

Working with population totals in the presence of missing data comparing imputation methods in terms of bias and precision

Thierry Onkelinx¹  · Koen Devos¹ · Paul Quataert¹

Received: 9 December 2015 / Revised: 30 August 2016 / Accepted: 10 October 2016 / Published online: 24 October 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Missing observations in water bird censuses are commonly handled using the Underhill index or the birdSTATs tool which enables the use of TRIM under the hood. Multiple imputation is a standard technique for handling missing data that is rarely used in the field of ecology, but is a well known statistical technique in the fields of medical and social sciences. The purpose of this paper is to compare these three methods in terms of bias and variance. The bias in the Underhill method depends on the algorithm and starting values. birdSTATs and multiple imputation are unbiased in the case of missing values that are missing completely at random; more missing values implies less information, and so wider confidence intervals are expected as the missingness increases. The Underhill method and birdSTATs tool underestimate the variance; omitting data from a complete dataset and applying the Underhill index or birdSTATs tool results in smaller confidence intervals. Multiple imputation with an adequate imputation model provides wider confidence intervals. Biased parameter estimates with underestimated variance can potentially lead to incorrect management and policy conclusions. Hence, we dissuade the use of Underhill indices or the birdSTATs tool to handle missing data, rather we suggest that multiple imputation is a more robust alternative, even in suboptimal conditions.

Keywords Missing data · Multiple imputation · Monitoring of biodiversity · Survey design and analysis

Zusammenfassung

Gesamtbestandszahlen trotz fehlender Daten – ein Vergleich von Imputationsmethoden hinsichtlich systematischer Abweichungen und Genauigkeit

Fehlende Beobachtungen bei Wasservogelzählungen werden üblicherweise gehandhabt, indem der Underhill-Index oder birdSTATs angewendet werden. Letzteres nutzt TRIM. Multiple Imputation ist eine Standardmethode für die Handhabung fehlender Daten, die in der Medizin und in den Sozialwissenschaften wohlbekannt ist, in der Ökologie jedoch kaum angewendet wird. Das Ziel dieses Artikels ist es, diese drei Methoden hinsichtlich systematischer Abweichungen und Varianz zu vergleichen. Systematische Abweichungen beim Underhill-Index hängen vom Algorithmus und von den Ausgangswerten ab. birdSTATs und multiple Imputation sind frei von systematischen Fehlern, falls Daten vollkommen zufällig fehlen (MCAR). Fehlen mehr Werte, bedeutet dies weniger Information, und folglich erwarten wir umso größere Konfidenzintervalle, je mehr Werte fehlen. Der Underhill-Index und birdSTATs unterschätzen allerdings die Varianz. Lässt man aus einem an sich kompletten Datensatz Daten aus und wendet den Underhill-Index oder birdSTATs an, werden die Konfidenzintervalle kleiner. Multiple Imputation mit einem geeigneten Imputationsmodell liefert hingegen größere Konfidenzintervalle. Systematisch abweichende Parameterschätzungen mit unterschätzter Varianz führen möglicherweise zu falschem Management und Leitlinienabschlüssen. Daher raten wir vom Gebrauch des

Communicated by F. Bairlein.

Electronic supplementary material The online version of this article (doi:10.1007/s10336-016-1404-9) contains supplementary material, which is available to authorized users.

✉ Thierry Onkelinx
thierry.onkelinx@inbo.be

¹ Research Institute for Nature and Forest (INBO),
Kliniekstraat 25, 1070 Anderlecht, Belgium

Underhill-Index oder birdSTATs zur Handhabung fehlender Daten ab. Multiple Imputation ist selbst unter suboptimalen Bedingungen eine robustere Alternative.

Introduction

R.A. Fisher wrote: ‘The best solution to handle missing data is to have none’ (Nakagawa and Freckleton 2008). However, in practice some missingness in data is inevitable. For example, long-term waterbird monitoring is prone to have missing counts because it requires a lot of human resources due to its large span in both time and space. Missingness in the data complicates data analysis and can introduce bias if not accounted for correctly.

As early as the end of the eighties, Rubin (1987) introduced the multiple imputation procedure as a general approach to handle missing values correctly. Although multiple imputation analysis is well established in fields such as medical and social sciences (9625 citations according to Google Scholar), its use is only emerging in the field of ecology (Nakagawa and Freckleton 2008), and its application in the analysis of population monitoring data is still limited (e.g. Blanchong et al. 2006; Rice et al. 2009).

In comparison, the Underhill index (Underhill and Prys-Jones 1994) and the TRIM (TRENds and Indices for Monitoring data) software package (Pannekoek and Van Strien 2005; Van Strien et al. 2001, 2004) are two more popular ways of handling missing data in population monitoring. A search on Google Scholar revealed 118 citations for Underhill and Prys-Jones (1994), including those of Cayford and Waters (1996), Goss-Custard et al. (1998), Perez-Arteaga (2004), Atkinson et al. (2006), Rendón et al. (2008) and Dalby et al. (2013) who apply the Underhill index on bird data and Kirkman et al. 2007 and Wright et al. 2013) who apply it to data on mammals. Dennis et al. (2013) apply a similar technique but with a different model on insect data. Another search on Google Scholar revealed 91 citations for Van Strien et al. (2001), including 34 for Van Strien et al. (2004) and 310 for various versions of Pannekoek and Van Strien (2005). A few examples are Gregory et al. (2005, 2007), Soldaat et al. (2007), Ward et al. (2009) and Musilová et al. (2014) on birds and Conrad et al. (2004), Van Dyck et al. (2009) and Staats and Regan (2014) on insects. TRIM indices for several countries are sometimes combined into supranational indices (Gregory et al. 2008; Vorisek et al. 2008). The European Bird Census Council requires member organisations to use TRIM or birdSTATs (van der Meij 2013), which is an access shell around TRIM, to produce national indices (<http://www.ebcc.info/index.php?ID=13>).

The aim of this paper was to investigate how well these two popular methods perform and effectively deal with missing data in comparison to the more generic multiple imputation technique for handling missing data (Rubin 1987). To our knowledge, no thorough evaluation of the birdSTATs and Underhill methods exist. Yet, as the many references cited above illustrate, they are applied frequently in population monitoring and related fields. ter Braak et al. (1994) mention Rubin (1987) and claim that Underhill and Prys-Jones (1994) use the principles of multiple imputation. In our opinion, this is not the case as we will demonstrate in this paper.

Materials and methods

We start by describing how census data are transformed into annual population indices in the presence of missing data. Then we describe how we simulate the census data and the patterns of missingness. Finally we introduce the different imputation methods and how their performance is evaluated.

The census data and population indices

Count data (C_{ijk})

In many bird monitoring programmes, sites are revisited multiple times per year. Hence, we denote the observed number of birds C_{ijk} with three indices: year $i : 1, \dots, n_y$, month $j : 1, \dots, n_m$ and site $k : 11, \dots, n_s$.

In this paper we assume the counts follow a negative binomial distribution (Eq. 1) with mean or expected value μ_{ijk} and size parameter θ . The variance (Eq. 3) depends on the mean and a size parameter θ , which governs the overdispersion. Small values of θ imply strong overdispersion, while for θ going to infinity, with the negative binomial distribution reduced to a Poisson distribution. We refer the interested reader to Zuur et al. (2012) for an introduction to negative binomial distribution.

$$C_{ijk} \sim NB(\mu_{ijk}, \theta) \quad (1)$$

$$\mathbb{E}(C_{ijk}) = \mu_{ijk} \quad (2)$$

$$\text{Var}(C_{ijk}) = \mu_{ijk} + \frac{\mu_{ijk}^2}{\theta} \quad (3)$$

The monthly population size (P_{ij})

If the monitoring programme covers most of the relevant sites in the region, the data collection is close to a census and it therefore is logical to interpret the sum (Eq. 4) over all

sites within 1 year i and month j as the monthly population P_{ij} . Otherwise, under the assumption that there are no major changes in the population distribution, P_{ij} can be considered as an indicator of the population size in the region.

$$P_{ij} = \sum_{k=1}^{n_s} C_{ijk} \tag{4}$$

The annual population index

For each year, there are n_m P_{ij} -values. To construct a single annual population index (API) from these monthly population totals, we fit a negative binomial generalised linear model (glm.nb) (Venables and Ripley 2002) with effects for year (γ_i) and month (ξ_j).

$$P_{ij} \sim \text{NB}(v_{ij}, \theta) \tag{5}$$

$$\log(v_{ij}) = \gamma_i + \xi_j \tag{6}$$

In this model, e^{γ_i} is an estimate of the API for year i at the reference month j (where $\xi_j = 0$). In fact, the above model corrects the P_{ij} for the average seasonal pattern, making the year-to-year variations more comparable to uncover trends or other patterns.

Note that API is dependent of the choice of the reference month. For the sake of simplicity, in this paper, we take the first month as the reference. However, for real-life applications, we recommend choosing the most relevant month based on expert knowledge, such as the month with the highest expected numbers.

From the fitted model, it is also possible to derive the confidence limits for the parameter γ_i (LCL $_i$; UCL $_i$) and (after transformation) for API.

The complete, observed and augmented dataset

In practice, not all sites are visited at every point in time, resulting in missing counts. However, to estimate API, it is necessary to have a complete dataset, and various imputation methods have been developed to fill in the missing data. In the paper, we distinguish between three types of datasets: complete, observed and augmented. A complete dataset has counts for every combination of year, month and site; the observed dataset has missing counts for some combinations of year, month and site; and the augmented dataset is an observed dataset in which the missing counts are replaced by imputed values by an imputation method.

Comparing imputation methods

To compare the imputation methods, i.e. methods used for filling in missing values, we first simulate complete

datasets and introduce missing counts according to a certain pattern to obtain observed datasets. Then, we apply various imputation methods to augment the data, from which the API along with confidence limits can be estimated. Finally, we compare the statistical properties of the estimates in terms of bias and precision to assess the performance of the imputation methods with respect to each other. In the following sections, we discuss each step successively, starting with data generation.

Simulating the complete and observed data

Long-term waterbird monitoring in Flanders

This paper is inspired by our work involving the long-term monitoring of waterbirds in Flanders, Belgium Waterbird Monitoring in Flanders (WMF). Volunteers count the number of wintering birds mid-monthly from October to March. The first coordinated counts go back to the late 1960s. In 1990, a reorganisation of the project resulted in an improved standardisation of the methods. In this paper, we use data that were collected from October 1990 until March 2014, spanning 24 winters, with data on 1225 sites. Overall 44 % of the counts are missing. We used this dataset to obtain sensible parameter values for our simulations. Also, the pattern of missingness is used to simulate missingness not at random (see section “[Setting the pattern of missing counts](#)” for a definition).

The data generating model for the complete data

We simulate 200 complete datasets consisting of 100 sites (n_s), 24 years (n_y) and 6 months per year (n_m). The counts follow a negative binomial distribution with mean μ_{ijk} and size parameter θ (Eq. 7).

The mean μ_{ijk} is on the log-scale linked to effects of year i , month j and site k (Eq. 8). The global effect of year i consists of an intercept β_0 , a linear trend $\beta_1 T_i$ and a random walk t_i (Eq. 9). Together, these terms reflect a long-term trend with year-to-year variation. The global effect of month j is a sine wave with a period of 12 months (M_j), fixed amplitude β_2 and variable phase shift ϕ_i (Eq. 10). The sine wave reflects a seasonal pattern allowing for a shift among the years. The site effect k is a combination of an intercept b_k Eq. (11) and a random walk along the year b_{ik} (Eq. 12). The intercept b_k allows for systematic differences in importance among sites, while the random walk b_{ik} allows the relative importance of the sites to change over time. ϵ_{ijk} is a random noise term in the log-scale.

$$C_{ijk} \sim \text{NB}(\mu_{ijk}, \theta) \tag{7}$$

$$\begin{aligned} \log(\mu_{ijk}) = & \beta_0 + \beta_1 T_i + t_i \\ & + \beta_2 \sin\left(2\pi \frac{M_j}{12} + \phi_i\right) \\ & + b_k + b_{ik} + \epsilon_{ijk} \end{aligned} \quad (8)$$

$$\Delta t_i = t_i - t_{i+1} \sim N(0, \sigma_{rw}) \quad (9)$$

$$\phi_i \sim N(\phi_0, \sigma_\phi) \quad (10)$$

$$b_k \sim N(0, \sigma_s) \quad (11)$$

$$\Delta b_{ik} = b_{ik} - b_{(i+1)k} \sim N(0, \sigma_{rw.s}) \quad (12)$$

$$\epsilon_{ijk} \sim N(0, \sigma_e) \quad (13)$$

We generate a new complete dataset for each of the 200 simulation runs. The parameters θ , β_0 , β_1 and β_2 are fixed for all simulations (Eq. 8). The other parameters are based on random numbers from independent normal distributions with zero mean and fixed standard deviation (Eqs. 9–13). Table 1 specifies the values for the fixed parameters and fixed standard deviations.

Figure 1 illustrates how API changes over time during 20 runs of the simulation. Each line corresponds to one complete dataset from which the API can be estimated directly because the data are complete. Some of the lines show dramatic changes, while others are quite smooth. Hence, the choice of model parameters covers a broad range of possible situations for which we will test the performance of the imputation methods.

Setting the pattern of missing counts

We refer to Nakagawa and Freckleton (2011) for an introduction on the differences between missing completely at random (MCAR) and missing not at random (MNAR).

First, we implement the MCAR scheme, i.e. the probability of a count being missing depends neither on observed nor on unobserved values. Hence, each count has the same probability of being missing. We can obtain the required fraction of missingness just by taking a simple random sample without replacement from the complete dataset. The fraction of missing observations is set to 0, 1, 5, 25, 50 or 75 % of the number of observations in the complete dataset.

In practice, the assumption of MCAR are likely to be violated (Nakagawa and Freckleton 2008), especially in a long-term monitoring project with many volunteers and many sites. Therefore, we also test an MNAR-pattern based on the observed missing counts in the WMF [for technical details, see section A of the Electronic Supplementary Material (ESM)]. For this scheme, the proportion of missingness is slightly variable. On average, 56 % of the counts is missing (range 45–64 %).

Table 1 Choice of θ , the regression parameters β and the standard deviations σ in the data generating model (Eqs. 7–13)

Regression	Value	Standard deviation	Value
β_0	2	σ_{rw}	0.1
β_1	0.01	σ_ϕ	0.2
β_2	1	σ_s	1
ϕ_0	0	$\sigma_{rw.s}$	0.02
θ	2	σ_e	0.01

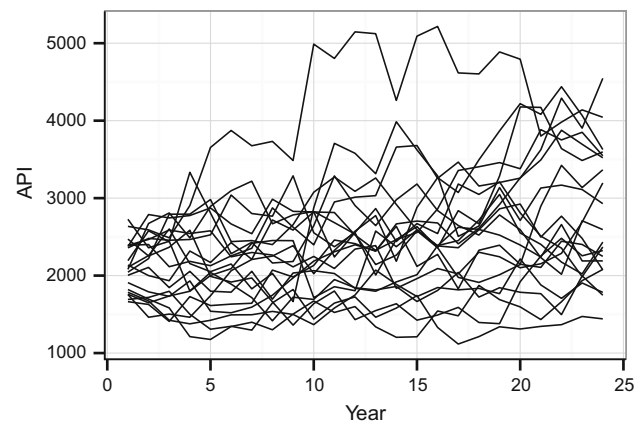


Fig. 1 Trend in annual population index (API) calculated from 20 simulated complete datasets for the parameters in Table 1 over a period of 24 years, with 100 sites visited 6 times a year (the winter months)

The single imputation methods

We first present two commonly used “single” imputation methods in population monitoring that create an augmented dataset only once. In the section “Multiple imputation”, we introduce the principle of “multiple” imputation and implement it for the bird census data. For any imputation method, an imputation model should be fitted to the available data (i.e. the observed dataset) in order to predict the missing observations and augment the observed dataset. We postpone the definition of the imputation model used here to the section “Multiple imputation”.

Underhill method

Underhill and Prys-Jones (1994) describe the Underhill method. The construction of an augmented dataset is based on an iterative algorithm with three main steps repeated until convergence. These steps are: (1) fit an imputation model to the augmented dataset; (2) predict with the imputation model the counts missing and round them to the nearest integer; (3) replace the previously imputed value with the new prediction only if the new prediction is larger

than the previously imputed value; otherwise keep the previously imputed value.

With the augmented dataset, it is possible to fit the model (Eqs. 5, 6) to estimate γ_i and its confidence interval (LCL_{*i*}; UCL_{*i*}), and, after transformation, API.

The algorithm requires an initialisation of the augmented dataset. Underhill and Prys-Jones (1994) suggest replacing the missing values with the mean of all available counts. An alternative is starting from zero. Underhill and Prys-Jones (1994) use an imputation model with independent effects for year, month and site with a quasi Poisson likelihood, fitted with an expectation–maximisation algorithm. Our imputation model uses the same effects but is based on a negative binomial regression model (glm.nb).

As acknowledged by Underhill and Prys-Jones (1994), the third step is susceptible to the introduction of bias because the algorithm only allows imputed values to increase. We propose an altered version for which in each iteration the previously imputed value is replaced with the new prediction irrespective of its value. For the simulations, we evaluate both alternatives, i.e. the original and altered, to demonstrate that the altered version is indeed an improvement.

birdSTATs

We prepared TRIM data files and command files according to the birdSTATs guidelines (van der Meij 2013). TRIM cannot handle multiple measurements per year–site combination. The workaround in van der Meij (2013) is to take the sum of the available counts per year–site combination and use the inverse of the number of available counts per year–site combination as weight factor.

We use the estimates in the ‘time TOTALS’ section from the TRIM output file which are expressed in the original count scale. In order to be able to compare birdSTATs with the Underhill method (and the multiple imputation method, see section “Multiple imputation”), we transform the TRIM output to the log scale with $\gamma_i = \log(I_i)$, $LCL_i = \log(I_i - 1.96 SE_i)$ and $UCL_i = \log(I_i + 1.96 SE_i)$ with I_i = the ‘Imputed’ parameter for year i and SE_i = the associated ‘std.err.’ parameter.

Multiple imputation

The general principle

The Underhill index and birdSTATs statistical programme are single imputation methods and, as a consequence, cannot acknowledge the extra uncertainty associated with data imputation. In contrast, Rubin (1987) suggests a general philosophy to assess the imputation variability by repeating the imputation step L times, yielding L

augmented datasets, from which the parameters in Eqs. 5 and 6 can be estimated L times. Rubin (1987) also provides the equations for combining the L estimates into one single estimate and its standard error. For example, for γ_i :

$$\bar{\gamma}_i = \frac{1}{L} \sum_{l=1}^L \hat{\gamma}_{il} \tag{14}$$

$$\sigma_i^2 = \frac{1}{L} \sum_{l=1}^L \hat{\sigma}_{il}^2 + \frac{L+1}{L} \sum_{l=1}^L \frac{(\hat{\gamma}_{il} - \bar{\gamma}_i)^2}{L-1} \tag{15}$$

The average of $\hat{\gamma}_{il}$ over all L imputation sets is (Eq. 14) an unbiased estimate of the true index γ_i . The squared standard error of this average ($\bar{\gamma}_i$) is (Eq. 15) the sum of two components (Rubin 1996). The first is the average of the squared standard errors $\hat{\sigma}_{il}$ of the individual $\hat{\gamma}_{il}$ over the L imputations and is a measure of within-imputation variability. The second is the variance of the $\hat{\gamma}_{il}$ among the augmented datasets, i.e. the between-imputation variability, multiplied with a correction factor $\frac{L+1}{L}$. This component will be large when the imputation step is highly variable. As the imputation model governs the variability of the predictions to augment the data, its choice is crucial to obtaining the correct confidence intervals.

The basic imputation model

With multiple imputation, there is a large flexibility to construct augmented datasets. A first “basic” imputation model (Eqs. 16–19) was chosen to make the results comparable with the Underhill method and birdSTATs. This model is a negative binomial generalised linear mixed model (glmm.nb) with year (β_i^*) and month (β_j^*) as fixed effects and site (b_k^*) as random effect. The observed dataset is sufficient to fit the model, so no initialisation is necessary.

$$\text{Count}_{ijk}^* \sim \text{NB}(\mu_{ijk}^*, \theta^*) \tag{16}$$

$$\log(\mu_{ijk}^*) = \eta_{ijk}^* \tag{17}$$

$$\eta_{ijk}^* = \beta_0^* + \beta_i^* + \beta_j^* + b_k^* \tag{18}$$

$$b_k^* \sim N(0, \sigma_s^*) \tag{19}$$

The imputed values are generated in two steps. The first step is to sample a value pair (η_{ijk}^*, θ^*) from their distributions $\eta_{ijk}^* \sim N(\hat{\eta}_{ijk}^*, \sigma_{\hat{\eta}_{ijk}^*})$ and $\theta^* \sim \text{Gamma}(\beta_{\hat{\theta}^*}, \hat{\theta}^*)$. These distributions combine the point estimates $\hat{\eta}_{ijk}^*$ and $\hat{\theta}^*$ with their model uncertainties. The second step is to sample the imputed value from a negative binomial distribution with mean $\mu_{ijk}^* = e^{\eta_{ijk}^*}$ and size $\hat{\theta}^*$. The first step takes the model uncertainty into account and the second step the natural variability.

More complex imputation models

Multiple imputation easily allows the incorporation of alternative, more complex imputation models (e.g. interactions, covariates, ...). To investigate the effects of the imputation model, we consider two alternative imputation models: “complex” and “true mean”.

The complex imputation model extends the basic model (Eq. 18) with two random intercepts, namely, b_{ij}^* , the interaction between year and month (Eq. 20), and b_{ik}^* , the interaction between year and site (Eq. 21). Then η_{ijk}^* becomes Eq. 22. The additional random effects allow the effects of month and site to change over the years without assumptions on their relation.

The true mean imputation model is not a model, but uses $\hat{\eta}_{ijk} = \log \mu_{ijk}$ and $\hat{\theta} = \theta$ of the data generating model (Eq. 7). Such information is of course only available in case of a simulation study. It is a surrogate for a perfect imputation model. The standard error will only be affected by the natural variability of the observations as there is no model uncertainty.

$$b_{ij}^* \sim N(0, \sigma_{ym}^*) \tag{20}$$

$$b_{ik}^* \sim N(0, \sigma_{ys}^*) \tag{21}$$

$$\eta_{ijk}^* = \beta_0^* + \beta_i^* + \beta_j^* + b_k^* + b_{ij}^* + b_{ik}^* \tag{22}$$

Evaluating the performance of the methods

Each run x of the simulation generates a complete and observed dataset according to a certain pattern of missingness. We calculate the API and its confidence interval for the complete dataset as the reference to compare with the API derived from the augmented datasets generated by the imputation methods. Thus, the API on the complete dataset of run x serves as a performance reference of the API as calculated by the imputation methods for run x .

We use two quality measures. The first is the bias (Eq. 23). An unbiased method will have $\mathbb{E}(\text{Bias}) = 0$. The second is the relative width of the 95 % confidence interval (RWCI) (Eq. 24) which assesses the precision. An augmented dataset should yield more uncertainty than the complete dataset. More uncertainty translates into wider confidence intervals. Hence, we expect that $\mathbb{E}(\text{RWCI}) > 1$ and RWCI should increase with the proportion of missing counts in the data.

$$\text{Bias}_i = \gamma_{i_{\text{method}}} - \gamma_{i_{\text{complete}}} \tag{23}$$

$$\text{RWCI}_i = \frac{UCL_{i_{\text{method}}} - LCL_{i_{\text{method}}}}{UCL_{i_{\text{complete}}} - LCL_{i_{\text{complete}}}} \tag{24}$$

Amano et al. (2012) combine bias and precision (RCWI) into a mean squared error (MSE) in order to have a single

performance measure. The downside of this strategy is that a high bias and low precision might give a MSE similar to a low bias and high precision. Looking at both the bias and precision allows for a finer interpretation of the results.

Software

All data were prepared and analysed with R version 3.0.1 R Core Team (2013) under Scientific Linux 6.1 with packages MASS (Venables and Ripley 2002) for the glm.nb and INLA (Rue et al. 2009) for the glmm.nb. TRIM is Windows-only software. Hence, only the preparation and post-processing of the TRIM files was done in R under Linux. The actual calculation of the indices used TRIM version 3.53 (Pannekoek and Van Strien 2005) run under Windows 7. The default number of replications L for the multiple imputation was set to 199.

Results

To facilitate comparison and to minimise the complexity of the figures, we present only one variant of each imputation method, unless otherwise specified. We selected the altered Underhill algorithm because it has a similar performance in terms of precision and the best performance in terms of bias. For the multiple imputation method (Eqs. 16–19), we start with the basic imputation model to match the complexity of the model as used by birdSTATs and the Underhill method. With respect to BirdSTATs, one should take into account that this method cannot estimate API, but only the yearly average. As a consequence, bias (Eq. 23) is less an issue. Still, the length of the confidence intervals (Eq. 24) remains a meaningful criterion to assess precision.

Overall comparison of the three methods

For an overall comparison of the three methods, we test for a real life setting, i.e. with an MNAR-pattern as derived from the WMF.

Bias

Boxplots of the bias for the different methods are given in Fig. 2a. The (altered) Underhill method has on average the smallest bias. So our proposal to change the algorithm is successful (see also Fig. 5). The spread of the bias over the simulation (width of the box plots) is similar to multiple imputation. The birdSTATs method yields a positive bias. This is the result of two effects. First, the API estimates the population at a reference month. The first month is used as reference month (section “The annual population index”)

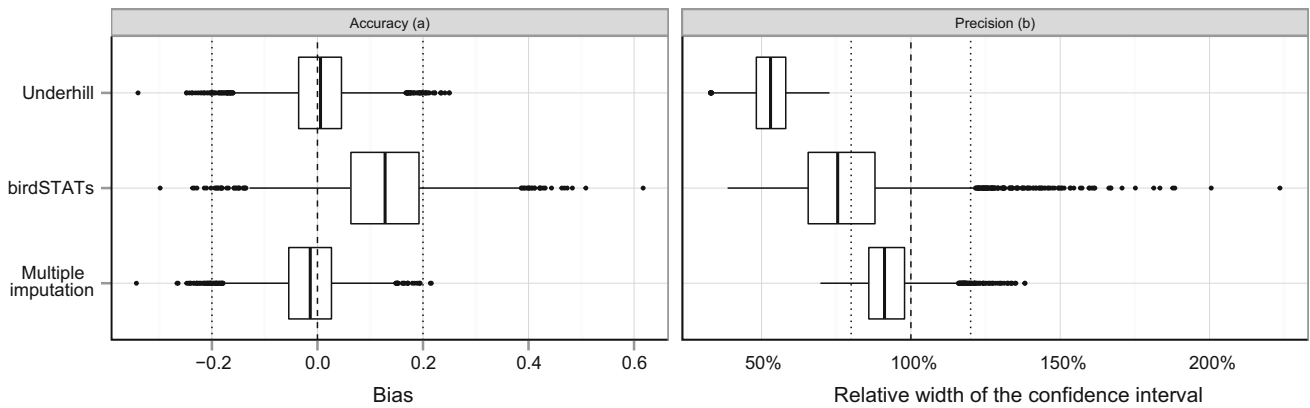


Fig. 2 Boxplots of the bias (a) and relative width of the confidence interval (b) of API for the different methods. Boxplots are based on 200 complete datasets with 24 years, 6 months per year and 100 sites. On average, 56 % of the counts is missing not at random (MNAR).

The Underhill method uses the alternative algorithm with mean as initialisation. Multiple imputations uses the basic imputation model with $L = 199$ imputations

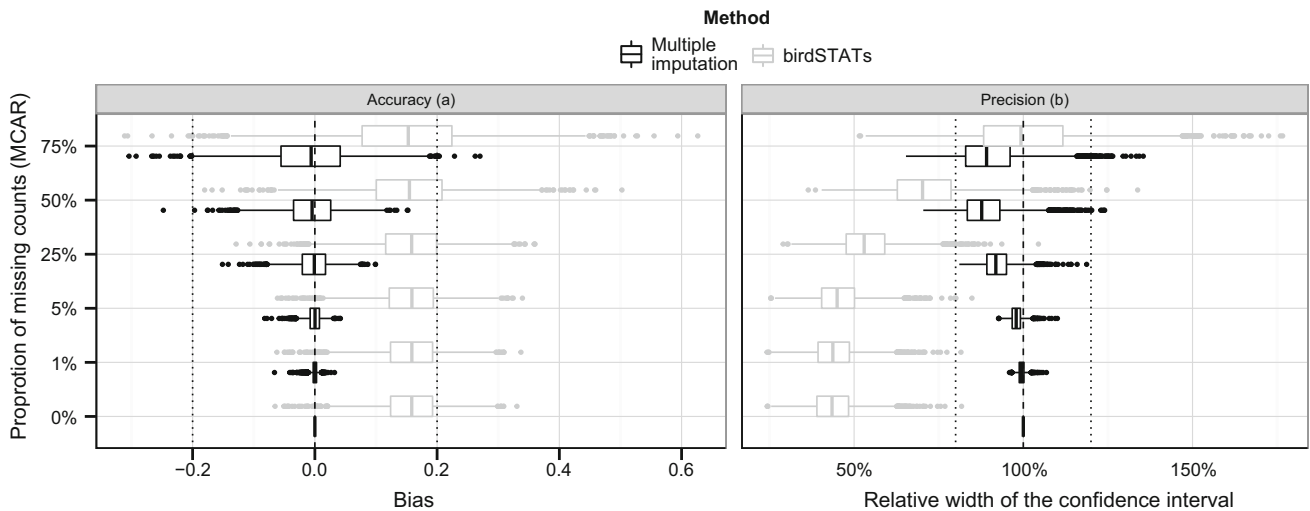


Fig. 3 Influence of the proportion of missing completely at random (MCAR) counts on the bias (a) and relative width of the confidence interval (b) of API when using either birdSTATs or multiple

imputation with $L = 199$ imputations from the basic imputation model. Results are based on 200 simulated complete datasets with 24 years, 6 months per year and 100 sites

and the first month has the lowest counts (section “[The data generating model for the complete data](#)”). Secondly, birdSTATs uses an arithmetic mean, whereas API uses a geometric mean which is smaller than the arithmetic mean. The multiple imputation method is slightly downward biased. Further results on this issue can be found section “[The impact of the imputation model](#)” where we propose a more complex and a perfect imputation model.

It is noted here that Figs. 2, 3 and 4 all have one dashed and two dotted vertical lines. The dashed line indicates the reference based on the complete dataset, and the dotted lines indicate arbitrary values based on the range of the values on x axes in Fig. 4. The sole purpose is to aid the comparison with Figs. 2 and 3 as the ranges of the x axes in these latter figures are much wider.

Precision

Boxplots of the relative width of the confidence interval are shown in Fig. 2b. The (altered) Underhill method strongly underestimates the width of the confidence intervals by about 50 % in comparison to the complete data. This method replaces missing values iteratively with predictions of the imputation model fitted on the available data. As a consequence, the variability in the augmented dataset is too low in comparison to real data. Hence, confidence intervals calculated on the augmented dataset will be too small. The birdSTATs tool also leads to a strong underestimation of the confidence intervals (70 %). The estimate is based on counts averaged per year-site combination which reduces the variation in the dataset, again resulting in too small

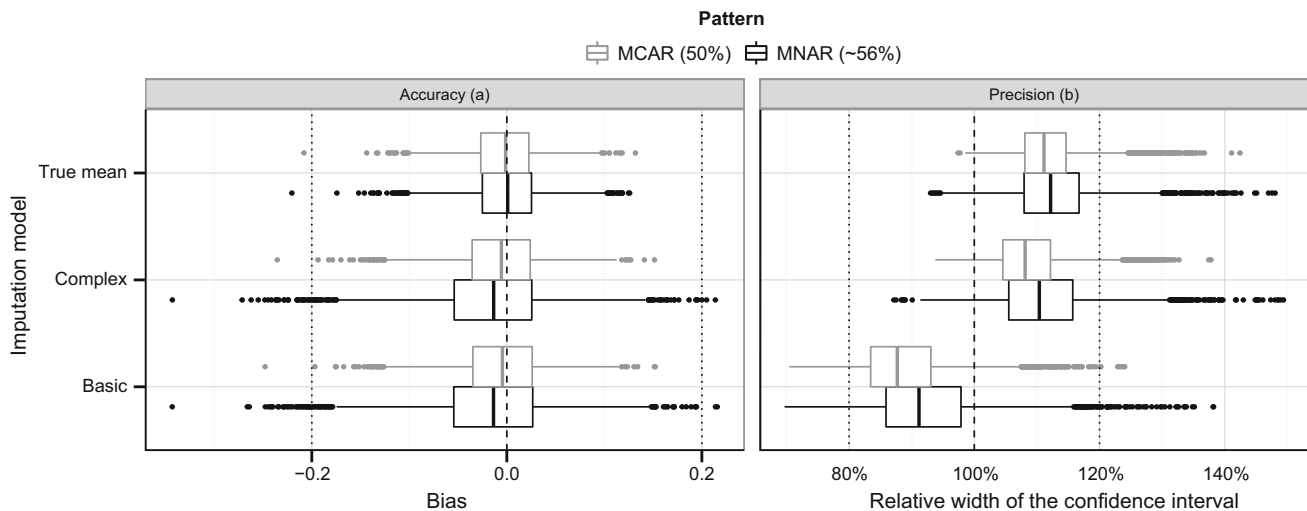


Fig. 4 Effect of the imputation model on bias (a) and precision (b) when using multiple imputation ($L = 199$ imputations). Based on 200 simulated datasets with 24 years, 6 months per year and 100 sites

confidence intervals. Multiple imputation outperforms both the Underhill method and birdSTATs. Yet, the confidence intervals determined using this method are smaller than those of the complete dataset. We explore this issue further in section “[The impact of the imputation model](#)” when we discuss the impact of the imputation model.

Influence of the proportion of missing counts

We excluded the Underhill method from further analysis because the computational burden was very high and it strongly underestimates the confidence limits (see section “[Precision](#)”). We therefore further only compare BirdSTATs and multiple imputation. To investigate the impact of the proportion of missingness on bias and precision, we choose MCAR as this technique is straightforward to set the proportion of missingness at a fixed value; this step is harder to control for MNAR.

Bias

The influence of the proportion of MCAR missing counts on the bias of API when using either birdSTATs or multiple imputation is shown in Fig. 3a as boxplots. On average, the bias of both methods is not affected by the proportion of missingness. Multiple imputation is unbiased. The bias of birdSTATs is explained in section “[Bias](#)”. The variability in bias among the simulations increases with the proportion of missing counts. This increase is stronger for multiple imputation than for birdSTATs.

Precision

The influence of the proportion of MCAR missing counts on the relative width of the confidence interval of API

when using either birdSTATs or multiple imputation is shown in Fig. 3b. On average, the confidence limits of birdSTATs increase with an increasing proportion of missing counts, but they are much too small. The average RCWI for 1 % missing counts is 44 %. Note that birdSTATs uses counts averaged over the 6 months per year. The standard error of the average of n numbers is $1/\sqrt{n}$ of the standard deviation of the n numbers. $1/\sqrt{6} = 0.41$, which is the same magnitude as the RWCI. For multiple imputation, the width of the confidence intervals is better but slightly decreases with the proportion of missing data. A high proportion of missingness implies an augmented dataset dominated by imputed values, which reduce the variability.

The impact of the imputation model

In this section, we explore the impact of the imputation model. Up to now we have chosen the basic imputation model to evaluate the methods on an equal footing. However, with multiple imputation, it is straightforward to adapt and improve the imputation model. This flexibility allows exploitation of the full potential of multiple imputation. Figure 4 shows a comparison of the basic, complex and true mean model (section “[More complex imputation models](#)”) for two patterns of missingness: MCAR pattern with 50 % missing counts and MNAR with on average 56 % missing counts.

Bias

As shown in Fig. 4a, the “true mean” model is unbiased. The two other models indicate a small downward bias for the observed pattern (on average -0.015). The MNAR

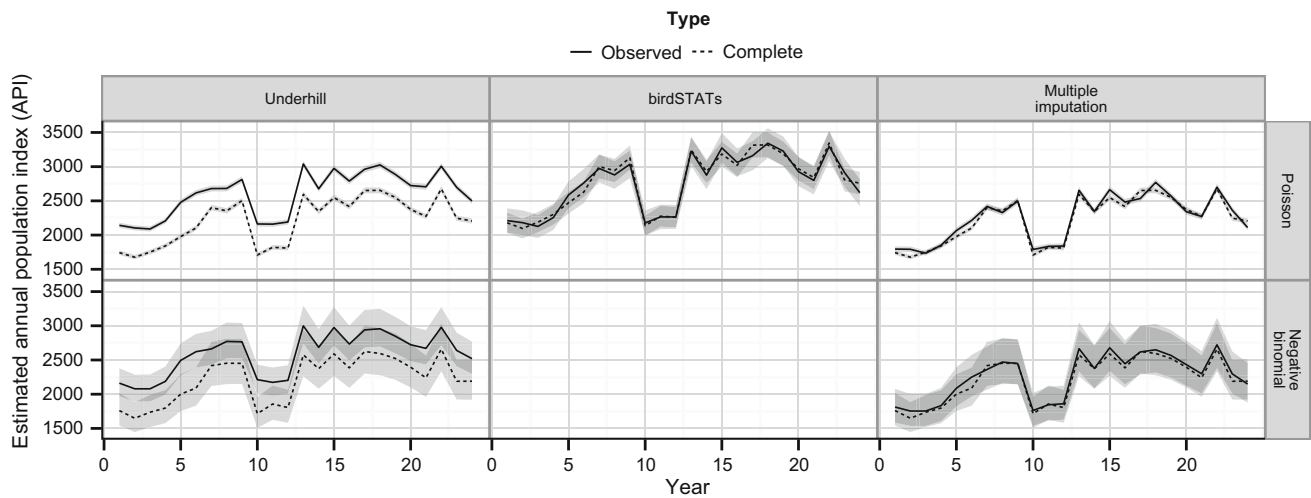


Fig. 5 Differences in confidence intervals of API based on several techniques for an example dataset with 24 years, 6 months per year and 100 sites. *Complete* Dataset is without missing data, *Observed* 25 % missing counts (MCAR). Poisson and Negative binomial refer

to the distribution used in the analysis. Multiple imputation uses the complex imputation model with $L = 199$ imputations; the Underhill method uses the original algorithm with the mean as starting value

pattern has several sites with short time series (see ESM Appendix A). For these sites, extrapolation risks to introduce bias.

Precision

As mentioned in section “More complex imputation models” the true mean model is only affected by the natural variation. Hence we assume that the RWCI of the true mean model reflects the correct value after imputation. This value is larger than 1 because imputing missing data increases the uncertainty compared to the complete dataset. In comparison to the true mean model, the basic imputation model underestimates the confidence intervals, as shown in Fig. 4b. The complex imputation model performs better. Still, the confidence intervals are slightly underestimated, as any model will always somewhat smooth the true variability.

Note that, for the true mean model, the width of the confidence intervals after imputation, assuming 50 % MCAR, is on average only 12 % larger than those of the complete dataset. This demonstrates that multiple imputation model is a very powerful tool.

Comparison of the methods on an example dataset

To obtain more insight and to visualise the implications of the above evaluation in the original count scale (and not in the log-scale), we now compare the yearly indices the different methods with one simulation using a balanced design consisting of 100 sites, 24 years and 6 months per year and with counts simulated according to Eqs. 7–13 with a MCAR-pattern of 25 % (Fig. 5).

We compare the original Underhill algorithm with the mean as a starting value, the birdSTATs method and multiple imputation with the basic imputation model ($L = 199$ imputations). Both the Underhill method and multiple imputation assume a negative binomial distribution for the counts (see Eqs. 5, 16). However, birdSTATs is based on TRIM which allows only for a Poisson distribution. To facilitate discussion, we also implemented the Poisson distribution for the Underhill and multiple imputation methods.

Switching between the Poisson and the negative binomial has a small effect on the estimates of the coefficient—and hence on the bias. However, the standard error of those coefficients is larger in the negative binomial model because it captures overdispersion (Agresti 2002), as demonstrated by Fig. 5 where the confidence intervals of the multiple imputation and Underhill method are very narrow with the Poisson distribution and wide with the negative binomial distribution.

The panels in Fig. 5 demonstrate that the misspecification of the distribution in birdSTATs is an important reason for the underestimation of the confidence intervals. A Poisson distribution does not allow for overdispersion, and in the simulated data there are two sources of extra-Poisson variability. First, the simulated observations themselves stem from a negative binomial. In addition, there is a seasonal effect; hence, within a year the counts are not uniformly distributed. This seasonal effect also causes extra variability which is not accounted for by birdSTATs because just the average over a year is used as input.

Finally, it should be noted from Fig. 5 that the original Underhill method overestimates API. Consequently, the

correction proposed to replace all imputed values in each cycle all imputed values, is necessary. In Fig. 2, the altered Underhill method is unbiased.

Discussion

The choice of the imputation method

Single or multiple imputation

It was not easy to compare the different methods on an equal footing as each method has its own particularities, thereby preventing a perfect match. For a fair comparison of all methods with each other, we started with a “basic” imputation model (Eqs. 16–19) that assumed a constant effect of year i over all months and sites, a constant effect of month j over all years and sites and a constant effect of site k over all years and months.

From these multiple perspectives, the main message is clear and straightforward. Multiple imputation is capable of capturing the extra uncertainty caused by the missingness in the data under different scenarios of missingness (MCAR and MNAR, increasing proportion of missingness). Even with the basic (and insufficiently complex) model, the confidence limits are (much) broader than those of the two other imputation methods, as it should be. In addition, the multiple imputation approach is very flexible and allows for improvements in the imputation model, such that the confidence limits are close to a true mean model.

Underhill method

The original Underhill method (Underhill and Prys-Jones 1994) is biased (Fig. 5), although it is was quite easy to adapt the algorithm to correct for the bias (Fig. 2; ESM section B). However, the Underhill method remains a single imputation approach, resulting in an underestimation of the variability in the model (Fig. 2).

birdSTATs

In the original setting, TRIM was designed for Poisson counts with one observation per year–site combination, with all sites having the same global trend. In this case, the estimates are unbiased for MCAR. Also, $RWCI > 1$ and increases with the proportion of missing counts (ESM section C). The situation totally changes in the case of multiple observations per year–site combination, when there is seasonality or when overdispersion is present. The confidence intervals are too small because on the one hand observations are aggregated and on the other hand seasonality and overdispersion are ignored. In addition,

Amano et al. (2012) state a uniform trend model is a strong and potentially wrong assumption that could lead to inaccurate estimates of population indices.

Averaging multiple counts per year–site combination artificially reduces the fraction of missingness to only those year–site combinations without any counts at all. For example, a year–site combination with only one count out of six is not considered as missing. In addition, the seasonality in the data is not accounted for; each year–site combination with one count is treated in the same way, regardless whether the count is from the start, middle or end of the season.

The flexibility of multiple imputation

In addition to its excellent statistical properties, multiple imputation can easily incorporate new elements.

The choice of imputation model

Regarding the multiple imputation method, Fig. 4 shows that the quality of the imputation model is important. The critique of Amano et al. (2012) on assuming the same temporal trend for all sites applies to imputation models as well. The risk of bias is not affected by the imputation model, but the width of the confidence interval is (Rubin 1996). Creating a perfect imputation model-like “true mean” is impossible in practice, and any practical model will lead to some reduction of variability in the augmented dataset in comparison to the complete dataset. A good model captures most of the patterns in the data and will minimise the variance reduction with respect to the “true mean” model. There is a fine line between sufficient complexity to describe the patterns accurately and too much complexity, leading to overfitting and possibly increased model uncertainty.

The proportion of missing data

The use of multiple imputation is not restricted by the proportion of missing data. A diminishing number of observations due to an increased proportion of missing counts leads to more model uncertainty in the imputation model, resulting in larger confidence intervals (Fig. 2b; Nguyen 2016). If the number of observations is too low to make meaningful statements, the confidence intervals will be wide. Hence, we can apply multiple imputation regardless of the number of observations, conditionally by taking the confidence intervals into account when making statements on the parameter estimates.

Other parameter estimates

We limited this paper to the estimation of yearly indices, but the multiple imputation technique can be applied to any

kind of parameter estimates, such as year maxima, smoothed yearly indices, pairwise comparison of yearly indices, linear trends in moving windows, proportion of the population in special protection areas, among others. Any analysis which can be performed on a complete dataset can be applied on an observed dataset after it has been augmented with multiple imputation (Rubin 1996). Therefore, multiple imputation should be used more often in ecology in the case of missing observations (Nakagawa and Freckleton 2008).

Possible extensions and alternatives

Missingness in the covariates

In our comparison we assume that only the counts have missing data and that all covariates are observed. This is a reasonable assumption when only using covariates that are fixed by the design of the monitoring and hence are never missing. In our case these covariates are site, year and month of sampling. Nakagawa and Freckleton (2011) illustrate how missing data in the covariates can be handled.

Bayesian models

Amano et al. (2012) and Johnson and Fritz (2014) give examples on how to use Bayesian Hierarchical Models to estimate indices on population totals. The benefit of the Bayesian technique is that it handles missing observation gracefully without the need for imputation. The downside is that such models tend to be more complex and more computationally intensive to run. Johnson and Fritz (2014) make their algorithms available as an R package “agTrend”, but unfortunately these cannot handle multiple observations per year–site combination. Skilled users can of course write their own algorithms to fit the relevant Bayesian model and run it in software like BUGS, JAGS, STAN, among others.

The number of imputations

The computing power in the 1990s and the first decade of the present century was a bottleneck for computational intensive methods, such as multiple imputation. Nowadays vast computing power is readily available. We have successfully applied multiple imputation on a single core machine for datasets with up to 1000 sites and time series of 24 years with 6 samples per year using $L = 199$ imputations. The imputation step in the algorithm is a so-called ‘embarrassingly parallel problem’ (Burns 1990). It is straightforward to run the imputation step in parallel on multi-core computer systems. Examples on how to do this

can be found in the “multimput” R package (Onkelinx et al. 2016).

The required computing time depends on two factors: the size of the design (number of sites, years and months) and the number of imputations. The size of the design affects all methods in a similar way. The number of imputations will determine the required extra computing time. In this paper, the default number of imputation is ($L = 199$), which is fairly large. This number has hardly any influence on the bias and only a small influence on the variance of the relative width of the confidence intervals (ESM section D). Therefore, a smaller number of imputation sets can be sufficient.

Graham et al. (2007) recommend running at least $L = 100$ imputation sets when the computational effort is reasonable. For these authors, the required number of imputation sets depends on the proportion of missing data and on the acceptable level of power falloff. Only $L = 3$ imputation sets are sufficient with 10 % missing observations and an acceptable power falloff of <5 %. 70 % missing observations with an acceptable power falloff of <1 % require at least $L = 40$ imputation sets (Graham et al. (2007), Table 5).

Implementation in R

We provide the code of the paper in the “multimput” R-686 package to replicate the simulations (Onkelinx et al. 2016), which is freely available on GitHub under a GPL-3 license (<https://github.com/inbo/multimput>. README provides installation instructions and some usage examples). The goal of this package is twofold. First, it allows other researchers to reproduce our results (Mislán et al. 2016); secondly, a set of more generic functions allow ecologists to apply multiple imputation on population monitoring data. The vignette of the package provides some examples on how to use the package. The “multimput” package still requires some basic “R” knowledge from the user. It has no graphical user interface (GUI) unlike TRIM (Pannekoek and Van Strien 2005) or birdSTATs (vander Meij 2013). However, the GPL-3 license of the “multimput” package allows others to build and distribute a GUI on top of the “multimput” package.

There are other R packages for multiple imput, such as “Amelia” (Honaker et al. 2011), “mi” (Su et al 2011) and “mice” (van Buuren and Groothuis-Oudshoorn 2011). The focus of the “multimput” package is on generalised linear mixed models which are not available in the other packages.

Acknowledgements The computational resources (Stevin Super-computer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Centre), funded by Ghent

University, the Hercules Foundation and the Flemish Government—department EWI. We are very grateful to the two reviewers and the Subject Editor for their constructive suggestions and for their proposed corrections to improve this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agresti A (2002) Categorical data analysis. John Wiley and Sons, Hoboken
- Amano T, Okamura H, Carrizo SF, Sutherland WJ (2012) Hierarchical models for smoothed population indices: the importance of considering variations in trends of count data among sites. *Ecol Indic* 13(1):243–252. doi:10.1016/j.ecolind.2011.06.008
- Atkinson PW, Austin GE, Rehfish MM, Baker H, Cranswick P, Kershaw M, Robinson J, Langston RH, Stroud DA, Turnhout CV, Maclean IM (2006) Identifying declines in waterbirds: The effects of missing data, population variability and count period on the interpretation of long-term survey data. *Biol Conserv* 130(4):549–559. doi:10.1016/j.biocon.2006.01.018
- Blanchong JA, Joly DO, Samuel MD, Langenberg JA, Rolley RE, Sausen JF (2006) White-Tailed deer harvest from the chronic wasting disease eradication zone in South-Central Wisconsin. *Wildl Soc Bull* 34(3):725–731. doi:10.2193/0091-7648(2006)34[725:WDHFTC]2.0.CO;2
- Burns EM (1990) Multiple and replicate item imputation in a complex sample survey. In: Anderson-Brown M (ed) 1990 Annual Research Conference. U.S. Department of Commerce, Bureau of the Census, Arlington, pp 655–665
- Cayford JT, Waters R (1996) Population estimates for waders Charadrii wintering in Great Britain, 1987/88–1991/92. *Biol Conserv* 77(1):7–17. doi:10.1016/0006-3207(95)00114-X
- Conrad KF, Woiwod IP, Parsons M, Fox R, Warren MS (2004) Long-term population trends in widespread British moths. *J Insect Conserv* 8(2–3):119–136
- Dalby L, Soderquist P, Christensen TK, Clausen P, Einarsson A, ElMBERG J, Fox AD, Holmqvist N, Langendoen T, Lehtikoinen A, Lindstrom A, Lorentsen SH, Nilsson L, Poysa H, Rintala J, Sigfusson AP, Svenning JC (2013) The status of the Nordic populations of the Mallard (*Anas platyrhynchos*) in a changing world. *Ornis Fennica* 90(1):2–15
- Dennis EB, Freeman SN, Brereton T, Roy DB (2013) Indexing butterfly abundance whilst accounting for missing counts and variability in seasonal pattern. *Methods Ecol Evol* 4(7):637–645. doi:10.1111/2041-210X.12053
- Goss-Custard JD, Ross J, McGrorty S, le V dit, Duell SEA, Caldwell RWG, West AD (1998) Locally stable wintering numbers in the Oystercatcher *Haematopus ostralegus* where carrying capacity has not been reached. *Ibis* 140(1):104–112. doi:10.1111/j.1474-919X.1998.tb04546.x
- Graham JW, Olchowski AE, Gilreath TD (2007) How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci* 8(3):206–213. doi:10.1007/s11121-007-0070-9
- Gregory RD, van Strien A, Vorisek P, Gmelig Meyling AW, Noble DG, Foppen RPB, Gibbons DW (2005) Developing indicators for European birds. *Philos Trans R Soc Lond Ser B Biol Sci* 360(1454):269–288. doi:10.1098/rstb.2004.1602
- Gregory RD, Vorisek P, Noble DG, Van Strien A, Klvanová A, Eaton M, Gmelig Meyling AW, Joys A, Foppen RPB, Burfield IJ (2008) The generation and use of bird population indicators in Europe. *Bird Conserv Int* 18(S1):S223–S244. doi:10.1017/S0959270908000312
- Gregory RD, Vorisek P, Van Strien AJ, Gmelig Meyling AW, Jiguet F, Fornasari L, Reif J, Chylarecki P, Burfield IJ (2007) Population trends of widespread woodland birds in Europe. *Ibis* 149:78–97. doi:10.1111/j.1474-919X.2007.00698.x
- Honaker J, King G, Blackwell M (2011) Amelia II: a program for missing data. *J Stat Softw* 45(7):1–47. doi:10.18637/jss.v045.i07
- Johnson DS, Fritz L (2014) agTrend: a bayesian approach for estimating trends of aggregated abundance. *Methods Ecol Evol* 5(10):1110–1115. doi:10.1111/2041-210X.12231
- Kirkman S, Oosthuizen W, Meÿer M, Kotze P, Roux JP, Underhill L (2007) Making sense of censuses and dealing with missing data: trends in pup counts of Cape fur seal *Arctocephalus pusillus* for the period 1972–2004. *Afr J Mar Sci* 29(2):161–176. doi:10.2989/AJMS.2007.29.2.2.185
- Mislan K, Heer JM, White EP (2016) Elevating the status of code in ecology. *Trends Ecol Evol* 31(1):4–7. doi:10.1016/j.tree.2015.11.006
- Musilová Z, Musil P, Zouhar J, Bejcek V, Slastný K, Hudec K (2014) Numbers of wintering waterbirds in the Czech Republic: long-term and spatial-scale approaches to assess population size. *Bird Study* 61(3):321–331. doi:10.1080/00063657.2014.919573
- Nakagawa S, Freckleton RP (2008) Missing inaction: the dangers of ignoring missing data. *Trends Ecol Evol* 23(11):592–596. doi:10.1016/j.tree.2008.06.014
- Nakagawa S, Freckleton RP (2011) Model averaging, missing data and multiple imputation: a case study for behavioural ecology. *Behav Ecol Sociobiol* 65(1):103–116. doi:10.1007/s00265-010-1044-7
- Nguyen V (2016) Modelling uncertainty in population monitoring data. Dissertation, The University of Sydney, Sydney
- Onkelinx T, Devos K, Quataert P (2016) multimp: using multiple imputation to address missing data. doi: 10.5281/zenodo.48423. URL:<https://github.com/inbo/multimp>
- Pannekoek J, Van Strien A (2005) TRIM 3 Manual (Trends & Indices for Monitoring data). Statistics Netherlands, Voorburg, The Netherlands
- Perez-Arteaga A (2004) Wildfowl population trends in Mexico, 1961–2000: a basis for conservation planning. *Biol Conserv* 115(3):343–355. doi:10.1016/S0006-3207(03)00088-0
- R Core Team (2013) R: A language and environment for statistical computing. Version 3.0.1. R Foundation for Statistical Computing, Vienna URL:<http://www.r-project.org/>
- Rendón M, Green A, Aguilera E, Almaraz P (2008) Status, distribution and long-term changes in the waterbird community wintering in Doñana, south-west Spain. *Biol Conserv* 141(5):1371–1388. doi:10.1016/j.biocon.2008.03.006
- Rice CG, Jenkins KJ, Chang WY (2009) A sightability model for mountain goats. *J Wildl Manag* 73(3):468–478. doi:10.2193/2008-196
- Rubin DB (1987) Multiple imputation for nonresponse in surveys. John Wiley and Sons, New York
- Rubin DB (1996) Multiple imputation after 18+ years. *J Am Stat Assoc* 91(434):473–489
- Rue H, Martino S, Lindgren F, Simpson D, Riebler A (2009) INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximation. <http://www.r-inla.org/download>
- Soldaat L, Visser H, Roomen M, Van Strien A (2007) Smoothing and trend detection in waterbird monitoring data using structural time-series analysis and the Kalman filter. *J Ornithol* 148(S2):351–357. doi:10.1007/s10336-007-0176-7

- Staats WT, Regan EC (2014) Initial population trends from a 5-year butterfly monitoring scheme. *J Insect Conserv* 18(3):365–371. doi:10.1007/s10841-014-9644-6
- Su Y, Gelman A, Hill J, Yajima M (2011) Multiple Imputation with Diagnostics (mi) in R: Opening windows to the black box. *J Stat Softw*. doi:10.18637/jss.v045.i02
- ter Braak CJF, Van Strien AJ, Meijer R, Verstrael TJ (1994) Analysis of monitoring data with many missing values: which method? In: 12th Int Conference of IBCC and EOAC. Statistics Netherlands, Voorburg/Heerlen & SOVON, Beek-Ubbergen, pp 663–673
- Underhill LG, Prys-Jones RP (1994) Index numbers for waterbird populations. I. Review and methodology. *J Appl Ecol* 31(3):463–480. doi:10.2307/2404443
- van Buuren S, Groothuis-Oudshoorn K (2011) Mice: multivariate imputation by chained equations in R. *J Stat Softw* 45(3):1–67. doi:10.18637/jss.v045.i03
- van der Meij T (2013) birdSTATs. Species Trends Analysis Tool (STAT) for European bird data. <http://www.ebcc.info/wpimages/video/BirdSTATS21.zip>
- Van Dyck H, Van Strien AJ, Maes D, CaM Van Swaay (2009) Declines in common, widespread butterflies in a landscape under intense human use. *Conserv Biol* 23(4):957–965. doi:10.1111/j.1523-1739.2009.01175.x
- Van Strien AJ, Pannekoek J, Gibbons DW (2001) Indexing European bird population trends using results of national monitoring schemes: a trial of a new method. *Bird Study* 48(2):200–213
- Van Strien AJ, Pannekoek J, Hagemeyer W, Verstrael TJ (2004) A loglinear poisson regression method to analyse bird monitoring data. In: Anselin A (ed) *Bird Numbers 1995/Proc Int Conf and 13th Meeting of the European Bird Census Council*. European Bird Census Council, vol 13, pp 33–39
- Venables WN, Ripley BD (2002) *Modern applied statistics with S*, 4th edn. Springer, Berlin Heidelberg New York
- Vorisek P, Gregory RD, Strien AJV, Gmelig Meyling A (2008) Population trends of 48 common terrestrial bird species in Europe: results from the Pan-European Common Bird Monitoring Scheme. *Rev Catalana Ornithol* 24:4–14
- Ward DH, Dau CP, Tibbitts TL, Sedinger JS, Anderson BA, Hines JE (2009) Change in abundance of Pacific brant wintering in Alaska: evidence of a climate warming effect? *Arctic* 62(3):301–311
- Wright LJ, Newson SE, Noble DG (2013) The value of a random sampling design for annual monitoring of national populations of larger British terrestrial mammals. *Eur J Wildl Res* 60(2):213–221. doi:10.1007/s10344-013-0768-x
- Zuur AF, Saveliev AA, Ieno EN (2012) *Zero inflated models and generalized linear mixed models with R*. Highland Statistics, Newburgh