Noname manuscript No.
(will be inserted by the editor)

Working with population totals in the presence of missing data: Electronic Supplementary Material

Comparing imputation methods in terms of accuracy and precision

Thierry Onkelinx · **Koen Devos** · **Paul Quataert**

the date of receipt and acceptance should be inserted later

A Simulate missing not at random (MNAR)

The simulations are based on the observed pattern of missingness in the long-term water bird monitoring of Flanders, Belgium (WMF). We assume that this pattern is missing not at random (MNAR). The WMF dataset consists of 1225 sites, 24 winters and 6 months. About 44% of the observations is missing.

Research Institute for Nature and Forest (INBO), Kliniekstraat 25, 1070 Anderlecht, Belgium, E-mail: thierry.onkelinx@inbo.be
Tel.: +32 2 525 02 00, Fax: + 32 2 525 03 00

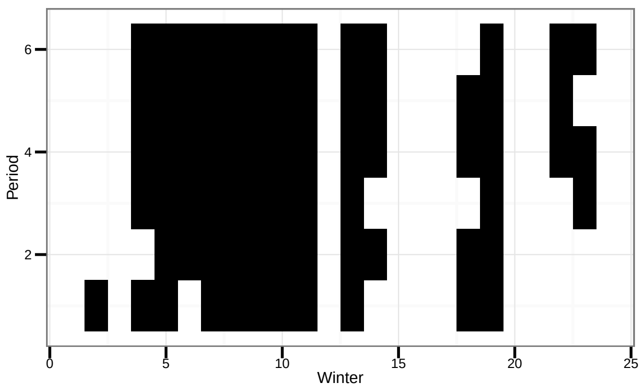

Fig. A.1 Pattern of observations in site 218. Dark squares indicate observations, voids are missing observations.

During the simulations we generate data with a different designs: different number of sites, different number of years and/or a different number of periods (months). The numbers might be smaller or larger than observed in the WMF. But still we want a MNAR pattern that is somewhat similar to the observed MNAR pattern.

We tackled this problem by using a simple random sampling with replacement of the WMF sites. Since we sample with replacement, the number of sites in the sample can be larger than observed in the WMF. Then we look at the observed time series (winter and month) for each selected site (e.g. Figure A.1). Within each selected site, we sample the required number of months from the time series of the site. Again, the sampling is done with replacement to allow for a larger number of periods should one require it (e.g. Figure A.2).

The resulting pattern is what we will use as the MNAR pattern in case of a time series of 24 winters. For shorter time series we will take the start of the time series. Since 24 winters of data is available in the WFC, we prefer the observed pattern of missingness over the ability to generate missingness for longer time series. This retains the observed decrease in missingness over time (Figure A.3).

Figure A.4 tries to illustrate the patterns in the missingness by sorting the sites in increasing order of time of the oldest observation, num-

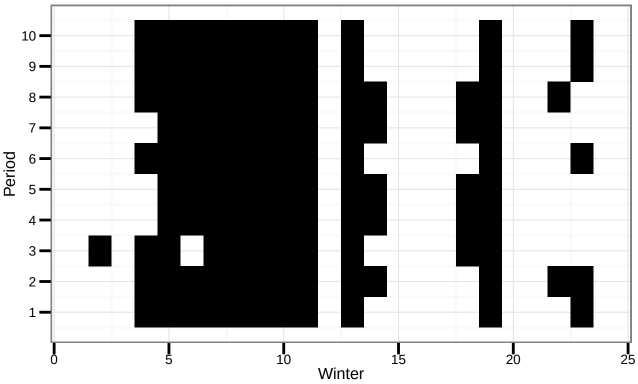

Fig. A.2 Pattern of observations in site 218 after resampling 10 periods with replacement. Dark squares indicate observations, voids are missing observations.

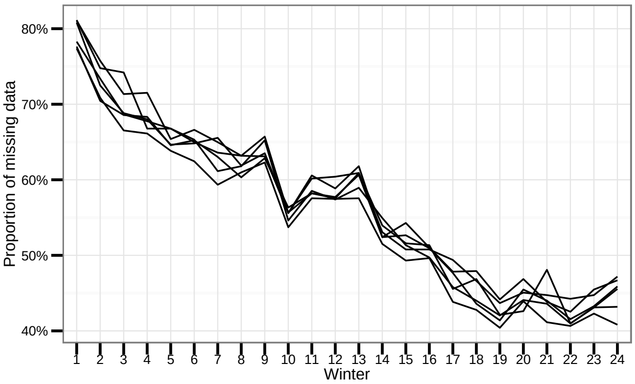

Fig. A.3 Evolution of the missingness over time in the WMF. Each line represents the proportion of missingness for a different month.

ber of winters with observations, the time of the most recent observation and the number of observations. Note that graph has many tiny dots. Each dot represents a site-year-month combination with observation. A site without missing data in a given winter will look like a horizontal line. The white space in between winters represent the summer months. The sorting places sites with a similar pattern of missingness close to each other, leading to vertical lines. The combination of both horizontal and vertical lines lead to areas. The blank area in the top left indicate sites were the observations have started recently. Blank horizontal lines in the bottom right area indicates sites were the monitoring has ceased early.

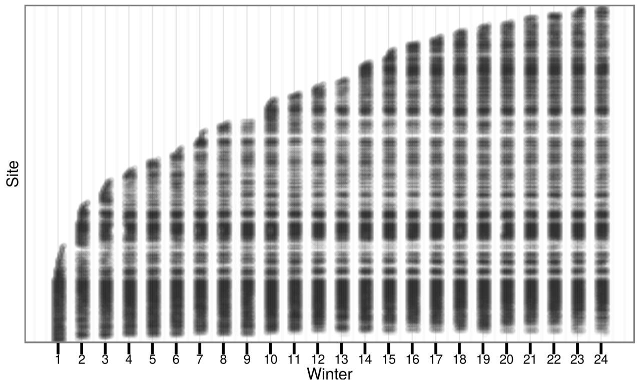

Fig. A.4 Evolution of the missingness over time in the WMF. Each tiny dot represents an observation of a site at that moment in time. Sites are ordered by the oldest observations, the number of winters with observations, the most recent observation and the number of observations.

B Compare different Underhill methods

We examine the Underhill method using the MNAR pattern. The bias of the Underhill method depends on the starting value (Figure B.1) and the algorithm. The original algorithm with mean as starting value clearly overestimates the winterly indices. Using zero as starting value yields a small downward bias. The altered algorithm is more robust against different starting values and yields less biased estimates.

The Underhill method replaces missing values with predictions of the model. This reduces the amount of variation in the augmented dataset. Hence the confidence intervals in the augmented dataset will be smaller than those from the complete dataset. Figure B.2 indicates that the Underhill method strongly underestimates the width of the confidence intervals. The size of the underestimation depends both on the starting value and the algorithm.

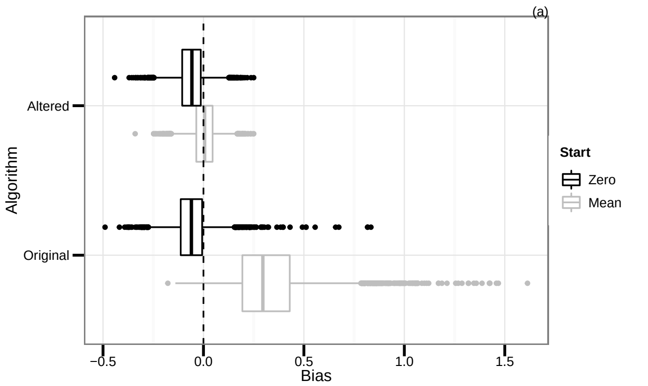

Fig. B.1 Boxplots of the bias of the different Underhill methods compared to the complete dataset.

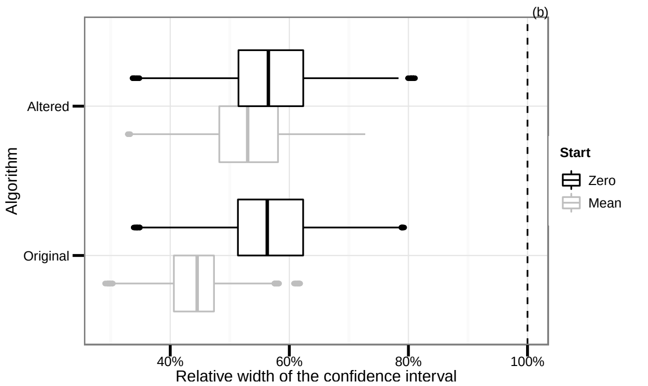

Fig. B.2 Boxplots of the relative width of the confidence interval of the different Underhill methods compared to the complete dataset.

C birdSTATs results with missing data compared with birdSTATs results on the complete dataset

We compare the birdSTATs indices on observed datasets with birdSTATs indices on the matching complete dataset. We find that the birdSTATs indices on observed datasets are unbiased in case of MCAR when compared to birdSTATs indices on complete datasets (Figure C.1). The MNAR pattern results in some bias. We see the expected pattern in *RWCI*: $RWCI \sim 1$ at 1% missingness and increasing *RWCI* as the degree of missingness increases (Figure C.2).

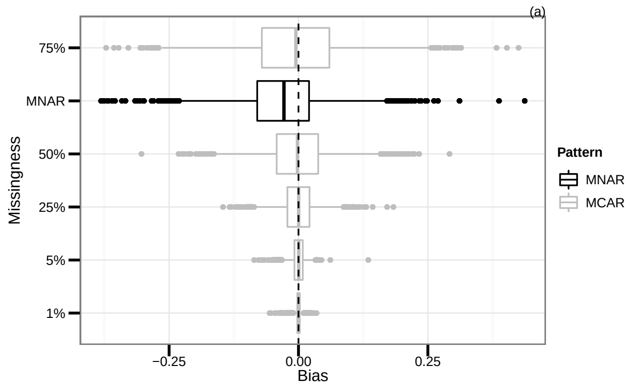

Fig. C.1 Boxplots of the bias of birdSTATs indices compared to the birdSTATS indices on the complete dataset.

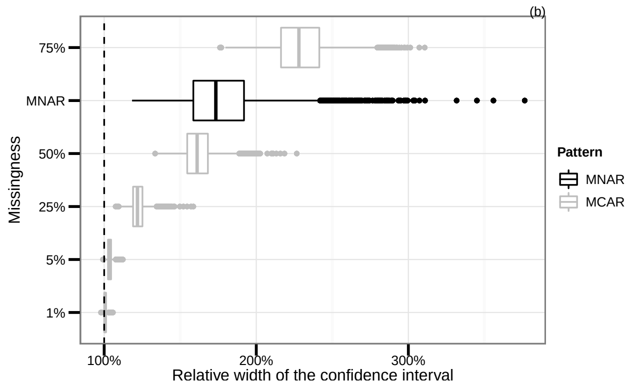

Fig. C.2 Boxplots of the relative width of the confidence interval of birdSTATs indices compared to the birdSTATS indices on the complete dataset.

D Influence of the number of imputation sets

The number of imputation sets has hardly any effect on the the bias (Figure D.1). The effect on the *RWCI* is slightly more pronounced. Figure D.2 shows that the *RWCI* decreases with increasing number of imputation sets. Note that the gain of adding extra imputation sets quickly diminishes with $L > 49$. $L = 19$ seems to be a reasonable compromise between keeping the number of imputation sets to a minimum while keeping the *RWCI* small.

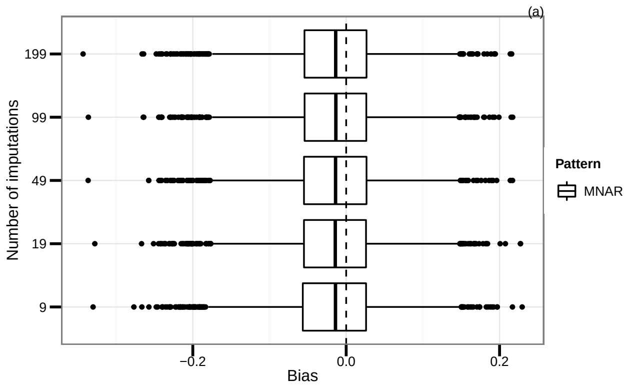

Fig. D.1 Boxplots of the bias of *API* after multiple imputation of missing data using INLA compared to *API* on the complete dataset. Results for the basic imputation model on 24 years of data for 100 sites with the MNAR pattern.

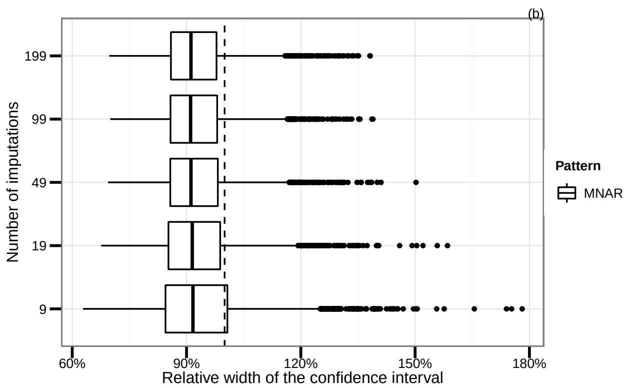

Fig. D.2 Boxplots of relative width of the confidence interval of *API* after multiple imputation of missing data using INLA compared to *API* on the complete dataset. Results for the basic imputation model on 24 years of data for 100 sites with the MNAR pattern.

References

Venables WN, Ripley BD (2002) Modern Applied Statistics with S, 4th edn. Springer