East Atlantic Flyway waterbird monitoring: some statistical challenges and Suggestions

Godeau U.[1], Gaget E.[2], Dami L.[2], Defos du Rau P.[1]

[1] Office Français de la Biodiversité, Tour du Valat, Le Sambuc, 13200 Arles, France
[2] Institut de Recherche de la Tour du Valat, Le Sambuc, 13200 Arles, France

This paper is not to be cited yet, contact Pierre du Rau Defos for more information

Abstract
The International Waterbird Census (IWC) is one of the longest-term and largest scale biodiversity monitoring programs of citizen science, providing data to assess waterbird population trends in support of several international agreements including the African Eurasian migratory Waterbirds Agreement, the Ramsar Convention and the Wadden Sea Flyway Initiative along the East Atlantic Flyway. Biodiversity data collected at large spatial and temporal scales are particularly relevant to provide policy makers with sound recommendations, but they can be challenging to analyze. Several modelling recommendations have been done to ensure the robustness of biodiversity models, and failing to address them can lead to erroneous conclusions Using modelling experiments on IWC data from five example waterbird species from the East Atlantic Flyway (4522 sites, 26 years), we addressed five statistical challenges potentially affecting the estimation of waterbird population trend: 1) zeroinflation, 2) sampling design, 3) detection bias, 4) imputation method and 5) spatial autocorrelation. We discuss these challenges and provide recommendations for future IWC data management and analyses in order to accommodate or mitigate these challenges.

Key words
East Atlantic Flyway, International Waterbird Census, sampling design, spatial autocorrelation, imputation, zero-inflation, detection bias

1. Introduction

Long-term biodiversity monitoring schemes are often more useful for answering important ecological or conservation questions than short-term ones (Magurran et al. 2010, White 2019). They are also generally more useful to support management decision-making and inform policy than short-term ones (Hughes et al. 2017). One of the longest-term, widest and most internationally widespread biodiversity monitoring scheme is the International Waterbird Census (IWC) (Stroud et al. 2022). The IWC is coordinated on a global scale by Wetlands International (www.wetlands.org). The IWC was launched in 1967 and covers today over 25 000 sites in more than 100 countries, making it one of the largest global monitoring schemes based largely on citizen science (Sayoud et al. 2017, Amano et al. 2018). Because it is a long-term global dataset, the IWC has permitted sound scientific inference on several significant advances in ecology and conservation (e.g. Green & Elmberg 2014, Amano et al. 2018, Gaget et al. 2018). In particular, the IWC is of particular importance and relevance in informing international agreements and treaties and supporting their policy decisions. It is for instance central in supporting policy of the EU Birds Directive (2009/147/EC), the Ramsar Convention on Wetlands, the Convention on the Conservation of Migratory Species of Wild Animals and especially the African-Eurasian migratory Waterbirds Agreement (AEWA Secretariat, 2013).

The 2010 Joint Declaration on the protection of the Wadden Sea is another international agreement supported by the IWC in the framework of the Wadden Sea Flyway Initiative (WSFI). The aim of the WSFI is to strengthen cooperation on conservation, management and research activities between the three Wadden Sea states and other countries along the East Atlantic Flyway (EAF) which play a significant role in conserving migratory waterbirds along this flyway (https://flyway.waddensea-worldheritage.org/). With a low minimum 20 million waterbirds belonging to 250 different species sampled over 115 sites (van Roomen et al. 2020), the EAF is of major importance for waterbirds at the global scale. By benefiting from the long-term support of the Trilateral Wadden Sea Cooperation, bird and wetland scientists and conservationists in the 31 EAF countries can infer sounder recommendations using such large spatio-temporal dataset (van Roomen et al. 2022).

Among the quantitative outputs inferred from the EAF dataset, one statistic is of particular interest conservation-wise: population trend of the targeted waterbird species. Methods for statistical estimation of both these parameters is still an area in progress (Buckland & Johnston 2017, Nagy et al. 2022), especially as large-scale schemes based on citizen-science may suffer from some additional statistical bias and challenges compared to more standardized biodiversity schemes run by e.g. professionals (Kosmala et al. 2016, Johnston et al. 2021). Zero-inflation (Zipkin et al. 2014), sampling design (Aubry et al. 2023), detection bias (Kéry et al. 2009), choice of imputation approach (Dakki et al. 2021) and spatial bias (Rocchini et al. 2023) are among the statistical issues affecting the estimation or modelling of wildlife population trend that still present potential for progress.

By modelling and simulating monitoring data of five waterbird species at the EAF scale, we evaluate here some impacts of these different statistical issues on population trend estimations and make suggestions for potential analytical improvement for the future of the EAF waterbird monitoring and possibly the whole IWC.

2. Methods

2.1 International Waterbird Census data

We used waterbird count data from five unnamed waterbird species from the EAF as examples and modelled their time-series from 1995 to 2020 (Table 1, Figure 1). We choose to keep species identity undisclosed in order to focus on statistical challenges rather than ecological inferences although we acknowledge that integrating population ecology into any population dynamics modelling is central in any proper conservation biology. IWC ideally takes place in mid-January every year. Each observers team, composed by skilled ornithologists, apply the field protocol for waterbird counting recommended by Wetlands International (2010). The count data were obtained by scanning multi-species flocks of waterbirds with telescopes or binoculars as appropriate and counting individuals of each species, or by estimating species-specific abundance using 'blocks' of known size in the appropriate order of magnitude. Counting sites of the EAF are distributed from Northern Scandinavia to South Africa all along the East Atlantic part of Europe and Africa (Figure 2) but not all are counted every January, resulting in a non-negligible number of missing site-year entries (NAs). In order to minimize such missing data, we first discarded count data before 1995 (when not all sites had been identified within the EAF database) and after 2020 (when not all surveys have been validated and integrated into the EAF database). We also filtered out sites with less than 15 surveys within time-series in order to decrease missing data to a reasonable ratio (less than 25% percent, Table 1).

Species code	number of sites	mean count/site	SE count/site	Q1 count/site	Median count/site	Q3 count/site	Max count/site	number of site.year entries	Number of NAs	% NAs
A	2165	54.56	2,89	0	0	1	45300	56290	8035	14.3
B	1340	16.62	0,89	0	0	3	6169	34840	5371	15.4
C	3445	288.5	6,74	0	0	70	96000	89570	14958	16.7
D	141	2.636	0,62	0	0	0	1291	3666	860	23.5
E	4467	352.4	5,51	10	57	220	85576	116142	19611	16.9

Table 1: description of the count data for the five EAF species

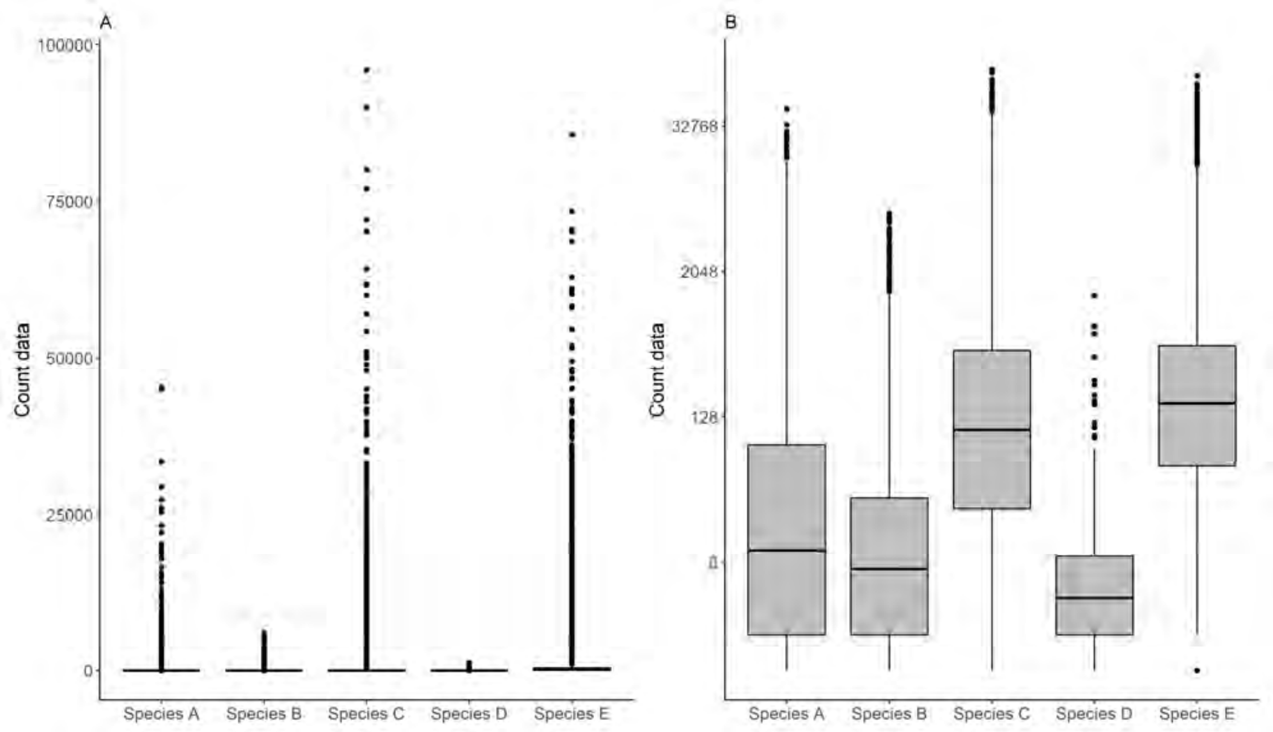

Figure 1: count data distribution for the five waterbird species over linear scale (A) or logarithmic scale (B).

Figure 2: monitoring sites for the five EAF studied species

2.2 time-series modelling

Time-series data were modelled by using generalized linear mixed effect models (GLMM) to estimate time (i.e. year) fixed effect in order to assess time trend as the slope of the effect of years on untransformed count data (O'Hara & Kotze 2010). We added a site random effect on the GLMM intercept to account for repeated samplings (Bird et al. 2014) and selected an appropriate response probability distribution (among Poisson, negative binomial and Tweedie error distribution) to accommodate the observed overdispersion. We used this design as the "base GLMM" to be compared and refined according to each statistical issue.

2.3 potential statistical challenges

In addition to overdispersion, sampling design, detection bias, imputation method and spatial autocorrelation can potentially impact time trend analysis. For sampling design, detection bias and imputation analyses, we plotted and compared the slope estimates of the different models compared, through the mean estimate of the year coefficients and its 95% confidence interval (as mean ± 1.96 * standard error).

Choice of response distribution

Poisson distribution is widely used to model wildlife, including bird, count data (e.g. van Strien et al. 2004, O'Hara & Kotze 2010). Yet, overdispersion and particularly zero-inflation can cause violation of the assumption of Poisson distribution ($E(Y)=Var(Y)=\lambda$, where Y is the observed count and λ the mean or expected count). Since our data seemed heavily zero-inflated (Table 1 and Figure 1), we compared zero-inflation under Poisson, negative binomial and Tweedie distributions using the "testZeroInflation" function in package DHARMa (Hartig 2022) to decide on a most appropriate distribution. The response distribution with the lowest zero-inflation was selected for each species as the distribution form to be used in the "base GLMM" and all refined GLMMs.

Sampling design

Population declines are best identified when species abundance values are high (Ficetola et al. 2018). We ran a GLMM on subsamples of each species-specific count dataset defined according to mean site-specific abundance in order to investigate which subsample would be most important to identify potential time-trend. Locating where would, in each full species-specific sample, most time-trend information be located could in turn potentially contribute to design a more cost-effective sampling scheme if need be. We therefore compared time-trends for each species, i.e. slope of the base full sample GLMM to GLMM based on the three subsamples respectively lacking the 10%, 20%, 50% sites with the smaller mean counts and a subsample lacking the 10% sites with the larger mean counts.

Detection bias

Imperfect detection has been a major wildlife monitoring issue since the 90s (Nichols et al. 2000), with explicit impacts on both inference of population metrics (Kellner & Swihart 2014) and costs of policy decisions (Moore & Kendall 2004) but seems to remain difficult to accommodate in large-scale long-term monitoring schemes (Sanz-Pérez et al. 2020) despite existing solutions like the double sampling approach proposed by Pollock et al. (2002). In order to explore impacts of detection bias on population trend estimation, we simulated an "unbiased" monitoring output by (i) randomly affecting one to five undetected individuals (with a discrete uniform probability of corrected count size) to 30% of the zero counts (ii) adding to all non-zero counts an ad-hoc uniformly distributed waterbird bonus randomly selected between 0% and 10% of the log-transformed count. For the sake of mere exemplification, we assumed a moderate rate of false-zero (30%), a systematic negative bias of count data due to imperfect detection and a positive effect of count size on counting error, larger clusters being more difficult to enumerate than smaller ones. We do not argue that the simulated detection correction we implemented is accurate, but we designed it to be reasonably limited with only 30% of false absence and an upper limit of count underestimation set at approximately 50% in the range of count values observed for our five species. Such detection bias is well within the range observed by e.g. Tyre et al. (2003) and Kellner & Swihart (2014). Correcting for negative bias in detection would obviously produce higher and likely very different abundance estimates hence flyway population estimates. Yet, we aimed at investigating whether such bias correction would also modify trend estimation. We then compared population trend estimations for each species, i.e. slope of the base GLMM on uncorrected count to range of slopes estimated from ten simulated GLMMs corrected for detection by randomly drawn unbiased counts.

Choice of imputation method

Within the whole EAF, there are several subregions where data collection is too costly or logistically difficult to be completed every year, which results in missing entries in the WI database. A substantial amount of missing data jeopardizes sound inference of population trends by unbalancing sampling in time and space (Nakagawa & Freckleton 2008, Łopucki et al. 2022), especially in areas where biodiversity is in stronger need of monitoring (Stephenson et al. 2017). Missing data in biodiversity monitoring is generally dealt with by case removal or missing value imputation, including through model-based imputation of spatio-temporal count data (Onkelinx et al. 2017). The TRIM (TRends and Indices for Monitoring data) model is the most widely used example of such methods, particularly for waterbirds (Lehikoinen et al., 2013; Van Strien et al., 2004). Recently an alternative model based on penalized Poisson parametrization (LORI as Low-Rank Interactions) was proposed by Dakki et al. (2021) to impute and analyze incomplete monitoring data and allowing integration of a large number of covariates, a feature not implemented in TRIM. Concerning the TRIM method, only factor covariates are allowed (with a limited number of levels) and we therefore used the location of the site in one of these four large geographical regions: Northern Europe, Southern Europe, North Africa and sub-Saharan Africa. Concerning the LORI method, only numerical covariates are allowed, and we therefore used the two covariates: the longitude and latitude of the site. In the same logic than for detection bias correction, imputing missing data would potentially also produce different abundance estimates hence flyway population estimates. Yet, we primarily aimed at investigating whether imputation method would also modify trend estimation. We then compared time-trends for each species, i.e. slope of the base GLMM over non-imputed data to slopes estimated over datasets imputed by either TRIM (using the rtrim package, Bogaart et al. 2016) and LORI (using the lori package, Robin et al. 2019) under the same GLMM design as the base GLMM with non-imputed data.

Spatial autocorrelation

Spatial autocorrelation (SA) is a pattern frequently found in species distribution data, i.e. locations close to each other exhibit more similar or dissimilar values than would be predicted by chance alone (Dormann et al. 2007). Waterbird count data could therefore display some SA when occurrence or abundance in neighbouring sites are negatively or, more often, positively correlated because of habitat or waterbird gregariousness or ecology. SA is a serious statistical issue in ecological models as it generally violates their assumption of independence of individuals and can artificially result in small confidence intervals on estimated parameters (Legendre 1993, Dale & Fortin 2002). We tested in several models, different spatial structures to accommodate and mitigate SA. Moran's I (Rangel et al. 2010) was assessed for each model thanks to the function testSpatialAutocorrelation() from the package DHARMa. Tested spatial structures were:

- Random intercept (RI): building on the methodology presented by Folliot et al. (2022), we grouped the sites using a 100x100km grid cells and added the cell identity corresponding to each site as random effect on the intercept. The spatial structure is then included as a hierarchical level to account for correlation among cells.
- Gaussian random field (GRF): we used the Stochastic Partial Differential Equations (SPDE) to approximate the gaussian random field, as introduced by Lindgren et al. (2011). The spatial mesh matrix needed to apply the SPDE approach was created from site coordinates using the package sdmTMB (Anderson et al. 2022). To explore the effect of the mesh parameters on SA mitigation, we used meshes, either unconstrained or constrained by the coastline as physical barrier. In addition, we tested the effect of minimal distance between mesh vertices, using cutoff distance of 0.05, 0.1, 0.15, and 0.5 decimal degree. As a result, we tested the effect of the gaussian random field in 10 models per species.
- Basis penalty smoothing (BPS): we used the package 'mgcv' (Wood 2011) to fit a GAMM with a two-dimensional smoothing function on site coordinates to model surface trends (Dormann et al. 2007).

3. Results

3.1 Choice of response distribution

The Poisson distribution was the least efficient in accommodating response zero-inflation whereas the Tweedie distribution was consistently accommodating the most zero-inflation across all five studied EAF waterbirds (Figure 3). We therefore fitted all subsequent models with Tweedie response distribution.

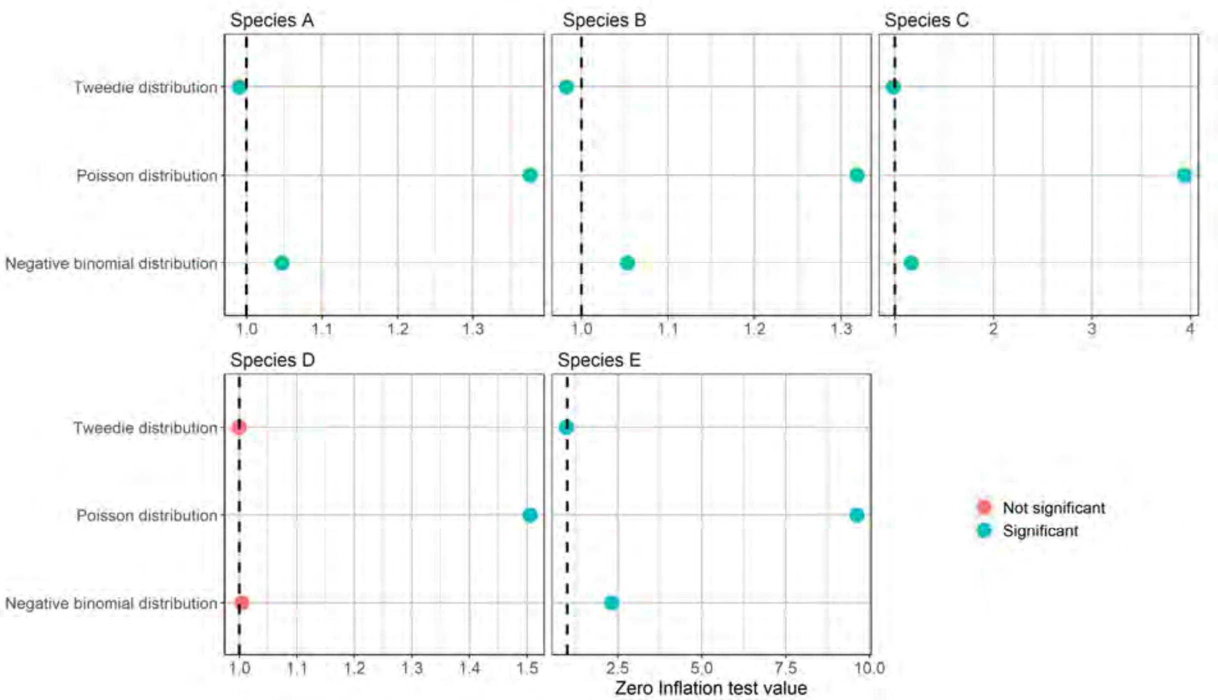

Figure 3: Zero-inflation test value and significance for three response distribution families and the five studied species.

3.2 Sampling design

GLMMs discarding the 10% sites with larger mean counts had the most different slope overall except for species E which showed a different pattern in time trends of subsamples (Figure 4). Slopes of GLMMs discarding sites with smaller mean counts were generally similar among each other, again with the exception of species E. Distribution of count data for species E was different from the data distribution of the other four species (Table 1) with a larger mean count/site overall and particularly in the smaller counts (both Q1 count/site and median count/site non-null), thus increasing influence of the sites with smaller counts on the overall time slope. For the other four species, sites with larger counts had the most influence on the slope.

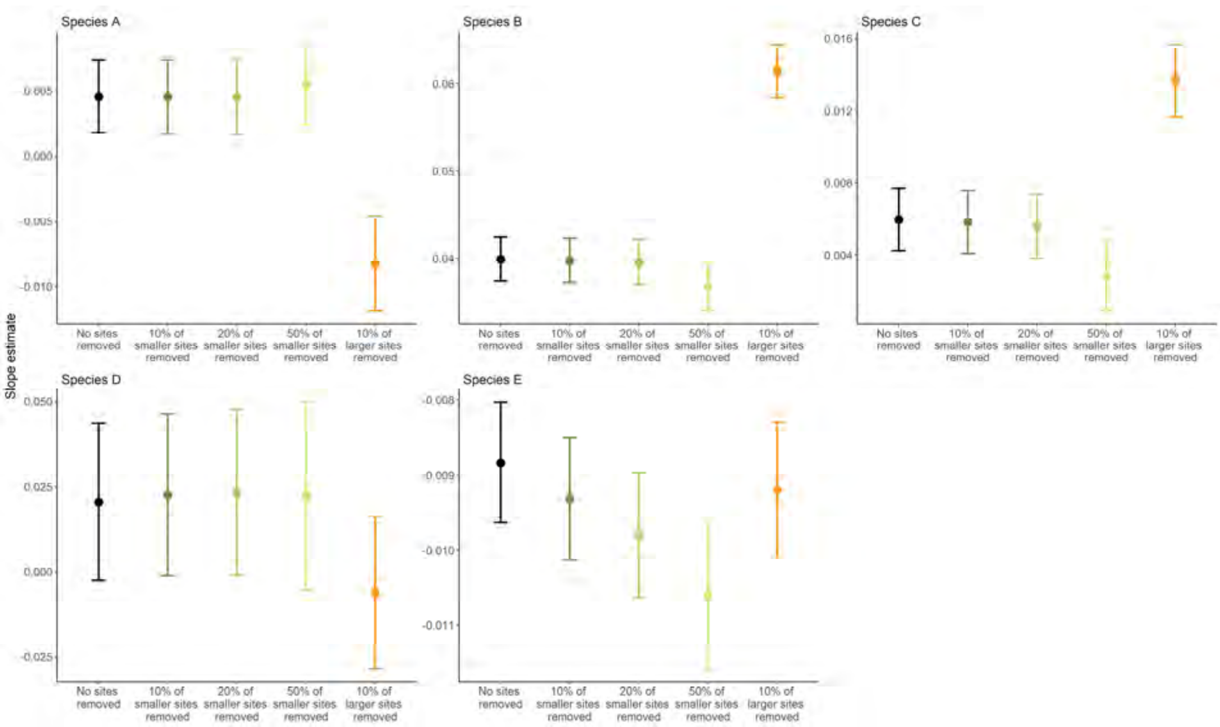

Figure 4: slope estimated (and 95% CI) from GLMM for the five studied species, using the full dataset (black), the dataset excluding 10%, 20% and 50% sites with smaller counts (green; dark, medium, pale) and excluding the 10% sites with larger counts (orange).

3.3 Detection bias

Slopes estimates differed to a large extent between the basic GLMM and the GLMMs using data arbitrarily corrected for detection probability (Figure 5).

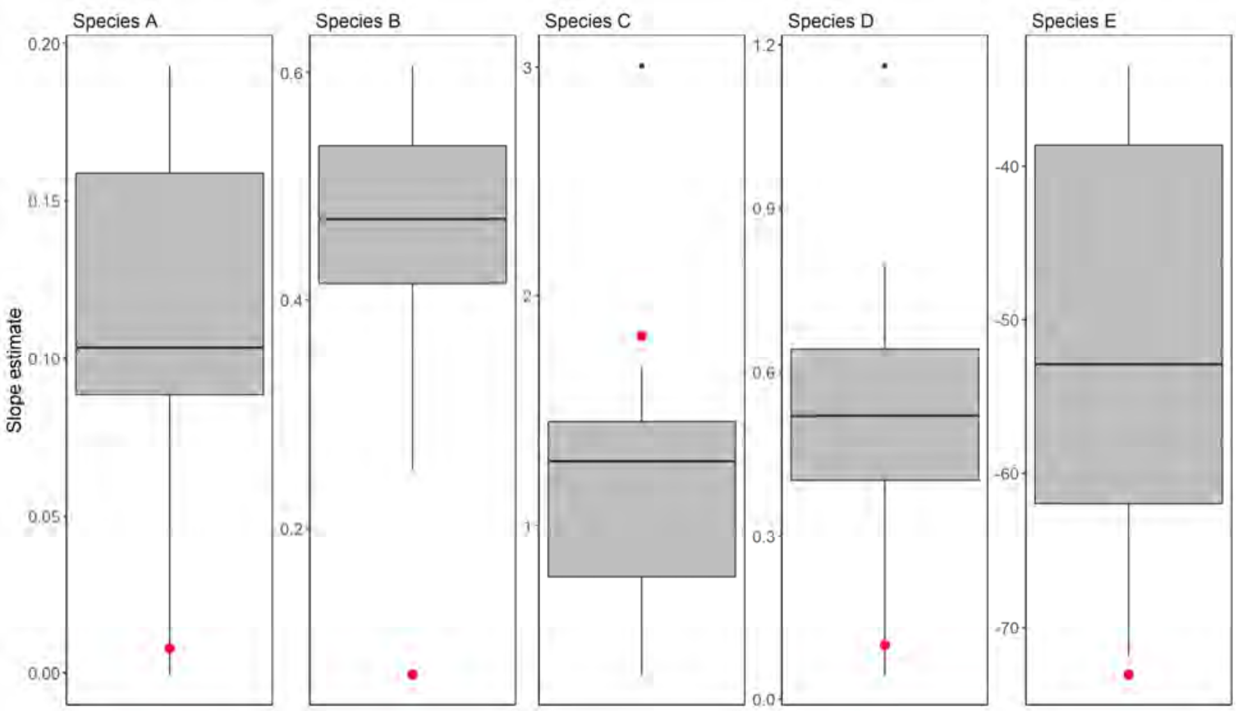

Figure 5: Slope estimates for year effect resulting from the detection bias Models. Red dots present the slopes estimates for the uncorrected data and boxplots present slope estimates for the ten simulated datasets corrected for detection by randomly drawn unbiased counts.

3.4 Imputation method

Slope estimates for time effect in the basic initial GLMM, the TRIM-imputed GLMM and the LORI-imputed GLMM all largely differed (Figure 6), illustrating the potential impact of the choice of imputation method on trend estimation. There was no consistent pattern in differences among slopes, the LORI GLMM producing more precise estimates, closer to those estimated from the initial non-imputed GLMM.

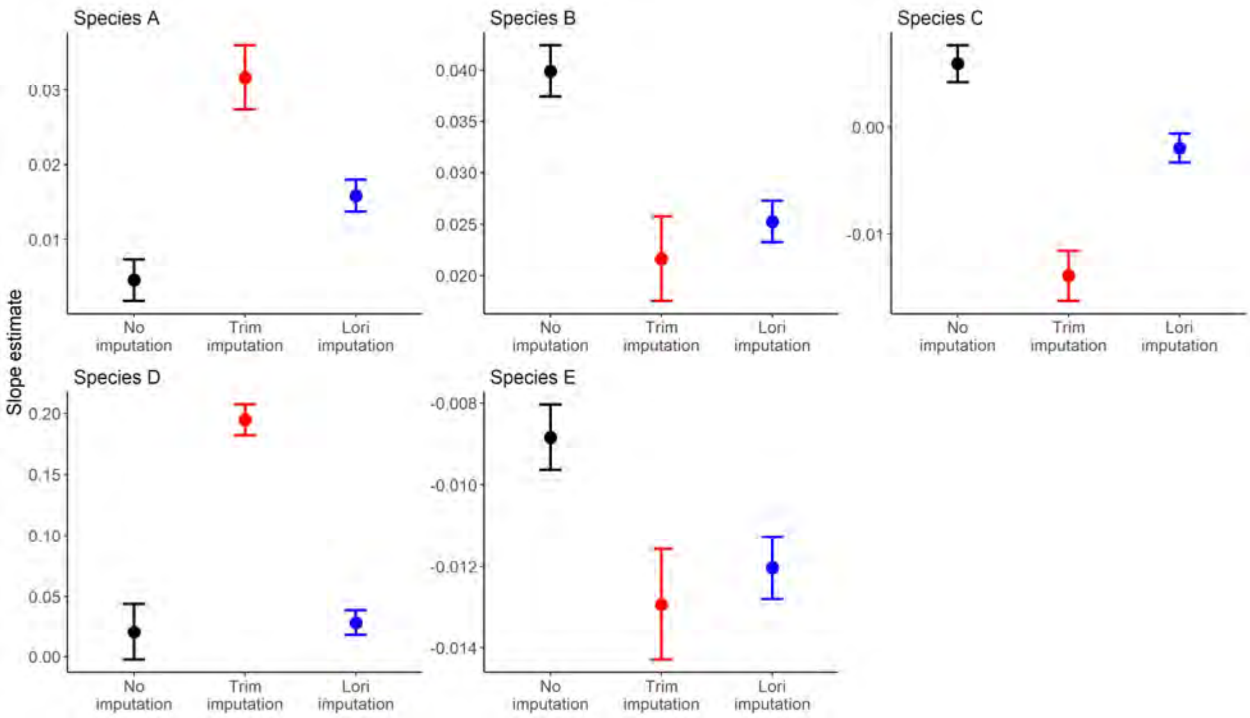

Figure 6: Slope estimates (and 95% CI) of year effect (time-trend) for the five studied species for non-imputed data GLMM (black), the TRIM-imputed GLMM (red) and the LORI-imputed GLMM (blue).

3.5 Spatial autocorrelation

Spatial autocorrelation was found positive and limited (0> Moran's I<0.1) yet always significant in residuals of the five EAF count data GLMMs (Figure 7). Autocorrelation was relatively variable among spatial models and among species, with most modelling options remaining unable to remove autocorrelation in the absence of spatial predictors other than geographical coordinates. Only in two species could autocorrelation be removed, either with a relatively small-sized mesh resolution (species C) or with any spatial modelling option (species D). Indeed, in general, integrating spatial components into models of counts contributed to decrease the regime of autocorrelation, with BPS performing relatively less well than RI.

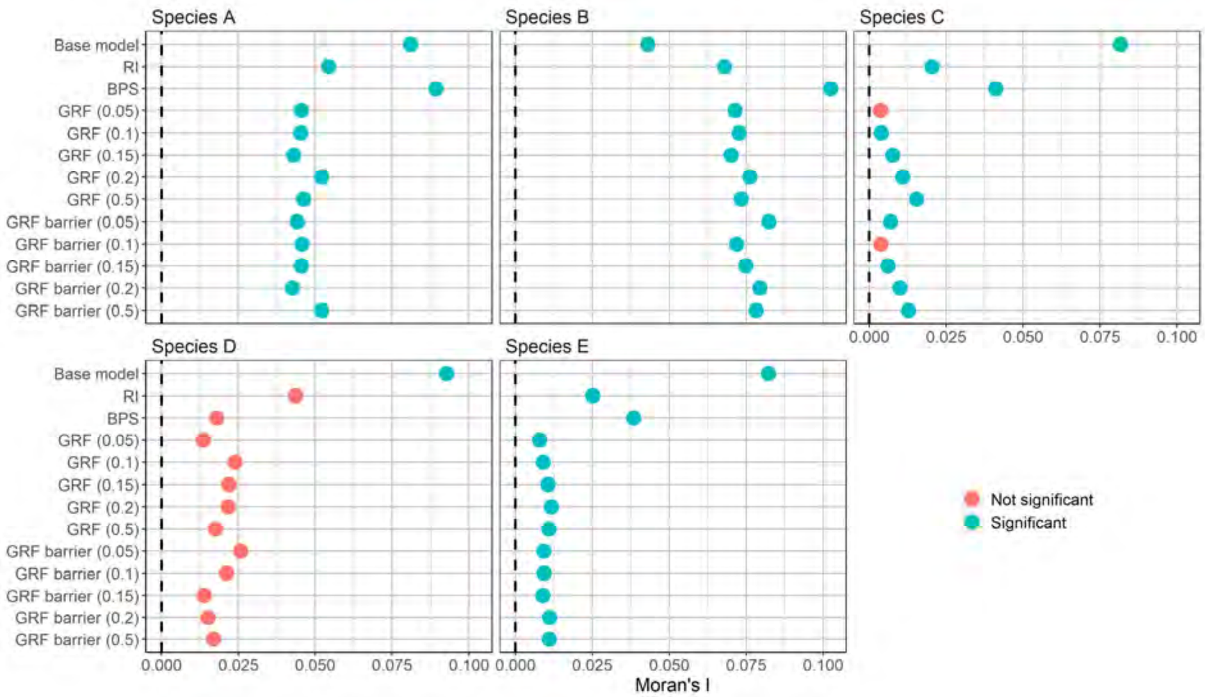

Figure 7: Moran's I value and test significance for four main model families (including 10 INLA models with two different mesh constraints and five parameterization each) for the five studied species. Random intercept (RI), Gaussian random field (GRF), Basis penalty smoothing (BPS).

4. Discussion

By analyzing IWC data from five random species from the EAF, we wished to illustrate that trend estimation based on IWC count data is not trivial and required advanced methodological approaches, with still potential for improvement. Investigating five species also shows that analytical requirements can and should differ between species. Because we randomly studied unknown species, we believe that trend estimation for any IWC waterbird species may thus require species-specific analytical requirements. A first recommendation of the present work would be to consider the statistical challenges addressed here when estimating population trends of all waterbird species. Depending on their zero-inflation, count data distribution and spatial autocorrelation, IWC count data may require different analytical scheme or modelling specifications between species.

We focused this work on time trend estimation, although population size estimation is equally important conservation-wise. Studying population size would require more in-depth knowledge regarding detection bias, which we do not have since it depends on the species studied. Having chosen to study unknown species, the study of trends appears more relevant. However, these statistical challenges also appear when studying the size of populations; our results shed light on and largely apply to the similar challenges of population size estimation. Other statistical challenges concerning the study of IWC data could also ultimately be studied with the aim of highlighting their impacts and proposing solutions, like e.g. relevance of generalized linear vs additive models (Atkinson et al. 2006), but we argue that we illustrated here some of the most impacting ones.

4.1 zero-inflation: we recommend focusing on the choice of distribution for count modelling
Although being widely used in wildlife count models, the Poisson distribution was the least efficient in accommodating response zero-inflation whereas the Tweedie distribution was consistently accommodating the most zero-inflation across all five studied EAF waterbirds (Figure 3) with the negative binomial distribution coming close. In order to avoid additional modelling assumptions, we did not investigate hurdle approaches, i.e. two-stage modelling for presence-absence and abundance conditional on presence. Hurdle models could be further explored to model IWC count data. The Tweedie distribution is relatively widespread and simple in use (Miller et al. 2013) and supported in several R packages. Zero-inflation in count, i.e. observed, data is a regular, if not frequent, pattern in wildlife science and conservation (Dénes et al. 2015). In line with Tirozzi et al. (2022), we recommend a more careful choice among response distributions when modelling EAF count data. As recommended by Miller et al. (2013), we suggest that Tweedie can often be a solution instead of the more restrictive Poisson distribution when more complex two-stages approaches (hierarchical modelling of occurrence then abundance) is to be avoided for the sake of simplicity.

4.2 Sampling design: we recommend investing extra effort in sampling sites with larger counts
For four of the five studied species, sites with larger counts had the most influence on the slope estimate. Consequently, depending on the species, most of the needed information on trend might lie in the sites with larger mean counts. This rather intuitive result can relate to a basic macro-ecological rule predicting that higher local abundances are located on larger habitat patches or larger sites (Suet et al. 2021). Yet, comprehensive waterbird censuses are more difficult or limited in larger wetlands, which are often partly, or even largely invisible and/or inaccessible. A recommendation for EAF sampling design would be to increase sampling efforts on large wetlands relative to smaller wetlands (Sayoud et al. 2017); in particular, Africa has a number of huge wetlands, including the Merja Zerga in Morocco, the Banc d'Arguin in Mauritania, the Senegal Delta in both Mauritania and Senegal or the Bijagos archipelago in Guinea Bissau. Because sites with larger counts could be the most important in driving trends, we recommend that these few larger sites be given priority for correction of detection bias, count error estimation and power analysis of census effort in the framework of a locally agreed sampling design. Indeed, we question approaches aiming at obtaining total count for such larges sites versus approaches favouring sampling (Le Moullec et al. 2017). Finally, we also recommend checking or possibly conduct (by e.g. remote-sensing, Suet et al. 2021) wetland inventories in order to include in priority wetlands of larger surface area.

4.3 detection bias: we recommend investigating options for a double-sampling design
We encompassed both under-detection of presence and underestimation of abundance in this general detection issue. Each topic could deserve a discussion per se. Both kind of detection bias are obvious barriers to proper population size (Williams et al. 2002, Kellner & Swihart 2014) and trend (Kéry et al. 2009) estimations. However, as shown here, any mathematical link between true abundance and count other than mere proportionality (like, e.g. increased counting error in larger bird clusters) may also increase bias in trend estimation. A number of methods have been developed to mitigate or estimate detection bias (Williams et al. 2002, Dénes et al. 2015), and one possibility for large scale surveys such as the EAF monitoring scheme seems to be the double sampling approach (Pollock et al. 2002, Bart et al. 2004). Double sampling is currently applied in several waterbird schemes in America (e.g. Jiménez et al. 2023) and large-scale citizen science project (Johnston et al. 2018). In short, such an approach integrates a correction of counts from a random subsample of sites within the large-scale standard uncorrected monitoring scheme. Correction of counts can be designed through e.g. replicates of counts among observers (Bird et al. 2014) or short time-intervals. Correction of counts can

also be improved by incorporating observation covariates like counted surface area, number of observers, count duration, expertise of observers (Johnston et al. 2018).

4.4 Choice of imputation method: we recommend investing in statistical research and software programming

Temporal slope estimates differed markedly between TRIM and LORI imputation models. It should be noted that the LORI imputation method used in this work is particularly designed and therefore adapted to integrate a large number of environmental covariates to predict waterbird counts (Dakki et al. 2021). Thus, adding only geographical coordinates as covariates like we did in this study, is not very informative. LORI method would probably be much more relevant if we would have included environmental variables documenting (e.g. habitats, fragmentation etc.) and years (climatic covariates of breeding success and winter movements). We argue that this capacity to integrate as many covariates as needed is an asset by comparison to other imputation packages, However, both TRIM and LORI only model Poisson-distributed count data and are therefore less suited to model more overdispersed data with zero-inflation like IWC count data. A more overdispersed model like the Poisson-lognormal is currently being investigated to improve imputation performance over LORI's (Chiquet et al. 2021). Overall, we believe integrating large numbers of predictors into large-scale datasets like the EAF monitoring scheme as well as building more flexible modelling frameworks than the imputation models currently available is a robust approach for imputing IWC data (Dakki et al. 2021). Further developments are needed to improve imputation performances dedicated to EAF or more generally IWC data. Finally, depending on whether the EAF and, more generally, the IWC monitoring schemes, require imputation performance vs ecological inference, an option may be to use machine learning (Fink et al. 2023). We argue this debate is beyond the scope of the present work, but we may stick on the ecological inference side for now because our ultimate goal being conservation or sustainable management, we may need to permanently understand the complex interactions between birds and their environment.

4.5 Spatial autocorrelation: we recommend including a spatial structuration in the modelling framework

Waterbird count distribution across EAF wetlands showed little ($0 >$ Moran's $I < 0.1$) but significant spatial autocorrelation in model residuals for all five species. The Gaussian random field was often the most efficient way to reduce the spatial autocorrelation in the model residuals, independently from the mesh resolution or constrain. Integration of GRF in R packages is rapidly increasing, facilitating its use (cf mgcv and sdmTMB; Wood 2011, Anderson et al. 2022). Alternatively, both basis penalty smoothing and random intercept resulted in decent effects on spatial autocorrelation and could be used more extensively (Folliot et al. 2022). However, if all the tested models could mitigate spatial autocorrelation, only two of them totally controlled for it. It is noteworthy that none of the spatial models used in this work integrated environmental predictors besides latitude, longitude and year. However, one major issue in reducing spatial autocorrelation is to collect and integrate relevant environmental predictors and covariates into population models (Dormann et al. 2007). We believe that a next step to this work would be include environmental predictors that can explain spatial dependencies (Zuur and Ieno 2021).

As environmental predictors and covariates become increasingly available, they can be used to improve analyses of large waterbird count datasets like the EAF monitoring scheme by contributing simultaneously to inform:

- autocorrelation models,
- imputation models
- detectability models

Collecting and securing large datasets of environmental predictors and covariates, including sampling effort covariates, stands thus as a major priority for the EAF monitoring scheme.

1.6 *further challenges and improvements*

Uncertainty estimation and propagation is one of the most challenging issues not addressed in our study. Uncertainty can result from methodologic (epistemic) and natural (aleatoric) processes. Uncertainty occurs at multi scale and should be propagated to avoid overconfidence in population trend estimation. Both kinds of uncertainty can be propagated using the same method of Monte Carlo (Soize 2017, Zhang 2021) which we recommend introducing in analyses of the EAF monitoring scheme, as proper estimation of uncertainty is crucial for sound policy decision making.

Finally, the growth, maintenance, optimization and quality control of the EAF database and more generally the IWC database is probably the most important recommendation we want to highlight. Citizen science databases are now becoming more and more numerous and rich but also possible subject to more errors, biases and lack of precision than standard professional or expert databases (Bird et al. 2014). We argue that daily quality control of incoming data must be implemented to the highest standards in terms of expert staff and automation if the EAF monitoring scheme is to be cost-effective in the long term. Overall, this work must be seen as a plea for maintaining and strengthening the IWC. The general methodological recommendations we make here are only aiming at improving the IWC and their value for decision making. The rich data already collected by thousands of observers contain a wealth of information to be used for the conservation of birds and wetlands.

Acknowledgements

We would like to thank all the (thousands!) field observers who participated in the IWC in countries of the EAF, and who made this study possible by providing the data for this analysis; we would like also to thank all the national coordinators and Tom Langendoen from Wetlands International for taking care of the data quality and sharing and to Marc van, Roomen and Sjoerd Duijns from SOVON for initiating and supporting this work. We are also grateful to Khalil Baddour for his help with data management.

This work was produced upon request of Sovon, Dutch Centre for Field Ornithology as part of the project ‘Innovations for Migratory Bird Monitoring Along the East Atlantic Flyway’. This project is coordinated by the Coastal & Marine Union (EUCC) and financed by the Technical Support Instrument (TSI) as operated by the European Commision’s Directorate General for Structural Reform Support (DG REFORM). The project is requested by the Dutch Ministry of Agriculture, Nature and Food Quality also on behalf of its counterparts in Germany and Denmark."

References

Adde, A., C. Casabona i Amat, M. J. Mazerolle, M. Darveau, S. G. Cumming, and R. B. O'Hara. 2021. Integrated modeling of waterfowl distribution in western Canada using aerial survey and citizen science (eBird) data. Ecosphere 12(10): e03790. 10.1002/ecs2.3790

AEWA Secretariat, 2013. Report of the 5th Session of the Meeting of the Parties (MOP5) to the African-Eurasian Migratory Waterbird Agreement (AEWA), 14 – 18 May 2012, La Rochelle, France.

Amano, T., Székely, T., Sandel, B., Nagy, S., Mundkur, T., Langendoen, T., Blanco, D., Soykan, C. U., & Sutherland, W. J. (2018). Successful conservation of global waterbird populations depends on effective governance. Nature, 553(7687), 199. https://doi.org/10.1038/nature25139

Anderson, S. C., Ward, E. J., English, P. A., & Barnett, L. A. (2022). sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. bioRxiv, 2022-03.

Atkinson, P. W., Austin, G. E., Rehfisch, M. M., Baker, H., Cranswick, P., Kershaw, M., ... & Maclean, I. M. (2006). Identifying declines in waterbirds: the effects of missing data, population variability and count period on the interpretation of long-term survey data. Biological Conservation, 130(4), 549-559.

Aubry, P., Quaintenne, G., Dupuy, J., Francesiaz, C., Guillemain, M., & Caizergues, A. (2023). On using stratified two-stage sampling for large-scale multispecies surveys. *Ecological Informatics*, *77*, 102229.

Bart, J., Droege, S., Geissler, P., Peterjohn, B., & Ralph, C. J. (2004). Density estimation in wildlife surveys. Wildlife Society Bulletin, 32(4), 1242-1247.

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., ... & Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. Biological Conservation, 173, 144-154.

Bogaart, A.P., van der Loo, M. & Pannekoek, J. 2016. Package "rtrim". https://CRAN.R-project.org/package=rtrim.

Buckland, S. T., & Johnston, A. (2017). Monitoring the biodiversity of regions: Key principles and possible pitfalls. *Biological Conservation*, *214*, 23-34.

Chiquet, J., Mariadassou, M., & Robin, S. (2021). The Poisson-lognormal model as a versatile framework for the joint analysis of species abundances. Frontiers in Ecology and Evolution, 9, 588292.

Dakki M, Robin G, Suet M, Qninba A, El Agbani MA, Ouassou A, El Hamoumi R, Azafzaf H, Rebah S, Feltrup-Azafzaf C, Hamouda N, Ibrahim WA, Asran HH, Elhady AA, Ibrahim H, Etayeb K, Bouras E, Saied A, Glidan A, Habib BM, Sayoud MS, Bendjedda N, Dami L, Deschamps C, Gaget E, Mondain-Monval JY & Defos du Rau P (2021) Imputation of incomplete large-scale monitoring count data via penalized estimation. *Methods in Ecology and Evolution*. *12*(6), 1031-1039.

Dale M.R.T. & Fortin M-J. (2002) Spatial autocorrelation and statistical tests in ecology, *Écoscience*, 9:2, 162-167. http://dx.doi.org/10.1080/11956860.2002.11682702

Dambly, L. I., Isaac, N. J., Jones, K. E., Boughey, K. L., & O'Hara, R. B. (2023). Integrated species distribution models fitted in INLA are sensitive to mesh parameterisation. Ecography, e06391.

Dénes, F. V., Silveira, L. F., & Beissinger, S. R. (2015). Estimating abundance of unmarked animal populations: accounting for imperfect detection and other sources of zero inflation. Methods in Ecology and Evolution, 6(5), 543-556.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609-628.

Ficetola, G. F., Romano, A., Salvidio, S., & Sindaco, R. (2018). Optimizing monitoring schemes to detect trends in abundance over broad scales. Animal Conservation, 21(3), 221-231.

Fink, D., Johnston, A., Strimas-Mackey, M., Auer, T., Hochachka, W. M., Ligocki, S., ... & Rodewald, A. D. (2023). A double machine learning trend model for citizen science data. Methods in Ecology and Evolution, 14(9), 2435-2448.

Folliot, B., Caizergues, A., Tableau, A., Souchay, G., Guillemain, M., Champagnon, J., & Calenge, C. (2022). Assessing spatiotemporal variation in abundance: A flexible framework accounting for sampling bias with an application to common pochard (Aythya ferina). Ecology and Evolution, 12(4), e8835.

Gaget, E., Galewski, T., Jiguet, F., & Le Viol, I. (2018). Waterbird communities adjust to climate warming according to conservation policy and species protection status. *Biological conservation*, *227*, 205-212.

Green, A.J., Elmberg, J., 2014. Ecosystem services provided by waterbirds. Biological Reviews 89, 105-122.

Hartig, F. 2022. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. Version 0.4.

Hughes, B. B., Beas-Luna, R., Barner, A. K., Brewitt, K., Brumbaugh, D. R., Cerny-Chipman, E. B., ... & Carr, M. H. (2017). Long-term studies contribute disproportionately to ecology and policy. *BioScience*, *67*(3), 271-281.

Jiménez, A., Brush, J., Zambrano, R., Warraich, N., & Korosy, M. (2023). Using Double Sampling to Estimate the Population of Least Terns (Sternula antillarum) Nesting on Florida's Rooftops. Waterbirds, 45(4), 440-449.

Johnston, A., Fink, D., Hochachka, W. M., & Kelling, S. (2018). Estimates of observer expertise improve species distributions from citizen science data. Methods in Ecology and Evolution, 9(1), 88-97.

Johnston, A., Hochachka, W. M., Strimas-Mackey, M. E., Ruiz Gutierrez, V., Robinson, O. J., Miller, E. T., ... & Fink, D. (2021). Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. *Diversity and Distributions*, *27*(7), 1265-1277.

Kellner, K. F., & Swihart, R. K. (2014). Accounting for imperfect detection in ecology: a quantitative review. PloS one, 9(10), e111436.

Kéry, M., Dorazio, R. M., Soldaat, L., Van Strien, A., Zuiderwijk, A., & Royle, J. A. (2009). Trend estimation in populations with imperfect detection. *Journal of Applied Ecology*, *46*(6), 1163-1172.

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, *14*(10), 551-560.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm?. Ecology, 74(6), 1659-1673.

Lehikoinen, A., Jaatinen, K., Vähätalo, A., Clausen, P., Crowe, C., Deceuninck, B., Hearn, R., Holt, C. A., Hornman, M., Keller, V., Nilsson, L., Langendoen, T., Tománková, I., Wahl, J., & Fox, A. D. (2013). Rapid climate driven shifts in winter distributions of three common waterbird species. Global Change Biology, 19, 2071–2081.

Le Moullec, M., Pedersen, Å. Ø., Yoccoz, N. G., Aanes, R., Tufto, J., & Hansen, B. B. (2017). Ungulate population monitoring in an open tundra landscape: distance sampling versus total counts. Wildlife Biology, 2017(1), 1-11.

Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73(4), 423–498.

Łopucki, R., Kiersztyn, A., Pitucha, G., & Kitowski, I. (2022). Handling missing data in ecological studies: Ignoring gaps in the dataset can distort the inference. Ecological Modelling, 468, 109964.

Magurran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. M., Elston, D. A., Scott, E. M., ... & Watt, A. D. (2010). Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in ecology & evolution*, *25*(10), 574-582.

Miller, D. L., Burt, M. L., Rexstad, E. A., & Thomas, L. (2013). Spatial models for distance sampling data: recent developments and future directions. Methods in Ecology and Evolution, 4(11), 1001-1010.

Moore, C. T., & Kendall, W. L. (2004). Costs of detection bias in index–based population monitoring. Animal Biodiversity and Conservation, 27(1), 287-296.

Nagy, S., Langendoen, T., Frost, T. M., Jensen, G. H., Markones, N., Mooij, J. H., ... & Suet, M. (2022). Towards improved population size estimates for wintering waterbirds. Ornithologische Beobachter, 119(4).

Nakagawa, S., & Freckleton, R. P. (2008). Missing in action: The dangers of ignoring missing data. Trends in Ecology & Evolution,23(11), 592–596. https://doi.org/10.1016/j.tree.2008.06.014

Nichols, J. D., Hines, J. E., Sauer, J. R., Fallon, F. W., Fallon, J. E., & Heglund, P. J. (2000). A double-observer approach for estimating detection probability and abundance from point counts. The Auk, 117(2), 393-408.

O'Hara, R. B., & Kotze, D. J. (2010). Do not log-transform count data. Methods in Ecology and Evolution, 1, 118-122.

Onkelinx, T., Devos, K., & Quataert, P. (2017). Working with population totals in the presence of missing data comparing imputation methods in terms of bias and precision. Journal of Ornithology, 158(2), 603–615. https://doi.org/10.1007/s1033 6-016-1404-9

Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L., & Sauer, J. R. (2002). Large scale wildlife monitoring studies: statistical methods for design and analysis. Environmetrics: The official journal of the International Environmetrics Society, 13(2), 105-119.

Rangel, T.F.L.V.B., Diniz-Filho, J.A.F., Bini, L.M., 2010. SAM: a comprehensive application for Spatial Analysis in Macroecology. Ecography 33, 46-50.

Robin, G., Josse, J., Moulines, É., & Sardy, S. (2019). Low-rank model with covariates for count data with missing values. Journal of Multivariate Analysis, 173, 416-434.

Rocchini, D., Tordoni, E., Marchetto, E., Marcantonio, M., Barbosa, A. M., Bazzichetto, M., ... & Malavasi, M. (2023). A quixotic view of spatial bias in modelling the distribution of species and their diversity. *npj Biodiversity*, *2*(1), 10.

Sanz-Pérez, A., Sollmann, R., Sardà-Palomera, F., Bota, G., & Giralt, D. (2020). The role of detectability on bird population trend estimates in an open farmland landscape. Biodiversity and Conservation, 29(6), 1747-1765.

Sayoud, M. S., Salhi, H., Chalabi, B., Allali, A., Dakki, M., Qninba, A., ... & Defos du Rau, P (2017). The first coordinated trans-North African mid-winter waterbird census: the contribution of the International Waterbird Census to the conservation of waterbirds and wetlands at a biogeographical level. *Biological Conservation*, *206*, 11-20.

Soize, C. (2017). Uncertainty quantification. Springer International Publishing AG.

Stephenson, P. J., Bowles-Newark, N., Regan, E., Stanwell-Smith, D., Diagana, M., Höft, R., Abarchi, H., Abrahamse, T., Akello, C., Allison,H., Banki, O., Batieno, B., Dieme, S., Domingos, A., Galt, R., Githaiga, C. W., Guindo, A. B., Hafashimana, D. L. N., Hirsch, T.,… Thiombiano, A. (2017). Unblocking the flow of biodiversity data for decision-making in Africa. Biological Conservation, 213, 335–340. https://doi.org/10.1016/j.biocon.2016.09.003

Stroud, D. A., Pirot, J. Y., & Smart, M. (2022). The International Waterfowl (and Wetlands) Research Bureau: c. 1945–1995. *Wildfowl*, *72*, 3-60.

Suet, M., Lozano-Arango, J. G., Defos Du Rau, P., Deschamps, C., Abdalgader Mohammed, M. A., Elbashary Adam, E., ... & Mondain-Monval, J. Y. (2021). Improving waterbird monitoring and conservation in the Sahel using remote sensing: a case study with the International Waterbird Census in Sudan. Ibis, 163(2), 607-622.

Tirozzi, P., Orioli, V., Dondina, O., Kataoka, L., & Bani, L. (2022). Population trends from count data: Handling environmental bias, overdispersion and excess of zeroes. Ecological Informatics, 69, 101629.

Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecological Applications, 13(6), 1790-1801.

van Roomen M., Agblonon G., Langendoen T., Citegetse G., Diallo A. Y., Gueye K., van Winden E. & Luerssen G. (eds.) 2020. Simultaneous January 2020 waterbird census along the East Atlantic Flyway: National Reports. Wadden Sea Flyway Initiative p/a Common Wadden Sea Secretariat, Wilhelmshaven, Germany, Wetlands International, Wageningen, The Netherlands, BirdLife International, Cambridge, United Kingdom.

van Roomen M., Citegetse G., Crowe O., Dodman T., Hagemeijer W., Meise K., & Schekkerman H. 2022 (eds). East Atlantic Flyway Assessment 2020.The status of coastal waterbird populations and their sites. Wadden Sea Flyway Initiative p/a CWSS, Wilhelmshaven, Germany, Wetlands International, Wageningen, The Netherlands, BirdLife International, Cambridge, United Kingdom

van Strien, J. Pannekoek, W. Hagemeijer & T. Verstrael, 2004 - A loglinear Poisson regression method to analyse bird monitoring data. In: Anselin, A. (ed.) Bird Numbers 1995, Proceedings of the International Conference and 13th Meeting of the European Bird Census Council, Pärnu, Estonia. Bird Census News 13 (2000):33-39

Wetlands International, 2010. Guidance on waterbird monitoring methodology: Field Protocol for waterbird counting. wpe.wetlands.org (accessed 5.10.12).

White, E. R. (2019). Minimum time required to detect population trends: the need for long-term monitoring programs. *BioScience*, *69*(1), 40-46.

Williams, B. K., Nichols, J. D., & Conroy, M. J. (2002). Analysis and management of animal populations. Academic press.

Wood, S.N. 2011 Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Volume 73, Issue 1, Pages 3–36, https://doi.org/10.1111/j.1467-9868.2010.00749.x

Zhang, J. (2021). Modern Monte Carlo methods for efficient uncertainty quantification and propagation: A survey. Wiley Interdisciplinary Reviews: Computational Statistics, 13(5), e1539.

Zipkin, E. F., Leirness, J. B., Kinlan, B. P., O'Connell, A. F., & Silverman, E. D. (2014). Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation. *Statistical Methodology*, *17*, 67-81.

Zuur, A.F. and Ieno, E.N. 2021. AEWA Waterbird Trend Analysis: Methodology Review. Contracted report to JNCC (Contract No. C20-0419-1492).