Journal: Ecology; **Article Type**: Statistical Innovations

Accounting for missing data in autoregressive models of ecological time series

Alice E. Stears[*, 1, 2], Melissa DeSiervo[*, 3, 2], Dustin Gannon[4, 2], Amy Patterson[5, 2], Alice M. Carter[6,7], Joanna R. Blaszczak[8], Matt Trentman[9], Eliza Grames[10], Robert O. Hall, Jr[7], Joshua P. Jahner[11, 2], Saheed O. Jimoh[2], Courtenay A. Ray[2], Christa L. Torrens[7], Lauren Shoemaker[◦2], Christopher Weiss-Lehman[◦2]

Author affiliations

[1] Center for Adaptable Western Landscapes, Northern Arizona University, Flagstaff, AZ

[2] Botany Department, University of Wyoming, Laramie, WY

[3] Biology Department, Union College, Schenectady, NY

[4] Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR

[5] Department of Biology, University of Maryland, College Park, MD

[6] Department of Mathematics and Statistics, Utah State University, Logan, UT

[7] Flathead Lake Biological Station, University of Montana, Polson, MT

[8] Department of Natural Resources and Environmental Sciences, University of Nevada, Reno, Reno, NV

[9] O'Connor Center for the Rocky Mountain West, University of Montana, Missoula, MT

[10] Biological Sciences, Binghamton University, State University of New York, Binghamton, NY

[11] Department of Biology, New Mexico Institute of Mining and Technology, Socorro, NM

[*] Denotes equal contribution as lead author

[◦] Denotes equal contribution as primary investigator

Corresponding author: Alice Stears, alice.e.stears@gmail.com

Open Research Statement: All R code to generate simulated data and conduct analyses are available in a GitHub repository (https://github.com/melissadesiervo1031/missing-data), and will be publicly archived upon acceptance of the manuscript.

Key words: Time series; missing data; autoregressive models; forecasting; simulations

Abstract

Long time series are valuable and increasingly available tools for understanding ecological systems. Unfortunately, sustaining data collection efforts over time is challenging, and gaps in time series data can complicate data analysis. While several methods exist to account for missing data, we know of no established best practice for handling missing data in time series models. Using simulated and empirical time series of primary productivity in a river and great tit population size, we compared the performance of six missing data approaches across scenarios with varying amounts and types of missing data. In each scenario, we fit statistical models across all approaches and assessed their ability to recover simulation parameters and forecast future dynamics. When data were missing completely at random, our results indicate that researchers have multiple methods available to them. Parameters were recovered well even with as high as 50% missingness. Conversely, no method can adequately account for data missing not at random (e.g., temperature sensors fail more often on hot days) by default. Our study emphasizes that identifying the mechanism of missingness is the most critical component of handling missing data, at which point there may be multiple suitable analysis options one could implement.

Introduction

Long-term time series contribute to our understanding of many ecological phenomena, from the impacts of species diversity on predator-prey dynamics to patterns of nutrient cycling in ecosystems (Hughes et al., 2017; Likens et al., 1970; Sinclair et al., 2003). There has been a concerted effort in recent years to facilitate the collection of long-term ecological datasets (e.g., the U.S. National Science Foundation Long Term Ecological Research Program), as well as to re-analyze historical long-term datasets with modern methods (Adler et al., 2009; Buma et al., 2017). However, even the most rigorously maintained long-term datasets are likely to have missing observations, due in part to unpredictable barriers to data collection that are difficult if not impossible to overcome (Łopucki et al., 2022; Nakagawa & Freckleton, 2008). Challenges include faulty sensors (Hossie et al., 2021), inaccessibility of field sites due to safety concerns or travel restrictions, funding lapses, and human error during data entry or collection. It is impossible to retroactively collect observations to fill gaps in time series, instead requiring statistical methods that account for missingness. Doing so is critical, as missing data can have cascading negative effects on subsequent analyses including reduced statistical power (Kang, 2013; Moritz & Bartz-Beielstein, 2017) and possibly biased estimation of parameters, leading to both inaccurate and imprecise conclusions (Aleryani et al., 2018; Junger & Ponce de Leon, 2015; Kim et al., 2018).

Challenges in accounting for missing values are compounded when dealing with time series data. For example, ecological time series are often autoregressive, meaning that the value of a data point depends in part on the values of previous observations. Thus, a single missing observation could lead to the effective deletion of multiple data points in downstream analyses. Further, many ecological time series consist of discrete data such as Poisson-distributed counts or binomial presence-absence data, which precludes the use of primary time series approaches (e.g. ARIMA models). As such, approaches that directly impute missing values or are model-based must be adjusted to account for time series that do not conform to classical models that assume Gaussian error distributions.

An additional challenge arises via the underlying mechanism driving missingness in time series, i.e. the statistical relationship between observations and the probability that a given observation goes missing. Missing data can be Missing Completely at Random (MCAR), Missing at Random (MAR), or Missing Not at Random (MNAR) (Nakagawa, 2015; Rubin, 1976, Appendix S1: Fig. S1). Data are MCAR when the probability of a missing observation is independent of both the numerical value of the observation itself, as well as any external information available (e.g., the value of a covariate) (Horton & Kleinman, 2007; Nakagawa, 2015; D. A. Newman, 2014). Data are MAR when the probability of missingness is not related to the numerical value of the missing observation, but the probability that an observation goes missing can be explained by one or more observed predictor variables (e.g., a buried soil temperature sensor malfunctions when soil moisture is high, causing wetter study sites to have more missing soil temperature observations) (Ellington et al., 2015; Nakagawa, 2015; D. A. Newman, 2014). Data are MNAR when the probability of missingness depends on the value of either the missing observation (e.g., a sensor cannot record values above a threshold) or an unmeasured predictor (e.g., high stream flow may be associated with high turbidity, but also leads to missing observations in both datasets because of sensor disruption—missing turbidity values are dependent on the missing stream flow values). Data that are MAR can often be treated like data that are MCAR after conditioning on the known variables driving missingness (Nakagawa, 2015; Nakagawa & Freckleton, 2011). MCAR and MNAR represent the two extremes of missing data in terms of the statistical complications they pose, and as such our analyses focus on these two missingness types.

Multiple methods exist for dealing with missing data (Fig. 1). Broadly speaking, these can be categorized into data processing approaches that occur prior to statistical analysis or model-based approaches that directly incorporate a model for missingness into the analysis. Data processing approaches are often simple and intuitive, including dropping all missing time points or imputing the data to fill them in before analysis. Even with more advanced imputation techniques (Kang, 2013; Nakagawa, 2015; Nakagawa & Freckleton, 2011; Onkelinx et al., 2017; Rubin, 1988, 1996) the general workflow is still straightforward: repair the missingness in the data, then analyze it as you would a complete dataset. By contrast, model-based approaches involve specifying a model for missingness as part of your statistical analysis where missing values are treated as parameters to be estimated (Kalman, 1960; Kang, 2013; Kong et al., 1994; Li et al., 2019; Nadjafi & Gholami, 2022). Many approaches in both categories have been applied to ecological data (K. Newman et al., 2023; Soldaat et al., 2007), but which method to choose and how their relative performance depends on time series attributes remains unclear. This is particularly true when confronted with time series that have different amounts and types of missing data or that have discrete values or non-Gaussian error distributions.

Here, we evaluate the performance of six different approaches for dealing with missing data in autoregressive models of ecological time series. We compare approaches using both real and simulated datasets with both Gaussian and Poisson error distributions. We artificially introduce different mechanisms and amounts of missingness and then quantify the performance of different missing data approaches in terms of both parameter recovery (simulated datasets) and forecasting accuracy (empirical datasets). Though no single method emerged as the overall best approach, we provide a detailed discussion of the relative merits of each with recommendations for when and how to use them in different contexts. Thus, we hope our results will be a resource for ecologists and environmental scientists in search of robust, reproducible methods for analyzing time series with missing data.

Methods

We first discuss the motivating case studies that we used to simulate time series with either Gaussian-distributed or Poisson-distributed error and their two corresponding empirical examples. We next overview creating data sets with different mechanisms, amounts, and autocorrelation in missingness, and finally describe how we compare performance of the six approaches for accounting for missingness.

Real-valued time series: Sensors that collect daily or hourly readings of environmental data have become ubiquitous in environmental research, and are an ideal example of data typically modeled with a Gaussian error distribution. We thus refer to these time series as “real-valued” moving forward. Despite the prevalence of sensor data, these types of data are highly prone to missingness (Chen et al., 2013). We evaluated the impact of missingness using both simulated and empirical data that represent daily measures of environmental and response variables from a sensor. We simulated and analyzed such real-valued time series using a first-order auto-regressive (AR(1)) error model with explanatory covariates, such that:

$$Y_t = \mathbf{x}_t'\beta + \phi(Y_{t-1} - \mathbf{x}_{t-1}'\beta) + \varepsilon_t \tag{1a}$$

$$\varepsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{1b}$$

where the data (Y_t) depend on the modeled effects of covariates ($\mathbf{x'}_t$), their associated parameters (β), and on an autoregressive error term (ϕ) which determines the influence of the previous time point on the current value of the time series. After accounting for the autoregressive structure, residual error (ε_t) at each time point is normally distributed with a variance of σ^2. We used Eq. 1a to simulate 1000 datasets, each with two covariates and 365 observations, representing a year of daily sensor data.

We further applied Eq. 1a as our statistical model to an empirical dataset consisting of estimates of daily river gross primary productivity (GPP; units: g O_2 m^{-2} d^{-1}) across three years in the Au Sable river in Michigan, USA (Appling et al., 2018). We note that while this times series is missing GPP estimates for 50 days, it still provides 1046 non-missing values, and indeed we were unable to find long term empirical GPP time series without any missing values. In fitting Eq. 1a to these data, we used estimates of incoming light (μmol m^{-2} s^{-1}) and flow (m^3 s^{-1}) as covariates, since they have previously been identified as primary drivers of GPP (Bernhardt et al., 2022).

Time series of counts: Time series with integer-valued error distributions (subsequently referred to as "time series of counts"), such as annual censuses of population size, are also very common in ecology. Approaches for dealing with missing data in these types of time series are not as well developed as they are for data with continuous values that can be modeled using Gaussian error structure. For these reasons, we additionally evaluated the impact of missing data on estimator performance using both simulated and empirical data that represent annual counts of individuals in a population. To simulate and analyze a time series of counts, we used a stochastic Ricker population model (Ricker, 1954) such that:

$$\eta_{t+1} = N_t e^{(r-\alpha N_t)} \tag{2a}$$

$$N_{t+1} \sim f(\eta_{t+1}) \tag{2b}$$

where η_{t+1} represents the expected population size at time $t+1$, r is the intrinsic, density-independent growth rate of the population, and $\alpha > 0$ is the intraspecific competitive effect that induces negative density dependence. Realized population size, N_{t+1}, is a random draw from the distribution $f()$ with mean η_{t+1}. Throughout our simulations and subsequent analyses, we set $f()$ to be a Poisson distribution with rate parameter η_{t+1}.We used Eq. 2a to simulate 1000 datasets of 60 observations each, representing 60 years of annual census data.

Paralleling our above approach, we also fit Eq. 2a to an empirical dataset consisting of a 59-year sequence of annual counts of great tit (*Parus major*) broods in the Wytham Woods in Oxford, UK (Ben Sheldon, personal communication, https://wythamtits.com). These census data are continuous from 1960 - 2018 with no missing values.

Introducing missingness into time series: To assess how missing data approaches perform across differing amounts and types of missing data, we systematically removed observations from both the empirical and simulated time series to create "missing datasets" (Appendix S1: Fig. S1).

Datasets differed in both the amount of missingness and mechanism of missingness, where we examined data missing completely at random (MCAR) and missing not at random (MNAR), as described below. These scenarios are meant to represent processes that could commonly occur in ecological data collection but are not intended to be exhaustive.

For all simulated and empirical time series, we created MCAR datasets with varying proportions of missing data and degrees of autocorrelation in missingness (Appendix S1: Fig. S1 B–E) by viewing a time series as a Markov-modulated Bernoulli process where the response variable could have two states: missing or not missing (Edwards, 1960; Gharib et al., 2014). We used a transition matrix to stochastically introduce increasing levels of missingness and different degrees of autocorrelation in missingness (where missing values were temporally clumped or evenly distributed across the time series) into our time series (see **Appendix S1: Introducing Missingness**). For each simulated time series, we created 150 MCAR datasets by combining 15 levels of missing data (from 5-75% by increments of 5%) with 10 levels of autocorrelation (from 0 to ~ 0.9 by increments of ~ 0.1). For both empirical time series, we created 450 MCAR datasets, representing 50 instances drawn from each combination of three levels of missingness (20±5%, 40±5%, and 60±5%) with three levels of autocorrelation (0.25 ± 0.05, 0.5 ± 0.05, and 0.75 ± 0.05).

We created real-valued MNAR datasets by removing observations at both the high and low tails of the distribution of data (Appendix S1: Fig. S1 F,G). For each of the simulated real-valued time series we created 15 MNAR datasets with missingness increasing from 5 to 75% by increments of 5%. We also created 15 total MNAR datasets from the empirical GPP time series, resulting in five MNAR datasets within each of three binned levels of of missingness: 20±5%, 40±5% and 60±5% (see **Appendix S1: Introducing Missingness**). We did not create time series of counts with MNAR data because that has no real-world parallel: in population count data, a population's size is unlikely to affect the probability that a sampling event occurs that year. It is important to note that while very low population size may lead to false zeros, observation error is a problem that is distinct from missing data.

Comparing missing data approaches: We evaluated several, previously published approaches for accounting for missing data: simple and complete data deletion (Nakagawa & Freckleton, 2011), multiple imputation (MI) (Rubin, 1988), the Kalman filter (KF) (Kalman, 1960), the expectation maximization (EM) algorithm (Kang, 2013), and Bayesian data augmentation (DA) (Kong et al., 1994) (Fig. 1). Briefly, both types of data deletion and MI are data processing approaches, meaning we applied them *before* fitting the corresponding statistical models. In simple data deletion, only the missing values are removed, and the resulting time series is compressed, violating the assumption of equal spacing between observations. In contrast, complete data deletion removes the missing value and any subsequent observation(s) it would predict from the response variable (note that subsequent observations are retained as predictors of future observations). MI systematically fills in missing observations with imputed values, creating multiple versions of the dataset to be fit and averaged over. The KF, EM algorithm, and Bayesian DA are all model-based approaches and thus are implemented *simultaneously* with fitting the relevant statistical model. The KF derives the likelihood of a time series with missing observations by following a two-step procedure: forecasting a future state given an initial state, then using the next time point to update the forecast using Bayes' theorem. The KF assumes a Gaussian error distribution, so we only used this method with our real-valued time series. The EM algorithm is conceptually similar to the KF but can be used for count data. As such, we only employed the EM algorithm with our time series of counts to replace the KF. Finally, Bayesian DA uses a model-based framework to simultaneously estimate both the standard model parameters and the value of any missing data points which are treated as additional parameters.

See **Appendix S1: Missing data approaches** for more detailed descriptions of all these methods.

To assess the impacts of missing data approach and amount and type of missingness on parameter recovery, we applied each missing data approach to our MCAR and MNAR datasets from *simulated* time series. For our real-valued time series, we used the missing data approaches to fit Eq. 1a and assessed the recovery of the ϕ and β parameters. For our time series of counts, we fit Eq. 2a and assessed the recovery of the r and α parameters. For each parameter and model fit, we calculated three metrics: relative error (e_s), absolute relative error ($|e_s|$), and whether or not the estimated 95% Confidence Interval (CI) contained the true parameter value used for the simulation (i.e., coverage). Relative error was calculated as $e_s = \frac{\hat{\theta}_s - \theta_s}{\theta_s}$ where $\hat{\theta}_s$ is the estimated parameter value and θ_s is the true value for simulation s. Using relativized values facilitated comparison across different simulated datasets and among different parameters within the same model. We use both relative and absolute relative error as relative error measures bias in the parameter estimate (positive values indicate an overestimate and vice versa) while the absolute relative error facilitates comparisons between methods in the magnitude of error. We use coverage to assess how well methods accurately convey the uncertainty associated with parameter estimates across different missingness levels and mechanisms.

To aggregate these metrics across our simulations, we grouped results according to data type (real-valued or count), missing data approach, missingness type (MCAR or MNAR), proportion of missing data (rounded to the nearest 10%), and three levels of autocorrelation ($0.25 \pm 0.05\%$, 0.5 ± 0.05, and 0.75 ± 0.05). Within these groups, we calculated the median relative error, median absolute relative error, and coverage. We reported the median of the error distributions (rather than mean) to reduce the influence of rare, extremely large error values. If the models and missing data approaches worked perfectly, the median relative error would be 0 (indicating no bias), the absolute relative error would be low (indicating low magnitudes of error), and the coverage would be 0.95 (indicating the 95% CIs accurately capture the associated uncertainty in a given parameter estimate). Finally, we note that, since we relativized the estimates and because the two β parameters in Eq. 1a are of the same class (i.e. regression coefficients), we pooled their results.

Next, to quantify the impacts of missing data approach and amount and type of missingness on forecasting accuracy, we applied each missing data approach to our MCAR and MNAR datasets generated from the two *empirical* time series. To this end, we first split each empirical time series into training and testing sets such that models were fit to the training set and those fits were subsequently used to forecast the unobserved testing set. The Au Sable GPP time series (real-valued) used the first two years of daily GPP estimates (66% of the total time series) as the training set with the remaining year (33%) used for the testing set. In the great tit data (time series of counts), we used the first 49 years (83%) for the training set and the last 10 years of data (17%) as the testing set. Then for each model forecast we calculated the root mean squared error (RMSE) of the forecasted values compared to the true observations. As described above, we aggregated results from each time series according to the missing data approach used, type of missingness, proportion of missing data, and autocorrelation in missing data. Within each grouping, we then calculated the mean RMSE value.

Results

Parameter recovery in models of real-valued time series: For the real-valued MCAR data, all missing data approaches performed similarly well in recovering the covariate β parameters, with median error and median absolute error both remaining below 10% and coverage very near 95% across all levels of missing data. In contrast, estimates of the autoregressive parameter, ϕ, were more affected by missing data, and some missing data approaches performed better than others (Fig. 2). Notably, the DA and KF methods had the strongest overall performance, with low error and accurate coverage of ϕ with as much as 60% missing data. All other approaches lost both accuracy and coverage as the amount of missing data increased. The two data deletion methods, simple and complete-case, performed nearly identically in their ability to recover ϕ. Both these methods had increasing median error and median absolute error as the amount of missingness increased, while coverage dropped as low as $\sim 70\%$ (Fig. 2). MI had the worst ability to recover the ϕ parameter across every proportion of missing data. Coverage dropped sharply with more than 10% missing data, while median error became more negative (eventually underestimating ϕ by $\sim 50\%$) and median absolute error steadily increased. The combination of large magnitude errors and low coverage seen in the lowest-performing approaches demonstrates they could be especially misleading.

The degree of autocorrelation in MCAR data also affected parameter recovery, but only for the two data deletion methods and only for ϕ (Fig. 3; see Appendix S1: Fig. S2 for the effect of autocorrelation on β estimates). Specifically, higher levels of autocorrelation led to more accurate estimates of ϕ when either complete case or simple data deletion were used, while other approaches were consistently accurate. For this reason, we primarily present results from moderate autocorrelation MCAR datasets in the main text. However the corresponding results for low and high autocorrelation can be found in the supplement (Appendix S1: Fig. S2).

Unsurprisingly, misleading parameter estimates were an even greater concern when fitting models to MNAR data. In contrast to our results for MCAR data when data are MNAR, our analysis indicates these misleading estimates occur for both β and ϕ parameters across all missing data approaches, with increasing effects as missingness increases (Fig. 2). All missing data approaches underestimated ϕ by $\sim 75\%$ and overestimate β by $\sim 40\%$ at the highest proportion of missingness (60%). Coverage for β plummeted from $\sim 75\%$ with low missing data to below 50% with 10% missing data. Once 40% or more of data were missing, coverage of the β parameters approached 0%, indicating that almost none of the CIs included the true values for these parameters (Fig. 2).

Parameter recovery in models of count time series: When fitting the Eq. 2a population model to simulated time series of count data with observations that were MCAR, the simple data deletion approach did not recover the density-independent growth rate (r) or intraspecific competitive effect (α) as accurately as other approaches, particularly as the proportion of missing data increased (Fig. 4). Among the remaining approaches, EM consistently had low median error and median absolute error. Interestingly, complete data deletion also had low median error, similar to EM, but higher median absolute error than EM. However, this pattern was reversed for DA, which recovered parameters with relatively higher median error than EM, but lower median absolute error (i.e. the estimates were more stable, but slightly more biased). Finally, MI recovered parameters with higher median error and median absolute error, though was still much more accurate and precise when compared to simple data deletion. In addition to generating higher errors, both simple data deletion and MI had substantially reduced coverage of both parameters. Similar to the results for real-valued time series, autocorrelation in missing data led to more accurate parameter estimates for the simple data deletion method, but had little impact on other methods (Appendix S1: Fig. S3).

Forecasting with missing data: Examining the effect of missing data on forecasts of real-valued time series revealed several key patterns. First, and unsurprisingly, increasing the proportion of MCAR data resulted in higher forecast RMSE across multiple model runs (Fig. 5B). Second, when faced with MCAR data, all missing data approaches performed similarly in terms of forecast RMSE. This likely reflects the relative importance of environmental covariates versus autoregressive structure in the GPP time series as our simulations showed β estimates to be much less impacted by missing data compared to ϕ estimates (Fig. 2). Finally, when confronted with MNAR data, model forecasts had substantially increased RMSE, even with relatively low levels of missing data. In fact, with only 20% of data MNAR, model forecasts had higher RMSE across all methods than scenarios in which 60% of the data were MCAR.

As with the real-valued time series, forecasts for time series of counts with MCAR observations also had increasing RMSE with increasing proportions of missing data. There was also a dramatic rise in the width of the interquartile ranges of RMSE values with increasing proportions of missing data (Fig. 6). We expect this is due to instability of the estimates as the total number of observations decreases. Similar to our forecasts of real-valued time series, all missing data approaches performed broadly similarly when forecasting time series of counts with MCAR data.

Discussion

Missing data are ubiquitous in ecological studies and can be particularly problematic when fitting autoregressive time series models since missing values violate a key statistical assumption that observations are equally spaced in time. We evaluated parameter recovery and forecasting ability of six previously proposed approaches for dealing with missing data by fitting autoregressive models to simulated and empirical examples of two different types of time series common in ecology. Our results indicate that parameter estimation can be fairly robust to missing data so long as an appropriate method is used to account for the missing data and the data are missing completely at random (MCAR). Thus, missing values in time series do not necessarily have a catastrophic effect on model accuracy, precision, or forecasts. In fact, several methods, such as DA, the KF (for real-valued time series), and EM (for time series of counts), could recover simulation parameters relatively well, even when 50% of observations were MCAR (although we still suggest caution if fitting models to time series with this much missing data).

While several missing data approaches performed well for a range of data and missingness types, we identified other approaches that were uniformly unsuitable and that we advocate against using. For example, MI performed poorly, especially when applied to real-valued time series. This is particularly notable given that MI has been shown to work quite well in non-time series data (Graham, 2009) and is commonly recommended for ecology (Ellington et al., 2015; Nakagawa & Freckleton, 2008). Overall, though, the most consistently poor performing method was simple data deletion in which unequal spacing between observations is ignored, violating a key assumption of most time series models. Simple data deletion tended to cause high error and simultaneously low coverage, meaning it produced incorrect estimates, but with relatively narrow confidence intervals. This effect was most pronounced for time series of counts (Fig. 4) but still present for real-valued time series (Fig. 2). We hypothesize the adverse effects of simple data deletion may have been mitigated in our real-valued time series by including two additional covariates whereas our time series of counts contained no covariates. The poor performance of simple data deletion is alarming because it is easy to implement and can be commonly applied in ecological studies, potentially even unknowingly depending on how software handles missing observations. However, based on our results we echo previous advice that researchers should avoid simple data deletion (Łopucki et al., 2022; Nakagawa & Freckleton, 2011; Shoari & Dubé, 2018).

Before selecting a particular approach for dealing with missing data, we argue it is critical to first identify the mechanism leading to missing data. We found that when data were missing not at random (MNAR), none of the methods we tested provided robust parameter estimates or forecasts. Unfortunately, many ecological time series may contain data that are MNAR (Bowler et al., 2025), but this may be difficult or even impossible to verify (Nakagawa, 2015). However, if the processes behind MNAR data can be identified and modeled, the missing data can be treated as MCAR conditional on the missingness process, which then makes applying common missing data approaches more viable (Nakagawa, 2015; D. A. Newman, 2014). With MCAR data, our results suggest that within both real-valued time series and time series of counts, there are several methods that tend to perform consistently well. The best suited approach will depend on the type of time series being investigated and the goals of each particular analysis. For example, if researchers wish to estimate the value of missing data points in a time series, then DA would be the best choice. If a clear quantification of uncertainty in parameter estimates or forecasts is important, researchers should avoid EM which does not directly provide standard errors of the estimates (Figs. 1, 4).

Another key consideration when choosing among missing data approaches is the autoregressive nature of the time series and model. For example, if forecasting real-valued time series with covariates, the choice among missing data approaches may not matter as all were able to accurately recover the β parameters in simulations (Fig. 2), leading to similar accuracy in forecasts (Fig. 5). However, if the autoregressive model is itself of interest, our results suggest options for adequately dealing with missing data become more limited. Only two approaches had similarly strong recovery of the autoregressive parameter ϕ: KF and DA. In contrast, both data deletion methods struggled to accurately recover ϕ except at high levels of autocorrelation in missingness (Fig. 3). This is likely because high autocorrelation in missing values also preserved longer stretches of intact data, allowing easier estimation of the autoregressive structure in these instances. In contrast to our model for real-valued time series, our model for time series of count data contained no additional covariates and estimated current population abundance purely based on abundance in the previous time point (Eq. 2a). Thus, the forecasting performance of all missing data approaches with this model demonstrate the potential importance of autoregression, as RMSE increased much more dramatically with increasing proportions of missing data than compared to forecasts of real-valued time series (Figs. 5, 6). However, it should be noted that previous analyses of the great tit data suggest that more complex models than used here (e.g., incorporating varied functional relationships or spatial processes) greatly aid in adequately modeling dynamics in this system (Lebreton, 1990).

Our results provide important insights into the performance of a variety of methods for dealing with missing data in ecological time series, where our simulations can be viewed as benchmarks demonstrating both best and worst case scenarios, respectively. For example, when working with a specific dataset, several methods can be 'tuned' to achieve better or worse performance (e.g., MI might perform better if using more imputed datasets (Honaker & King, 2010), while the KF might perform worse when setting the Kalman gain below one to account for observation error (Kalman, 1960)). Further, we chose to explore the effect of missing data in the response variable and assumed no missingness in the covariates. However, ecological datasets often have missing data in the covariates as well as the response, which would complicate the modeling approach. There is much literature on the problem of missing covariates (albeit, not necessarily in the context of time series) that is beyond the scope of this manuscript (Little, 1992). Similarly, we chose an empirical real-valued time series with known, strong effects from environmental covariates (Bernhardt et al., 2022; Hall et al., 2015). In more exploratory scenarios or systems with weaker covariate effects, inaccurate estimates of autoregressive parameters might have larger negative effects on model forecasting ability. Finally, we only explored MCAR and MNAR data, representing the two opposite ends of a spectrum from best to worst case scenarios (D. A. Newman, 2014). In between these are scenarios with data missing at random (MAR), meaning the probability of missingness depends on one or more covariates or known processes (D. A. Newman, 2014). Properly accounting for such missing data requires knowing the underlying mechanism such that the likelihood of missingness can itself be modeled. This may be a tall order in many ecological systems (e.g., McCall et al., 2014; Shoari & Dubé, 2018; Sotto et al., 2011). Though we did not explore MAR data, we argue model performance in this case would fall in between the two extremes of MCAR and MNAR presented here and depend on how well the missingness mechanism is modeled.

Collectively, our comparison of missing data approaches provides a guide for dealing with missing data in ecological time series. Our results, especially for the real-valued data, suggest that parameter estimation and forecast accuracy can be robust and even relatively high with MCAR data – especially when using a suitable method for accounting for missingness. However, we acknowledge this is unlikely to hold for all data types or model structures. As a result, we suggest that researchers confronting missing data in time series with other characteristics not considered here (e.g., different model or error structures, different lengths, etc.) run similar simulations to those we performed to decide how to best deal with missing observations. We argue this step should become part of the analysis workflow, particularly in a field-based discipline such as ecology where missing observations are often unavoidable.

Acknowledgments

We thank Heili Lowman and the Modelscapes consortium for their contributions to the manuscript. This manuscript was supported by funding from the National Science Foundation (award # OIA-2019528).

Conflict of Interest Statement

The authors declare no conflicts of interest.

References

Adler, P. B., HilleRisLambers, J., & Levine, J. M. (2009). Weak effect of climate variability on coexistence in a sagebrush steppe community. *Ecology*, *90*(12), 3303–3312. https://doi.org/10.1890/08-2241.1

Aleryani, A., Wang, W., & De La Iglesia, B. (2018). Dealing with missing data and uncertainty in the context of data mining. *Hybrid Artificial Intelligent Systems: 13th International Conference, HAIS 2018, Oviedo, Spain, June 20-22, 2018, Proceedings 13*, 289–301. https://doi.org/10.1007/978-3-319-92639-1_24

Appling, A. P., Read, J. S., Winslow, L. A., Arroita, M., Bernhardt, E. S., Griffiths, N. A., Hall, R. O., Harvey, J. W., Heffernan, J. B., Stanley, E. H., Stets, E. G., & Yackulic, C. B. (2018). The metabolic regimes of 356 rivers in the United States. *Scientific Data*, *5*(1), 180292. https://doi.org/10.1038/sdata.2018.292

Bernhardt, E. S., Savoy, P., Vlah, M. J., Appling, A. P., Koenig, L. E., Hall Jr, R. O., Arroita, M., Blaszczak, J. R., Carter, A. M., Cohen, M., et al. (2022). Light and flow regimes regulate the metabolism of rivers. *Proceedings of the National Academy of Sciences*, *119*(8), e2121976119. https://doi.org/10.1073/pnas.2121976119j

Bowler, D. E., Boyd, R. J., Callaghan, C. T., Robinson, R. A., Isaac, N. J., & Pocock, M. J. (2025). Treating gaps and biases in biodiversity data as a missing data problem. *Biological Reviews*, *100*(1), 50–67. https://doi.org/10.1111/brv.13127

Buma, B., Bisbing, S., Krapek, J., & Wright, G. (2017). A foundation of ecology rediscovered: 100 years of succession on the William S. Cooper plots in Glacier Bay, Alaska. *Ecology*, *98*(6), 1513–1523. https://doi.org/10.1002/ecy.1848

Chen, C.-P., Chuang, C.-L., & Jiang, J.-A. (2013). Ecological monitoring using wireless sensor networks-—overview, challenges, and opportunities. *Advancement in Sensing Technology: New Developments and Practical Applications*, 1–21. https://doi.org/10.1007/978-3-642-32180-1

Edwards, A. W. F. (1960). The meaning of binomial distribution. *Nature*, *186*. https://doi.org/10.1038/1861074a0

Ellington, E. H., Bastille-Rousseau, G., Austin, C., Landolt, K. N., Pond, B. A., Rees, E. E., Robar, N., & Murray, D. L. (2015). Using multiple imputation to estimate missing data in meta-regression. *Methods in Ecology and Evolution*, *6*(2), 153–163. https://doi.org/10.1111/2041-210X.12322

Gharib, M., Ramadan, M. M., & Al-Ajmi, K. A. (2014). Characterization of markov-bernoulli geometric distribution related to random sums. *Journal of Mathematics and Statistics*, *10*, 186–191. https://doi.org/10.3844/jmssp.2014.186.191

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Hall, R. O., Yackulic, C. B., Kennedy, T. A., Yard, M. D., Rosi-Marshall, E. J., Voichick, N., & Behn, K. E. (2015). Turbidity, light, temperature, and hydropeaking control primary productivity in the Colorado River, Grand Canyon. *Limnology and Oceanography*, *60*(2), 512–526. https://doi.org/10.1002/lno.10031

Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, *54*(2), 561–581. https://doi.org/10.1111/j.1540-5907.2010.00447.x

Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*, *61*(1), 79–90. https://doi.org/10.1198/000313007X172556

Hossie, T. J., Gobin, J., & Murray, D. L. (2021). Confronting missing ecological data in the age of pandemic lockdown. *Frontiers in Ecology and Evolution*, *9*, 542. https://doi.org/10.3389/fevo.2021.669477

Hughes, B. B., Beas-Luna, R., Barner, A. K., Brewitt, K., Brumbaugh, D. R., Cerny-Chipman, E. B., Close, S. L., Coblentz, K. E., de Nesnera, K. L., Drobnitch, S. T., Figurski, J. D., Focht, B., Friedman, M., Freiwald, J., Heady, K. K., Heady, W. N., Hettinger, A., Johnson, A., Karr, K. A., … Carr, M. H. (2017). Long-term studies contribute disproportionately to ecology and policy. *BioScience*, *67*(3), 271–281. https://doi.org/10.1093/biosci/biw185

Junger, W. L., & Ponce de Leon, A. (2015). Imputation of missing data in time series for air pollutants. *Atmospheric Environment*, *102*, 96–104. https://doi.org/10.1016/j.atmosenv.2014.11.049

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35–45. https://doi.org/10.1115/1.3662552

Kang, H. (2013). The prevention and handling of the missing data. *Korean journal of anesthesiology*, *64*(5), 402–406. https://doi.org/10.4097/kjae.2013.64.5.402

Kim, S. W., Blomberg, S. P., & Pandolfi, J. M. (2018). Transcending data gaps: A framework to reduce inferential errors in ecological analyses. *Ecology Letters*, *21*(8), 1200–1210. https://doi.org/10.1111/ele.13089

Kong, A., Liu, J. S., & Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems [Publisher: Taylor & Francis]. *Journal of the American Statistical Association*, *89*(425), 278–288. https://doi.org/10.1080/01621459.1994.10476469

Lebreton, J.-D. (1990). Modelling density dependence, environmental variability, and demographic stochasticity from population counts: An example using wytham wood great tits. In *Population biology of passerine birds: An integrated approach* (pp. 89–102). Springer.

Li, C., Chng, K. R., Kwah, J. S., Av-Shalom, T. V., Tucker-Kellogg, L., & Nagarajan, N. (2019). An expectation-maximization algorithm enables accurate ecological modeling using longitudinal microbiome sequencing data. *Microbiome*, *7*(1), 1–14. https://doi.org/10.1186/s40168-019-0729-z

Likens, G. E., Bormann, F. H., Johnson, N. M., Fisher, D. W., & Pierce, R. S. (1970). Effects of forest cutting and herbicide treatment on nutrient budgets in the hubbard brook watershed-ecosystem. *Ecological Monographs*, *40*(1), 23–47. https://doi.org/10.2307/1942440

Little, R. J. A. (1992). Regression with missing x's: A review. *Journal of the American Statistical Association*, *87*(420), 1227–1237. https://doi.org/10.1080/01621459.1992.10476282

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. https://doi.org/10.1016/j.ecolmodel.2022.109964

McCall, M. N., McMurray, H. R., Land, H., & Almudevar, A. (2014). On non-detects in qpcr data. *Bioinformatics*, *30*(16), 2310–2316. https://doi.org/10.1093/bioinformatics/btu239

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, *9*(1), 207. https://doi.org/10.32614/RJ-2017-009

Nadjafi, M., & Gholami, P. (2022). Expectation-maximization algorithm to develop a normal distribution nhpp reliability growth model. *Engineering Failure Analysis*, *140*, 106575. https://doi.org/10.1016/j.engfailanal.2022.106575

Nakagawa, S. (2015). Missing data: Mechanisms, methods, and messages. In G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa (Eds.), *Ecological statistics: Contemporary theory and application* (pp. 81–105). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672547.003.0005

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

Nakagawa, S., & Freckleton, R. P. (2011). Model averaging, missing data and multiple imputation: A case study for behavioural ecology. *Behavioral Ecology and Sociobiology*, *65*(1), 103–116. https://doi.org/10.1007/s00265-010-1044-7

Newman, D. A. (2014). Missing data: Five practical guidelines [Publisher: SAGE Publications Inc]. *Organizational Research Methods*, *17*(4), 372–411. https://doi.org/10.1177/1094428114548590

Newman, K., King, R., Elvira, V., de Valpine, P., McCrea, R. S., & Morgan, B. J. T. (2023). State-space models for ecological time-series data: Practical model-fitting. *Methods in Ecology and Evolution*, *14*(1), 26–42. https://doi.org/10.1111/2041-210X.13833

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/s10336-016-1404-9

Ricker, W. E. (1954). Stock and recruitment. *Journal of the Fisheries Board of Canada*, *11*(5), 559–623. https://doi.org/10.1139/f54-039

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Rubin, D. B. (1988). An overview of multiple imputation. *Proceedings of the survey research methods section of the American statistical association*, *79*, 84.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association*, *91*(434), 473–489.

Shoari, N., & Dubé, J.-S. (2018). Toward improved analysis of concentration data: Embracing nondetects. *Environmental Toxicology and Chemistry*, *37*(3), 643–656. https://doi.org/10.1002/etc.4046

Sinclair, A., Mduma, S., & Brashares, J. (2003). Patterns of predation in a diverse predator–prey system. *Nature*, *425*, 288–290. https://doi.org/10.1038/nature01934

Soldaat, L., Visser, H., van Roomen, M., & van Strien, A. (2007). Smoothing and trend detection in waterbird monitoring data using structural time-series analysis and the kalman filter. *Journal of Ornithology*, *148*(Suppl 2), 351–357. https://doi.org/10.1007/s10336-007-0176-7

Sotto, C., Beunckens, C., Molenberghs, G., & Kenward, M. G. (2011). Mcmc-based estimation methods for continuous longitudinal data with non-random (non)-monotone missingness. *Computational Statistics Data Analysis*, *55*(1), 301–311. https://doi.org/10.1016/j.csda.2010.04.026

Figure captions

Fig 1: Conceptual figure showing the different approaches to handling missing data considered in this manuscript and the mechanisms for each approach.

Fig 2: Parameter recovery metrics for real-valued time series, showing the median error of parameter recovery, absolute median error of parameter recovery, and 95% coverage of parameter estimates, resulting from models fit using five missing data approaches (colors) across an increasing proportion of missing data. These models were fit to simulated, real-valued datasets with data missing completely at random (MCAR; left panel) with autocorrelation in missingness >0.3 and < 0.6, and data missing not at random (MNAR; right panel). The coverage panel shows the proportion of model runs where the 95% confidence interval of a parameter estimate includes the true simulation parameter (dotted line at 0.95). Each point in each panel shows the median value of error or coverage across all models fit to simulations that used the same missing data approach, had the same amount of missing data (within a 10% bin), approximately the same amount of autocorrelation in missingness, and the same type of missingness.

Fig 3: Median error of parameter recovery of ϕ depending on the proportion of missing data and autocorrelation in missingness for each of five missing data approaches, using simulated, real-valued datasets with data missing completely at random (MCAR). Cells in dark blue (closer to 0) show that the median of model estimates of ϕ were close to the actual simulation parameter. See Appendix S1: Figure S2 for a version of this figure expanded to include all parameter recovery metrics (error, absolute error, and coverage) for both ϕ and β parameters.

Fig 4: Parameter recovery metrics for time series of counts, showing the median error of parameter recovery, absolute median error of parameter recovery, and 95% coverage of parameter estimates, resulting from models fit using five missing data approaches (colors) across an increasing proportion of missing data. These models were fit to simulated time series of count data where observations were missing completely at random (MCAR) with autocorrelation in missingness >0.3 and < 0.6. The coverage panel shows the proportion of model runs where the

95% confidence interval of a parameter estimate includes the true simulation parameter (dotted line at 0.95). Each point in each panel shows the median value of error or coverage across all models fit to simulations that used the same missing data approach, had the same amount of missing data (within a 10% bin), approximately the same amount of autocorrelation in missingness, and the same type of missingness.

Fig 5: (A) Daily estimates of scaled gross primary productivity (GPP) from the Au Sable River from 2012 through 2015 (699 days). Red tick marks on the x-axis indicate days when measurements were missing. The gray box indicates the data (347 days) excluded from model fitting, which were then forecast using the resulting parameterized model. (B) Root mean square error (RMSE) of forecasts made using a model fit to the Au Sable GPP time series shown in panel A. Each point shows the mean RMSE across all forecasts with a given missing data approach (indicated by color), in a given level of proportion of missing data (0.2 ± 0.05, 0.4 ± 0.05, 0.6 ± 0.05), and with a given type of missingness (missing completely at random (MCAR) with autocorrelation of 0.5 ± 0.05 or missing not at random (MNAR)). Error bars indicate the inter-quartile range.

Fig 6: A) Annual counts of great tit (*Parus major*) broods in the Wytham Woods from 1960 – 2018. The gray box indicates the data (10 years) that were excluded from model fitting, which were then forecast using the resulting model. (B) Forecast root mean square error (RMSE) of the great tit dataset with an increasing proportion of missing data using 5 approaches that account for missingness. Each point shows the mean RMSE across all forecasts with a given missing data approach (indicated by color), in a given level of proportion of missing data (0.2 ± 0.05, 0.4 ± 0.05, 0.6 ± 0.05) with all bars shown for autocorrelation of 0.5 ± 0.05. Error bars indicate inter-quartile range

Figures

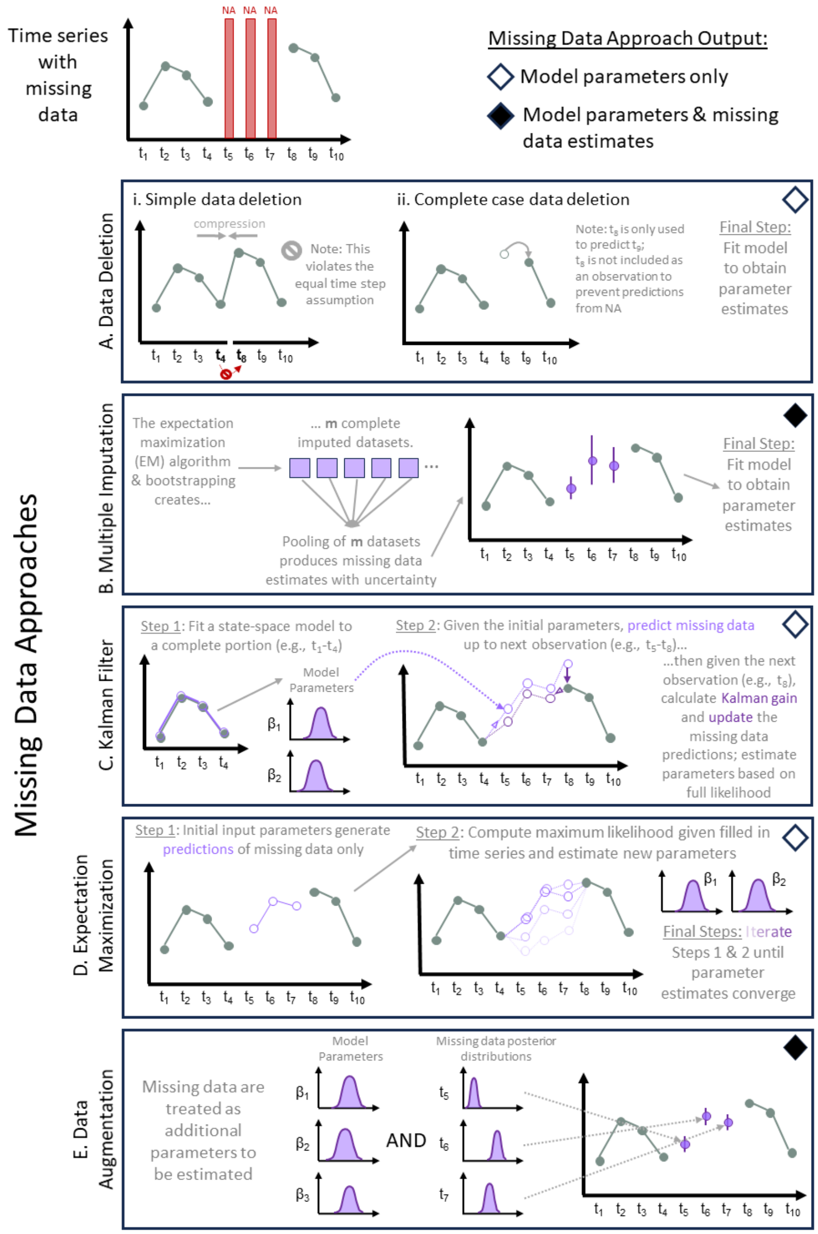
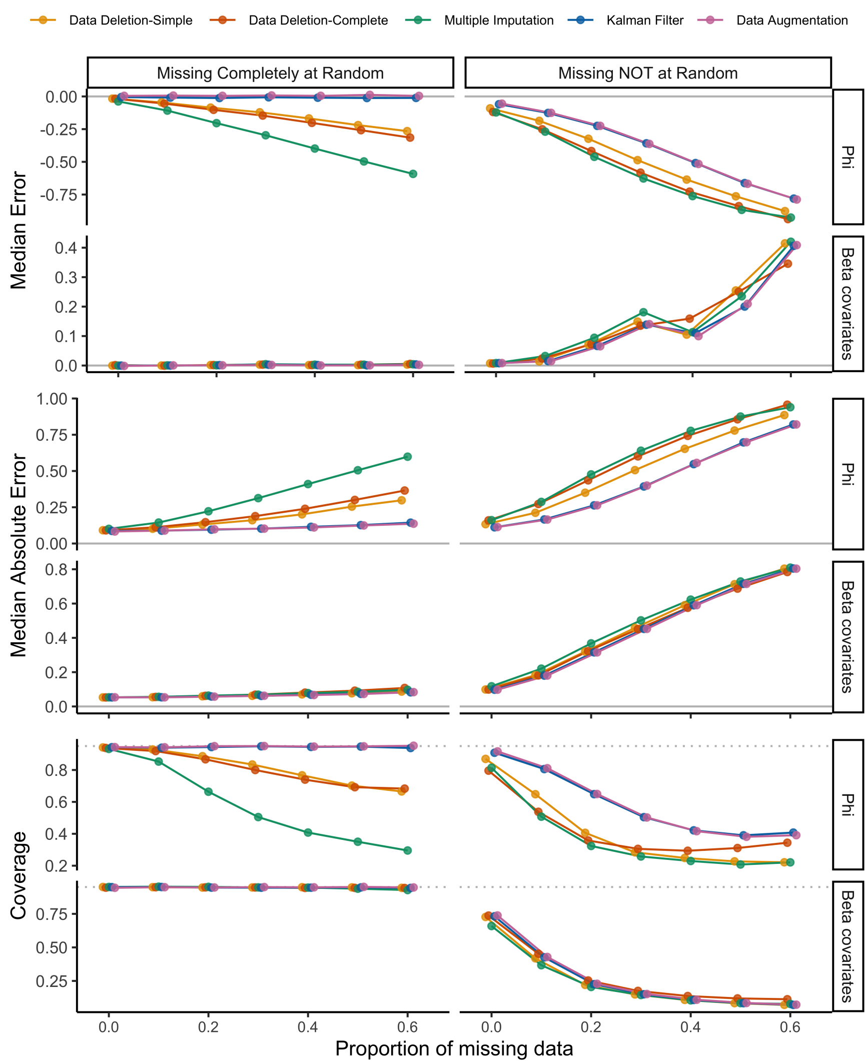


Figure 1:



Data Deletion-Simple
Data Deletion-Complete
Multiple Imputation
Kalman Filter
Data Augmentation
Missing Completely at Random
Missing NOT at Random
Median Error
Median Absolute Error
Coverage
Phi
Beta covariates
Proportion of missing data

Figure 2:

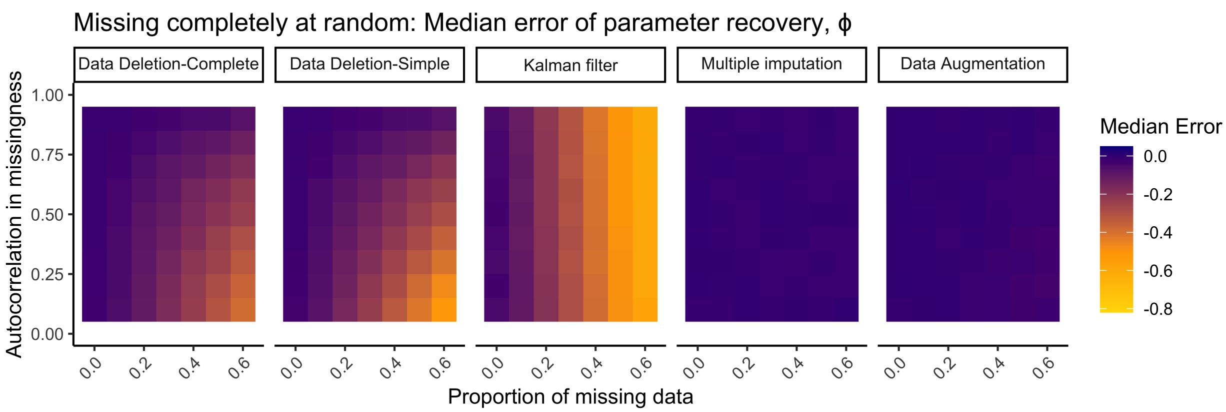
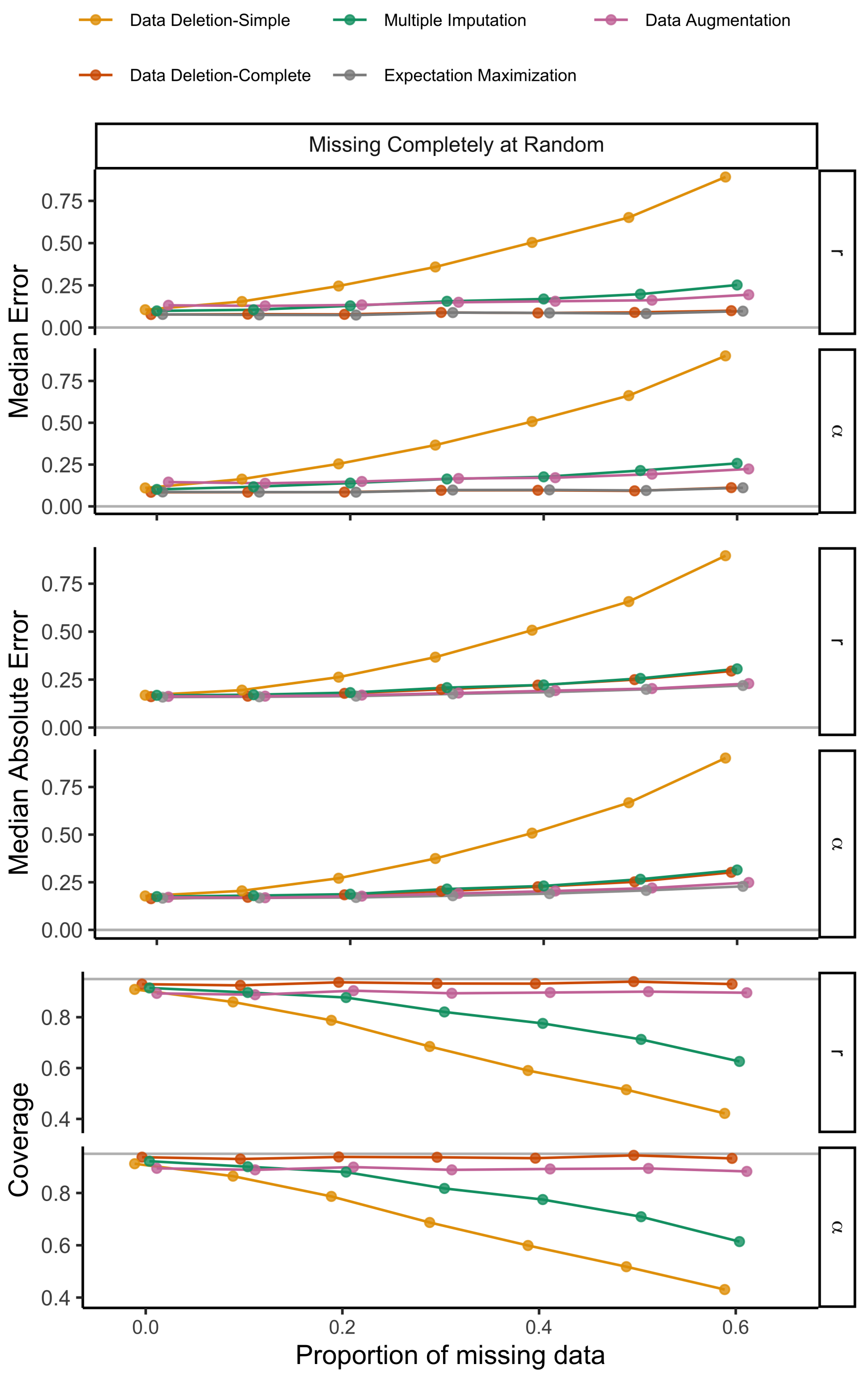

Missing completely at random: Median error of parameter recovery, ϕ
Data Deletion-Complete
Data Deletion-Simple
Kalman filter
Multiple imputation
Data Augmentation
Autocorrelation in missingness
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
Proportion of missing data
Median Error
0.0
-0.2
-0.4
-0.6
-0.8

Figure 3:



Data Deletion-Simple
Multiple Imputation
Data Augmentation
Data Deletion-Complete
Expectation Maximization
Missing Completely at Random
Median Error
Median Absolute Error
Coverage
r
α
0.00
0.25
0.50
0.75
0.4
0.6
0.8
0.0
0.2
0.4
0.6
Proportion of missing data

Figure 4:

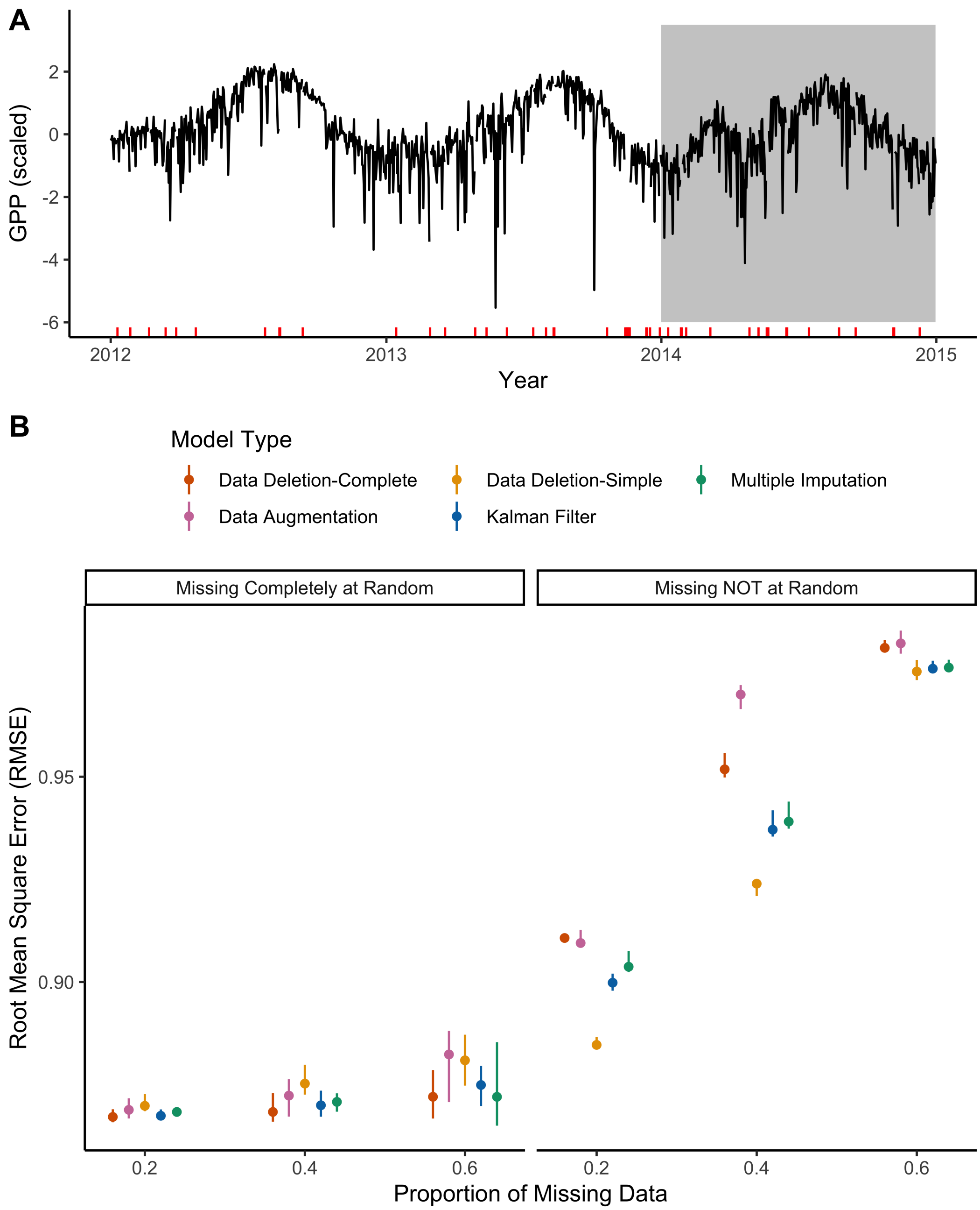

A
GPP (scaled)
2
0
-2
-4
-6
2012
2013
2014
2015
Year
B
Model Type
Data Deletion-Complete
Data Augmentation
Data Deletion-Simple
Kalman Filter
Multiple Imputation
Missing Completely at Random
Missing NOT at Random
Root Mean Square Error (RMSE)
0.95
0.90
0.2
0.4
0.6
Proportion of Missing Data

Figure 5:

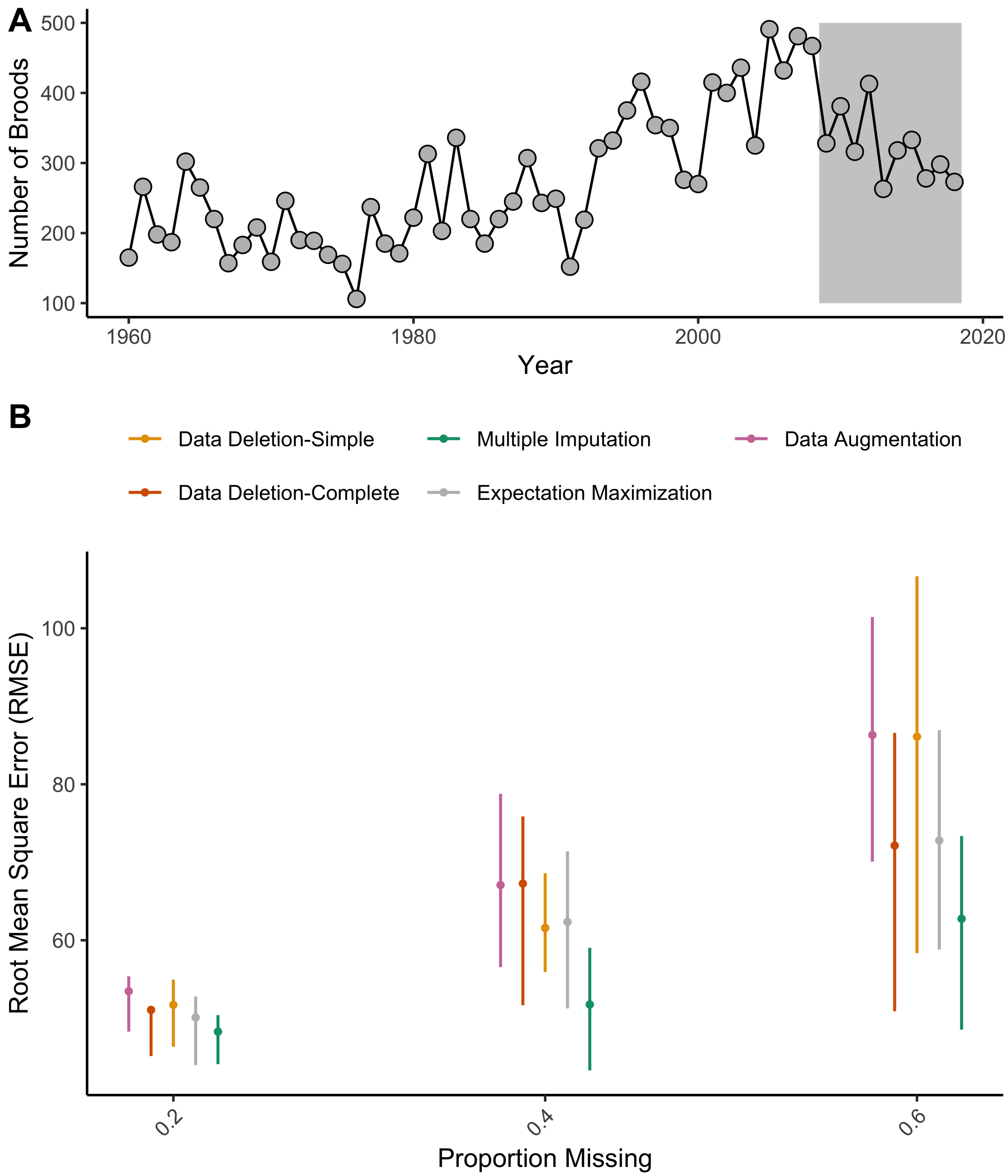

A
500
400
300
200
100
Number of Broods
1960
1980
2000
2020
Year
B
Data Deletion-Simple
Multiple Imputation
Data Augmentation
Data Deletion-Complete
Expectation Maximization
100
80
60
Root Mean Square Error (RMSE)
0.2
0.4
0.6
Proportion Missing

Figure 6: