate of change of species for each year based on the reference year (1970).

The modified LPI calculations are displayed as follows:

$$\overline{d_t} = \frac{1}{n_t} \sum_{i=1}^{n_t} d_{it} \tag{6}$$

where $\overline{d_t}$ is the average annual trend if there were multiple population time series of the same species. n_t refers to the number of populations. d_t is the annual rate of change for a population at time t, and is represented as follows:

$$d_t = log_{10}\left(\frac{N_t}{N_{t-1}}\right) \tag{7}$$

where N_t refers to the population size during year t.

The modified LPI calculations were applied to the summer population posterior values, for each year, estimated by the state-space model. The mean modified LPI was taken for each year and credible intervals were formed around that mean by calculating the 2.5th and 97.5th percentiles.

The exponentiated Shannon index and the Simpson index was also used to calculate further aspects of biodiversity in Barberspan. Namely the evenness and effective number of bird species. This is calculated using the Simpson's index and exponentiated Shannon's index respectivley (Nagendra 2002). The calculations for the Shannon's index and the Simpson's index is displayed as follows:

Shannon's index

$$x = -\sum_{i=1}^{S} p_i ln(p_i) \tag{8}$$

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.

Simpsons index

$$x = \sum_{i=1}^{S} p_i^2 \tag{9}$$

The Shannon and Simpson index were applied to the posterior values generated from the Gibbs sampler in similar manner as the modified LPI calculations. An exponentiated Shannon and Simpson index was calculated per year, using all 10 000 posterior count values for each bird species for that year. Thus creating 10 000 index values for each year. The mean index value was taken for both Shannon and Simpson values and the 2.5th and 97.5th percentiles were calculated to form the credible intervals.

Results

The combined bird counts of Barberspan from 1993 to 2018 are plotted in figure 2. The plot is broken up into summer and winter counts due to the biannual data capturing process that the CWAC protocols require. The plot shows fairly random variations of bird counts over the years, for both summer and winter counts, with multiple missing values in the dataset. Applying the state-space time series model to this data will approximate these missing values based on the previous existing values.
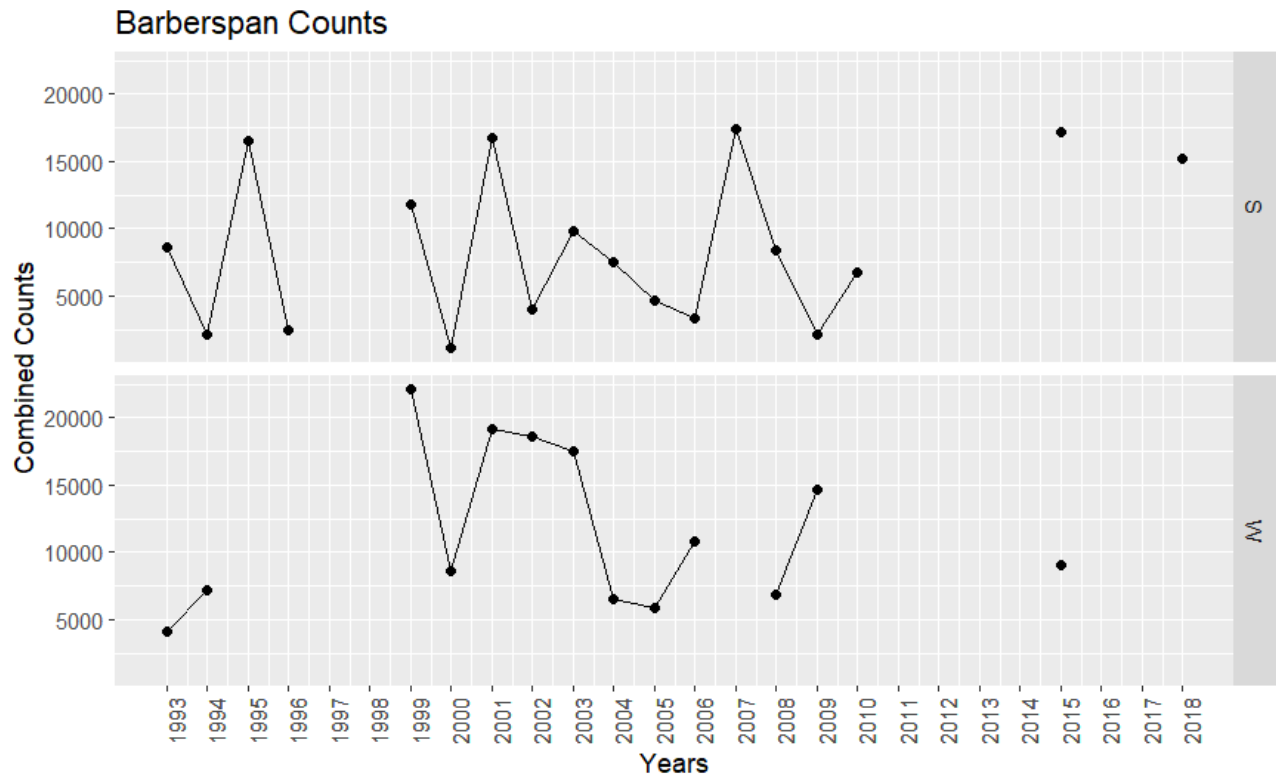
7

Figure 2: Total bird counts over time for the Barberspan wetland
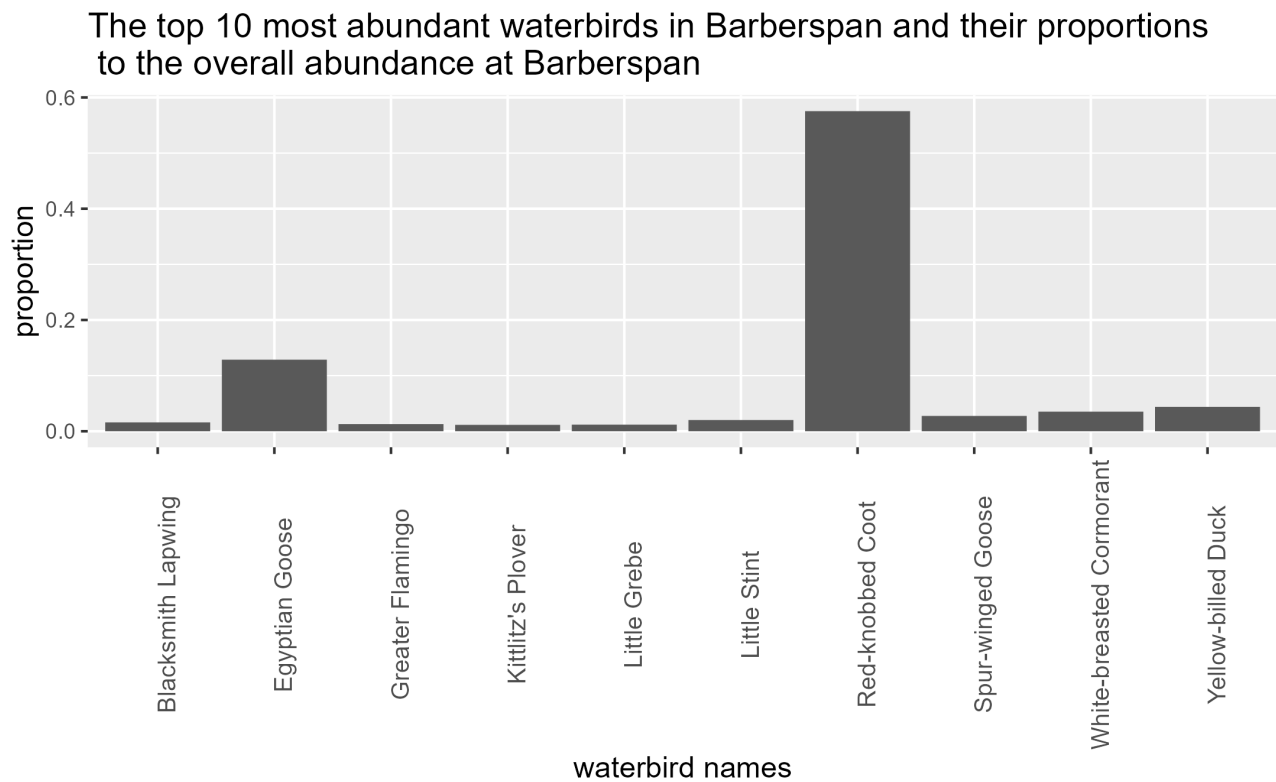

Figure 3: Top ten most abundant bird types in Barberspan wetland

A state-space time series model was fit to the combined bird counts of the Barberspan wetland using the JAGS (Just Another Gibbs Sampler) package. The model was fit on the summer and winter counts separately. The output is posterior distributions for summer and winter counts for each year. The state-space time series model output is displayed in figure 4.
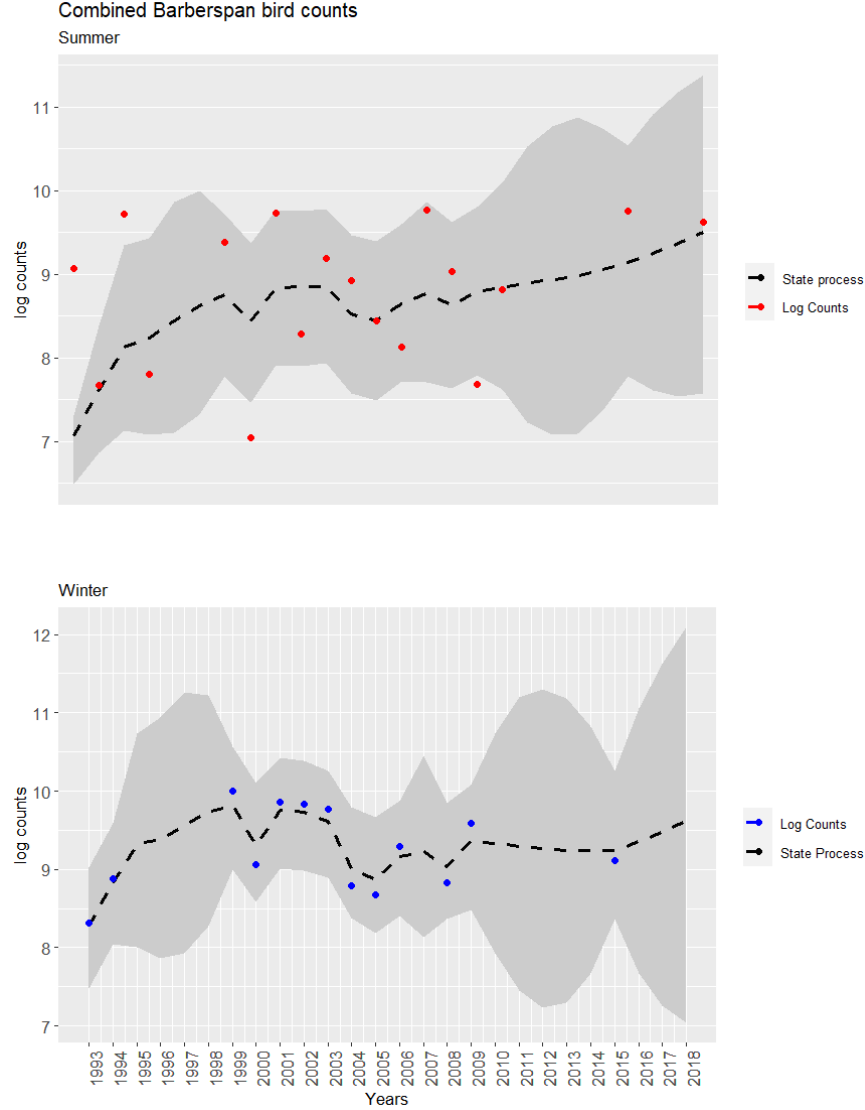

Figure 4: State-space time series output applied to the combined bird counts, separated by season, of Barberspan wetland. The grey area surrounding the dotted line is the 95% credible interval around the population level produced by the MCMC output.

From the 10 000 yearly count values generated from the Gibbs Sampler, we are able to calculate various bioindices. We demonstrate the benefits of using posterior output to calculate bioindices by calculating the LPI, exponentiated Shannon index and the Simpson index using the 10 000 yearly posterior values.

An LPI is calculated for each year and the mean, 2.5th and 97.5th percentile is generated from the 10 000 LPI values generated for each year. The output of this LPI calculations is shown in figure 5.
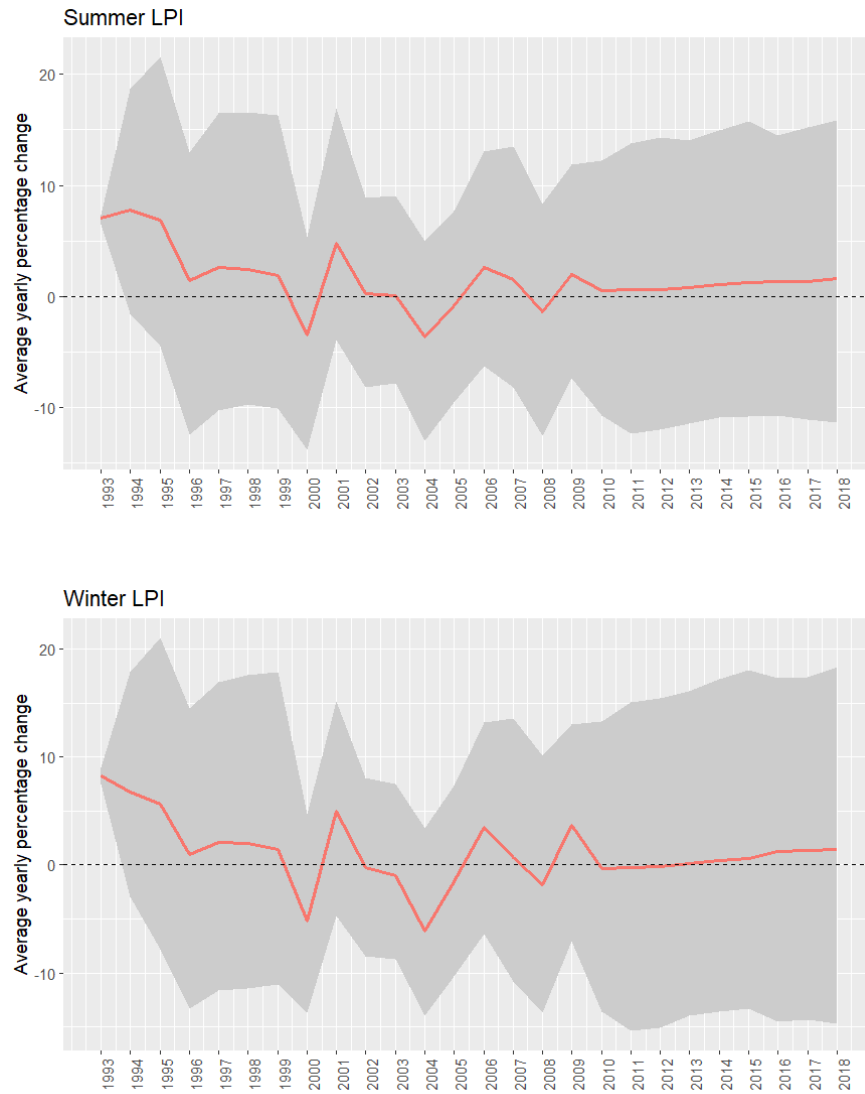
Figure 5: The LPI displaying the average yearly percentage change averaged across all birds in the Barberspan wetland. The grey area around the trend line represents the 2.5th and 97.5th percentile.

The exponentiated Shannon index was also calculated using the posterior output. An exponentiated Shannon index was calculated for each year. This resulted in 10 000 exponentiateed Shannon index values for each year. The mean, 2.5th and 97.5th percentile was then calculated from the 10 000 index values. The same approach was conducted when calculating the Simpson index for each year. The outputs of the exponentiated Shannon and Simpson index calculations are displayed in figure 6 and 7 respectively.
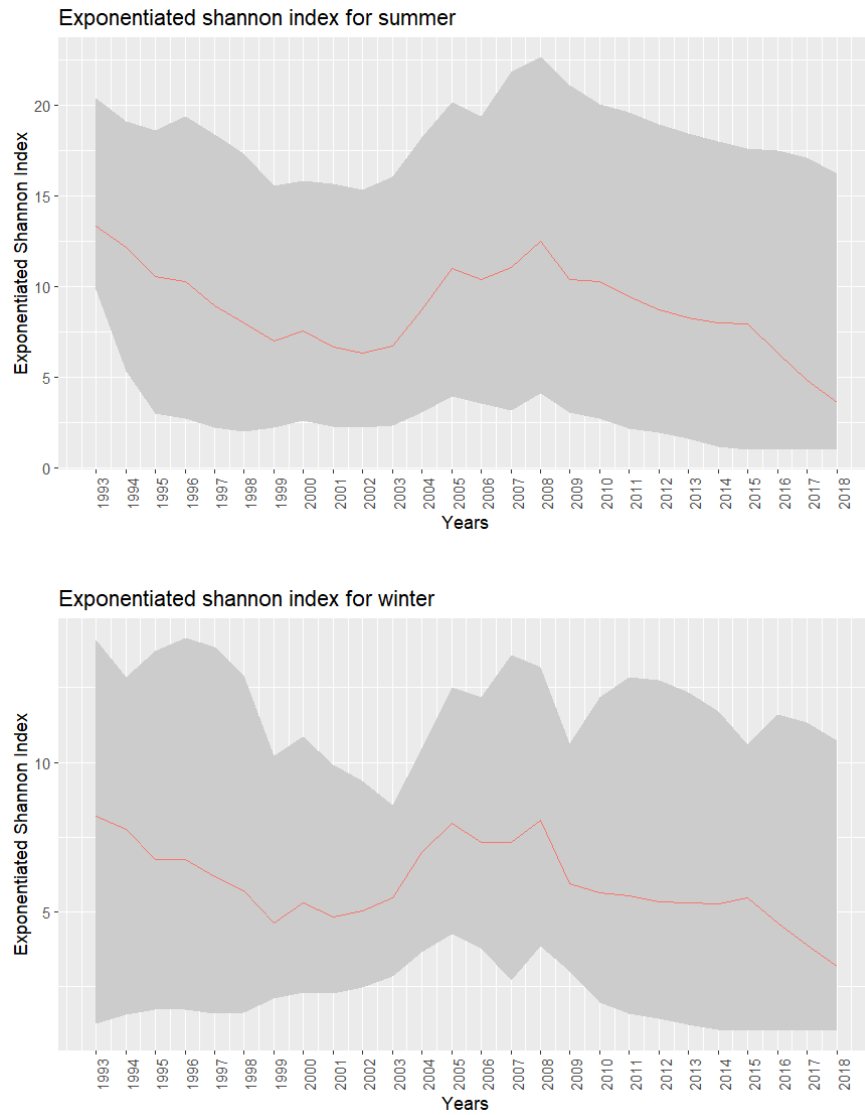
Figure 6: Exponentiated Shannon index for Barberspan, separated by season. The grey area around the trend line represents the 2.5th and 97.5th percentile

Figure 7: Simpson index for Barberspan, separated by season. The grey area around the trend line represents the 2.5th and 97.5th percentile

0.0.1 Discussion

Wetlands are valuable ecosystems for society (Clarkson, Ausseil, and Gerbeaux 2013). Unfortunately, there has been global wetland degradation with about half of the global wetlands being lost to accommodate for agricultural practices and urbanisation. The loss of wetlands has a negative impact on biodiversity, human well-being and has negative long-term impacts on economies (Clarkson, Ausseil, and Gerbeaux 2013). This has given rise to many wetland conservation projects, both locally (Working for Wetlands project) and globally (RAMSAR and AEWA).

A particularly effective method for tracking environmental changes within these ecosystems is through the use of bird species abundance as bioindicators. Birds are sensitive to various environmental changes and can thus provide direct reflections of ecosystem health (Mekonen 2017; Amat and Green 2010). In this study, we employed the state-space time series model to analyze bird abundance, utilizing the posterior output to calculate index values that measure population rate of change, diversity, and the expected number of bird species. This modeling approach offers a robust method for analyzing bird counts over time, as it distinguishes between latent and observation errors, thereby providing a clearer picture of true population dynamics (Auger-Méthé et al. 2021).

An additional advantage of using the state-space time series model is the ability to derive posterior distributions of bird counts, which can then be used to calculate index values with credible intervals. In our analysis, we calculated the Living Planet Index (LPI), Simpson Index, and Exponentiated Shannon Index to assess annual percentage population changes, community evenness, and the effective number of bird types at Barberspan, respectively. This flexibility allows conservation managers to compute any index relevant to their specific objectives, and the credible intervals associated with these indices enhance the reliability of the measurements.

Moreover, the state-space model allows for the incorporation of known biological patterns into the data analysis. In this study, we accounted for migratory patterns by including seasonal migrations of birds from South African wetlands during the winter. This aspect of the model helps to refine the accuracy of our population estimates and index calculations.

While our findings demonstrate the utility of state-space models in assessing bird abundance, there is room for further research. Future studies could apply these methods to different environments to broaden the applicability of our approach. Additionally, efforts should be made to reduce the variance in the models, as our index outputs occasionally displayed large error bands. By refining these models and expanding their application, we can improve our understanding of bird populations and contribute more effectively to wetland conservation efforts.

In conclusion, the state-space time series model proves to be a powerful tool in the analysis of bird abundance, offering detailed insights and reliable index values that can inform conservation strategies. As we continue to refine these models and expand their use, we enhance our ability to protect and preserve vital wetland ecosystems for future generations.

Supplementary Material

Acknowledgments

References

Amat, J. A., and A. J. Green. 2010. "Waterbirds as Bioindicators of Environmental Conditions." In *Conservation Monitoring in Freshwater Habitats: A Practical Guide and Case Studies*, edited by C. Hurford, M. Schneider, and I. Cowx, 45–52. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-9278-7_5.

Auger-Méthé, M., K. Newman, D. Cole, F. Empacher, R. Gryba, A. A. King, V. Leos-Barajas, et al. 2021. "A Guide to State-Space Modeling of Ecological Time Series." *Ecological Monographs* 91 (4). https://doi.org/e01470.

Barshep, Y., B. Erni, L. Underhill, and R. Altwegg. 2017. "Identifying Ecological and Life-History Drivers of Population Dynamics of Wetland Birds in South Africa." *Global Ecology and Conservation* 12: 96–107.

Brooks, Steve, Andrew Gelman, Galin Jones, and Xiao-Li Meng. 2011. *Handbook of Markov Chain Monte Carlo.* CRC Press.

Buckland, S. T., A. E. Magurran, R. E. Green, and R. M. Fewster. 2005. "Monitoring Change in Biodiversity Through Composite Indices." *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1454): 243–54. https://doi.org/10.1098/rstb.2004.1589.

Buckland, Stephen T, Angelika C Studeny, Anne E Magurran, Janine B Illian, and Stuart E Newson. 2011. "The Geometric Mean of Relative Abundance Indices: A Biodiversity Measure with a Difference." *Ecosphere* 2 (9): 1–15.

Buckland, ST, Yun Yuan, and Eric Marcon. 2017. "Measuring Temporal Trends in Biodiversity." *AStA Advances in Statistical Analysis* 101: 461–74.

Clarkson, B. R, A. E. Ausseil, and P. Gerbeaux. 2013. "Wetland Ecosystem Services. Ecosystem Services in New Zealand: Conditions and Trends." *Manaaki Whenua Press* Lincoln: 192–202.

"Convention on Biological Diversity." 2020. *Aichi Biodiversity Targets.* https://www.cbd.int/sp/targets/.

Coro, G. 2017. "Gibbs Sampling with JAGS: Behind the Scenes." https://www.researchgate.net/publication/313905185_Gibbs_Sampling_with_JAGS_Behind_the_Scenes.

FitzPatrick Institute of African Ornithology. 2020. "CWAC Coordinated Waterbird Counts." https://cwac.birdmap.africa/.

Gregory, R. D., and A. van Strien. 2010. "Wild Bird Indicators: Using Composite Population Trends of Birds as Measures of Environmental Health." *Ornithological Science* 9 (1): 3–22. https://doi.org/10.2326/osj.9.3.

Hill, M. O. 1973. "Diversity and Evenness: A Unifying Notation and Its Consequences." *Ecology* 54

(2): 427–32. https://doi.org/10.2307/1934352.

Jost, Lou. 2006. "Entropy and Diversity." *Oikos* 113 (2): 363–75.

Kellner, Ken. 2024. *jagsUI: A Wrapper Around 'Rjags' to Streamline 'JAGS' Analyses.* https://CRAN.R-project.org/package=jagsUI.

Kéry, Marc, and Michael Schaub. 2011. *Bayesian Population Analysis Using WinBUGS: A Hierarchical Perspective.* Academic Press.

Leung, Brian, Anna L. Hargreaves, Dan A. Greenberg, Brian McGill, Maria Dornelas, and Robin Freeman. 2020. "Clustered Versus Catastrophic Global Vertebrate Declines." *Nature* 588 (7837): 267–71. https://doi.org/10.1038/s41586-020-2920-6.

Link, William A, and James D Nichols. 1994. "On the Importance of Sampling Variance to Investigations of Temporal Variation in Animal Population Size." *Oikos*, 539–44.

Loh, J., R. E. Green, T. Ricketts, J. Lamoreux, M. Jenkins, V. Kapos, and J. Randers. 2005. "The Living Planet Index: Using Species Population Time Series to Track Trends in Biodiversity." *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1454): 289–95. https://doi.org/10.1098/rstb.2004.1584.

McRae, L., S. Deinet, and R. Freeman. 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." *Journal of Natural Sciences Research* 7 (21): 28–34. https://core.ac.uk/reader/234657570.

Nagendra, Harini. 2002. "Opposite Trends in Response for the Shannon and Simpson Indices of Landscape Diversity." *Applied Geography* 22 (2): 175–86. https://doi.org/10.1016/S0143-6228(02)00002-4.

Pereira, Henrique Miguel, Simon Ferrier, Michele Walters, Gary N Geller, Rob HG Jongman, Robert J Scholes, Michael William Bruford, et al. 2013. "Essential Biodiversity Variables." *Science* 339 (6117): 277–78.

Plummer, M. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria* 124.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Remisiewicz, M., and J. Avni. 2011. "Status of Migrant and Resident Waders, and Moult Strategies of Migrant Waders Using African Inland Wetland Habitats, at Barberspan Bird Sanctuary in South

Africa: Reports from BOU-Funded Projects." *Ibis* 153 (2): 433–35. https://doi.org/10.1111/j.1474-919X.2011.01112_1.x.

Rothmann J. 2015. "Barberspan." South {Africa}'s {Ramsar} {Sites}. *Barberspan.* http://www.saramsar.com/2015/06/barberspan.html.

Sekercioglu, Cagan H. 2006. "Increasing Awareness of Avian Ecological Function." *Trends in Ecology & Evolution* 21 (8): 464–71. https://doi.org/10.1016/j.tree.2006.05.007.