Received: 20 March 2025 | Accepted: 24 June 2025

DOI: 10.1111/1365-2664.70110

Journal of Applied Ecology

RESEARCH ARTICLE

To fill or not to fill: Comparing imputation methods for improved riverine long-term biodiversity monitoring

Phillip J. Haubrock[1,2,3] | Ismael Soto[2] | Rafael L. Macêdo[4,5,6]

[1]Bournemouth University, Poole, UK; [2]Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, University of South Bohemia in České Budějovice, Vodňany, Czech Republic; [3]CAMB, Center for Applied Mathematics and Bioinformatics, Gulf University for Science and Technology, Mubarak Al-Abdullah, Kuwait; [4]Institute of Biology, Freie Universität Berlin, Berlin, Germany; [5]Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany and [6]Graduate Program in Ecology and Natural Resources, and Department of Ecology and Evolutionary Biology, Federal University of São Carlos, Sao Carlos, Brazil

Correspondence
Phillip J. Haubrock
Email: phillip.haubrock@hotmail.de

Handling Editor: Francesca Pilotto

Abstract

1. The preservation of global biodiversity has become challenging due to intensifying anthropogenic pressures. This study addresses the complex challenges associated with long-term monitoring data (i.e. missing years and gap filling) on the accuracy of temporal biodiversity trends.
2. Here, we analysed over 20 years of annual river macroinvertebrate data, simulating missing entries and applying imputation methods with linear and non-linear models to fill gaps. Our findings show that increasing numbers of gaps lead to increased trend variability, lower Akaike Information Criterion scores and higher standard deviation in model-explained deviance, thus suggesting that models fit more easily to datasets with more missing values due to fewer data constraints, while also introducing greater uncertainty and unexplained variability in the inferred trends.
3. When evaluating different gap-filling algorithms, we found that their performance varied considerably, contributing to increased uncertainty in trend estimates. Random Forest Imputations and Random Sample from Observed Values performed best, introducing less variation and aligning more closely with the original trends, whereas Predictive Mean Matching and its weighted variant amplified deviations, particularly with increasing gaps. Importantly, even a small number of missing or imputed values could, in some cases, reverse the trend direction, highlighting the risk of misinterpretation from seemingly minor data loss.
4. *Synthesis and applications*. In the current era of large-scale biodiversity monitoring, our study highlights the risks of missing data and the need for cautious imputation. We show that, in many cases, retaining gaps may lead to more accurate trend estimates than imputing data. When imputation is unavoidable, methods, such as Random Forest and Random Sampling from observed values performed relatively well in our macroinvertebrate richness case study. However, the choice

to impute as well as the method used should be evaluated in light of the biodiversity metric and the type of trend being analysed.

KEYWORDS
biodiversity monitoring, gap filling, biodiversity trends, mice, missing data, resilience, time-series analysis

1 | INTRODUCTION

In the 21st Century, environmental and biodiversity research are moving into the age of large databases and intense data-driven science, characterised by increasing data availability (Farley et al., 2018; Hampton et al., 2013; Spengler, 2000). The growing demand for biodiversity monitoring data is crucial for addressing environmental challenges, such as global warming, water quality deterioration, biological invasions and habitat degradation (Ellison, 2010; Kelling et al., 2009). Efforts are required to improve data handling and consequently minimise the pace of the global biodiversity decline and deteriorations of ecosystem services (Bellard et al., 2016; Choi et al., 2023). In recent years, understanding of the decline of global biodiversity caused by numerous different threats (Fernández-Palacios et al., 2021; Forister et al., 2011; Pyšek et al., 2020) has been compounded by the emergence of complex methodological approaches (Blowes et al., 2019) and new tools—such as deep learning—that aim to improve the interpretation of biodiversity trends and underlying drivers (Reichstein et al., 2019). In turn, enhanced interpretation can promote clear communication, which sparks not solely the interest among scientists to analyse large databases but also draws the attention of society and politicians towards the importance of large spatial scale and long-term biodiversity monitoring data (Seibold et al., 2019; Wei et al., 2017).

Our understanding of how long-term biodiversity monitoring data can inform relevant stakeholders remains incomplete, as evidenced by the contrasting trends and numerous nuances observed in various aquatic and terrestrial biodiversity monitoring datasets (Haase et al., 2023; Pilotto et al., 2020; Van Klink et al., 2020). While some large-scale studies have used linear or generalised linear models to assess biodiversity change (Blowes et al., 2019; Hallmann et al., 2017; MacGregor et al., 2019; Seibold et al., 2019), others increasingly employ non-linear approaches to account for the often non-linear and complex nature of biodiversity trends (Grünig et al., 2020; Hallmann et al., 2020). However, discrepancies in reported trends are also shaped by differences in data quality, metrics used and how missing data are handled—highlighting the need for careful methodological consideration in trend estimation, given the distinct properties of linear and non-linear models that can lead to different modelled outcomes (Haubrock et al., 2023). This is particularly relevant for ecologists, because separating long-term trends from natural fluctuations or cycles is often challenging (Benton et al., 2002), such as in univoltine and seasonally dynamic aquatic invertebrate taxa (Gaston & Lawton, 1988; Li et al., 2022; Picapedra et al., 2020; Wallner, 1987). For this reason, trend analyses, especially to investigate phenomena of critical importance to humanity, such as the pace of global biodiversity decline, the named 'insect apocalypse' (Goulson, 2019; Singh, 2002), should rely on true annually sampled time-series data covering a sufficiently long period (i.e. at least 15 consecutive years; Didham et al., 2020; White, 2019). Such annually sampled time-series are, however, scarce and often prone to contain several gaps in sampled data. These gaps may arise due to logistical constraints, inconsistent survey efforts, financial limitations, adverse environmental conditions (e.g. extreme weather events) or methodological changes over time. Understanding these causes is essential, as they influence whether missing data occur randomly or systematically, which in turn affects the reliability of inferred biodiversity trends. The problem of missing data affects almost all datasets in ecology (Nakagawa & Freckleton, 2008), ultimately hindering accurate inferences from long-term data and potentially leading to misinterpretations of biodiversity trends. While several studies have assessed the effect of missing data in biodiversity time series for specific taxa, such as birds and butterflies (e.g. Wauchope et al., 2019, 2021), the influence of data imputation on perceived biodiversity trends remains underexplored, particularly for complex, multi-species community datasets like those from long-term riverine macroinvertebrate monitoring, where taxonomic richness, sampling variability and temporal resolution pose distinct challenges.

Indeed, many, if not most, trends inferred from time series are likely influenced by missing data due to sampling or laboratory issues. While the presence of the impact of missing data is not always explicitly addressed, such effects may have played a role in studies on ecological time series (e.g. Choi et al., 2023; Dornelas et al., 2014; Haase et al., 2023; Muñiz Pinto et al., 2020), where long-term monitoring data are subject to common issues like inconsistent effort or data loss. Improper handling of such limited datasets could potentially introduce bias into statistical analyses, especially when systematic differences exist between the observed and missing data (Rubin, 1976; Little & Rubin, 2019). In the context of long-term biodiversity monitoring of riverine ecosystems, abrupt climatic shifts or human-induced habitat alterations can cause sudden changes in species occurrences and abundances within time series, which can appear as isolated outliers in the data (Pavón-Jordán et al., 2019). When such data are not consistently reported over time, for example on an annually recurring basis as in the case for numerous time series collected under the umbrella of the European Water

Framework Directive (Haase et al., 2004; Haubrock et al., 2020), such information on temporal variations may not be accounted for when data are missing, referring to instances where there are gaps between available data points (Little & Rubin, 2019). However, the analysis of 'complete' time series, that is with continuously annual sampled data, is often infeasible as they are scarce or with restricted geographical distribution.

Thus, some bias in the analysis and inferences from extrapolation of trends are assumed (Little & Rubin, 2019). As a result, numerous data analysts address missing data by imputing missing values (Luo, 2022; Pili et al., 2024; Zhang, 2016) and subsequently treating the filled values as 'true' observations (Gutwinski et al., 2021; Toutain et al., 2024), overlooking the potential introduction of biases or artefacts. While various approaches exist to handle missing data in ecological monitoring, including methods specifically developed for species count data, these often rely on specific assumptions, such as synchronous inter-annual fluctuations (see e.g. Kéry & Royle, 2016). One of the most popular methods among data scientists that is flexible, model-free alternative that can accommodate a wide range of missing data patterns and variable types and being continuously developed to fill data gaps are *Multiple Imputations by Chained Equations* (MICE; Jolani et al., 2015; Van Buuren & Groothuis-Oudshoorn, 2011), which is usually implemented by specifying generalised linear models for univariate conditional distributions (Raghunathan et al., 2001; Royston & White, 2011; Su et al., 2011). One of the advantages of MICE is the versatility in handling different types of missing data (e.g. numeric, categorical, binary and ordinal). This approach is appealing in large-scale data because it is simple and flexible in imputing different types of variables but has a key theoretical drawback that the specified conditional distributions may be incompatible, that is, they do not correspond to a joint distribution (a comprehensive probability distribution for all variables considered together; Arnold & Press, 1989; Gelman & Speed, 1993). This specifically means that when using MICE in large-scale data, there is a theoretical limitation where the specified conditional distributions used for imputation may not accurately reflect the joint distribution of the variables, potentially leading to biased imputations. Despite this and other drawbacks (Azur et al., 2011), MICE works remarkably well in real applications and numerous simulations have demonstrated that it outperforms many theoretical alternatives (see van Buuren, 2018). However, popular software packages for implementing MICE, for example in the `mice` package in R (Van Buuren & Groothuis-Oudshoorn, 2011), are CPU intensive, especially when using a high number of iterations (Akande et al., 2017; White et al., 2011).

Conclusions drawn from data imputed using mice may be less reliable for several reasons. First, studies using MICE often rely on a limited number of publicly available benchmark datasets (e.g. Jäger et al., 2021; Woźnica & Biecek, 2020; Section 5), which are typically well-structured and do not reflect the complexity of real-world ecological monitoring data. These complexities include irregular sampling, environmental variability and non-random patterns of missingness. Second, evaluations commonly use a narrow set of performance metrics, such as predictive mean squared error or overall accuracy (Harvey et al., 1997). While these metrics provide a general measure of imputation quality, they do not necessarily reflect the ecological relevance of reconstructed trends and may lead to misleading conclusions about the reliability of imputed data. As a consequence, the suitability of imputing data for biological monitoring data, especially in the context of long-term biodiversity time series (Pili et al., 2024), should be critically examined.

In this study, we thus focus on long-term riverine macroinvertebrate biomonitoring datasets to examine whether maintaining missing values or imputing them via MICE yields trends that better resemble the original, complete data. Riverine macroinvertebrates are widely used in ecological assessments due to their sensitivity to environmental changes and their key roles in freshwater ecosystem functioning (Touron-Poncet et al., 2014). However, species-rich macroinvertebrate time series pose unique analytical challenges stemming from the high taxonomic diversity within communities, temporal variability in community structure and methodological heterogeneity across sampling strategies. These challenges are further compounded by differences in life cycles, e.g. larvae and nymphs' stages) as well as the presence of dispersing and diapausing individuals which are often less responsive to environmental changes targeted by monitoring programmes and which can also affect detection probabilities and temporal turnover (Stubbington et al., 2018). Moreover, many species may be missing from most samples, leading to lots of zeros or the number of individuals found can vary a lot from sample to sample, making it difficult to use standard statistical methods. Also, the data are often grouped, that is taken from the same river or collected over time. Given the high variability and complexity of these communities, examining and choosing appropriate methods to identify and fill data gaps is essential, as poor handling of missing data can lead to biased conclusions about biodiversity change over time.

We assess the impact of gap filling by: (a) establishing alpha diversity (species richness per year) trends from complete time series, (b) introducing gradients of randomly missing years and (c) imputing these gaps using MICE and multiple modelling approaches. To test this, we collected three time series exceeding 20 years of continuous annual monitoring data of riverine macroinvertebrate species and evaluated (a) the original alpha diversity (i.e. local richness, as the number of species per year) trend obtained from the full time series, (b) shifts in obtained alpha diversity trends following an increasing gradient of randomly introduced annual gaps and finally (c) evaluated alpha diversity trends when previously introduced gaps were imputed using MICE and a series of linear and non-linear modelling approaches. We hypothesised that (i) with an increasing number of gaps or (ii) an increasing number of imputed gaps, obtained species richness trends will deviate substantially from the original richness over time towards both more positive or more negative trends, hence affecting consequent inferences. While our case study centres on riverine macroinvertebrates, the methodological framework we apply has broader relevance for ecological time series where imputation may influence trend interpretation.

2 | METHODS

2.1 | Data acquisition

Riverine macroinvertebrates were chosen as a test system due to the availability of high-quality, long-term, annually resolved time series across multiple sites. While macroinvertebrate communities have distinct characteristics, such as seasonal life cycles, high species richness and sampling variability, our analysis is not dependent on these traits. Instead, we use this group as a representative and data-rich example to evaluate how imputation influences trend detection in complex ecological time series. Thus, to investigate the effect of missing and, respectively, imputed gaps on time-series analyses, we sourced time-series data of sufficient length—defined here as at least 20 consecutive years of annual sampling, which has been recommended for detecting directional biodiversity trends (Didham et al., 2020; White, 2019)—from Haase et al. (2023) (but see Welti et al., 2024 for a detailed description of the data). This database contains 1826 time series of riverine macroinvertebrate communities sampled across 22 European countries. Time series were collected as part of 41 independent monitoring projects, where all samples in a time series were collected within a period of three consecutive months of the year. Each sample consists of abundance data for all macroinvertebrate species recorded at a specific site in a given year. For the purposes of this study, we derived annual species richness (i.e. number of species per year) as the primary metric for trend analysis. Each time series contains at least eight sampling years albeit not necessarily continuous (i.e. can contain gaps). Each time series is standardised to contain only one sampling event per year. Although different time series used different sampling methods and protocols, methodologies were kept consistent within each time series during the sampling period.

From this dataset, we extracted three distinct time series from single localities within Denmark, the Netherlands and Sweden. Each of these series provided annual species-level samples, with a minimum temporal coverage of 20 years (exceeding the 15 years proposed by Didham et al., 2020). The time series from the Netherlands spanned 26 years (1991–2016), comprising 218 unique entries at both species and genus levels. We excluded all genus-level information due to potential inaccuracies and overlaps, focusing solely on taxa (species-level only) identified at the species level for final inclusion. This resulted in 139 unique species-level entries. The Swedish time series comprised 23 years (1997–2019), yielding 114 unique species. Following the same procedure, only taxa identified at the species level were retained, resulting in a total of 70 unique species reported. In the case of Denmark, a time series spanning 27 years (1992–2018) was used, encompassing 195 genus- or species-level entries from which a total of 129 species-level entries were retained for analysis after omitting genus-level information.

2.2 | Data analysis

Prior to formal analysis, we plotted the annually observed species richness of each of the three time series using the `ggplot2` R package (Wickham, 2006) and visualised each time-series trend using a `ggplot2`'s 'loess' function with a span of 0.75 and degree of 2. Then, we introduced deletions of 1, 2, 5 or 10 years randomly within each of the three time series, a process repeated 100 times for each time series, resulting in the creation of 500 distinct time series (i.e. 100 for each gap quantity). The first and last years remained unaltered, ensuring that all time series retained their initial and final points. Subsequently, we computed the annual species richness for both the original time series and the altered time series with randomly deleted years (1, 2, 5 or 10). Species richness was chosen as a metric due to its widespread recognition as a fundamental indicator of biodiversity (Prendergast & Eversham, 1997), which both facilitates meaningful comparisons across studies and facilitates the analysis and interpretation of the ecosystem dynamics under investigation. Its broad acceptance within the scientific community facilitates meaningful comparisons across studies, providing a comprehensive lens through which to analyse and interpret the complex dynamics of the respective ecosystems under investigation (Lavorel et al., 2017).

To evaluate how different levels of missing data can affect the detection and interpretation of temporal trends in species richness, to both the original time series and each of the replicated time series with randomly deleted years (i.e. with 100 iterations for each gap quantity), we fit two different regression methods: (i) linear models using the *lme* function of the `lme4` R package (Bates et al., 2009) and (ii) generalised additive models using the *gam* function of the R package `mgcv` (Hastie & Hastie, 2015). Each model was constructed with the total annually observed species richness as the response variable and year as the sole predictor to isolate and assess the temporal trends in species richness over time. This approach allows us to focus directly on the relationship between time and species richness without introducing additional variables that could confound the analysis. By using year as the predictor, we can effectively capture any underlying trends or patterns in species richness that may be influenced by temporal factors. We chose linear and generalised additive models because linear models underlie the assumption of monotonicity, whereas generalised additive models can offer greater flexibility in capturing non-linear relationships within the data. We then computed the Akaike information criterion (AIC), the root mean square error (RMSE; a measure of the accuracy of a predictive model, quantifying the difference between predicted values and actual values) and explained deviance for each model and replicated time series with gaps to assess model fit. The AIC accurately, however, measures model fit only when computed across models with different specifications using the exact same dataset. When a model is fitted using a different number of data points (e.g. after the introduction of gaps or their respective filling), the comparability of the AIC is limited. However, in this study, we intentionally use the AIC as a diagnostic metric to illustrate how the introduction and filling of gaps affect the *perceived* model fit (Nakagawa & Freckleton, 2011). Our goal is to explore the sensitivity of AIC to these changes within the same datasets rather than to compare model quality directly across different datasets. By doing so, we aim to advance on the work of Li et al. (2024) to highlight the potential pitfalls of relying on AIC values when dealing with incomplete data by showcasing the effects of both missing and imputed data gaps while clearly acknowledging the limitations of this approach. Variance, representing a measure of a random variable, and deviance, indicating model fit, although they may appear similar, bear distinct conceptual differences. Variance quantifies the spread of data points around the mean, offering insights into the variability of a random variable, while deviance assesses the adequacy of a statistical model in explaining observed data—particularly in likelihood-based modelling—and serves as a key criterion for model selection and assessment (Lei & Bae, 2013; Variyath et al., 2010). Despite their superficial similarities, these three metrics serve fundamentally different roles in statistical analysis. This comparison between different imputation methods and their metrics provides insights into their impact on data integrity. Specifically, it highlights how these algorithms influence the number of gaps in the dataset and the associated statistical models used for analysis. By evaluating these metrics, we can better understand the reliability and performance of various imputation methods in maintaining data quality.

2.3 | Influence of gap filling on conventional modelling methods

We employed the *mice* function from the R package `mice` to fill the randomly introduced data gaps (1, 2, 5 or 10), using a method called multiple imputation (van Buuren et al., 2015; Van Buuren & Groothuis-Oudshoorn, 2011). This approach fills in missing values for the presence and abundance of macroinvertebrate species by using species-specific abundance data from before and after the gap. The *mice* algorithm is a flexible imputation framework that can be applied to time-series data by treating time (e.g. year) as a predictor. While it does not inherently model temporal autocorrelation, it estimates missing values based on observed patterns within the same dataset. The *mice* function affords the flexibility to specify various imputation parameters, including the imputation method, number of imputations and convergence criteria. In this study, we applied *mice* to fill missing years using species richness data from each respective time series, without borrowing information from spatially related sites, allowing us to assess how different imputation strategies influence trend estimation while maintaining the structure of the original time series. We therefore opted for the method *Predictive Mean Matching*, a widely used approach deemed suitable for numeric data (Kleinke, 2017; Toutain et al., 2024). *Predictive Mean Matching* imputes missing values by selecting variables from a set of $k \geq 1$ values known as the donor pool (i.e. values occurring within the time series). This donor pool consists of values that are closest to the predicted value for the missing case (Kleinke, 2017). It thus also considers relationships within the time series to generate multiple complete versions of the dataset.

Following gap imputation, we re-fitted linear and generalised additive models for each time series and subsequently extracted AIC values and explained deviance. In addition, we compared the performance of *Predictive Mean Matching* with three additional gap-filling methods, namely *Weighted Predictive Mean Matching*, *Random Forest Imputations* and *Random Sample from Observed Values*. *Predictive Mean Matching* is an advanced method for handling missing data by leveraging predictive modelling (van Buuren, 2018). It first predicts missing values based on observed data and then adjusts these predictions by matching them to the closest observed values in terms of mean. *Weighted Predictive Mean Matching* is a lesser-known approach which extends *Predictive Mean Matching* by introducing a weighting mechanism during the adjustment process, aiming to enhance accuracy by automatically assigning weights to observed values. *Random Forest Imputations* are implemented through the `MICE` R package, which builds multiple decision trees and combines their predictions into a single output (Shah et al., 2014). The predictors used by the Random Forest algorithm for imputation were the same species-specific abundance data used in the other imputation methods, along with temporal relationships within the time series. *Random Sample from Observed Values* involves imputing missing values by randomly selecting observed values from the dataset, providing a simpler approach compared with predictive modelling methods (van Buuren, 2018; Van Buuren, 2011). These approaches also offer similar abilities to impute any kind of missing variable (i.e. numerical, binary, ordered and unordered) within a series of data points, but have not been assessed for their suitability to fill missing data in long-term biodiversity monitoring data. For this detailed imputation comparison, the Swedish time series was selected due to its longer duration and higher taxonomic richness, making it ideal for comparative evaluation.

2.4 | Influence of gap filling on meta-regression modelling

While the previous analyses focused on how individual time series were affected by missing or imputed data, the meta-regression approach provides a broader synthesis across multiple altered time series. This allows us to systematically evaluate the general effect of missing data on trend estimation across different datasets rather than examining deviations on a case-by-case basis. Thus, to expand the methods used in our study, we also incorporated a monotonic Mann-Kendall trend test analysis to perform a meta-regression on the 100 randomised time series comprising 1, 2, 5 and 10 gaps, respectively, as well as their corresponding filled counterparts. With this analysis, we aimed to discern if the (meta-) trend's slope (response variable, i.e. S-statistic of the monotonic trend) of each altered time series (with gaps) diverged from the time series after gap filling. Meta-regressions are a statistical tool to synthesise and investigate trends across different time series that enable the examination of how various factors or characteristics relate to the overarching trends or patterns observed across diverse time-series datasets (Soto et al., 2023). Here, we used the number of missing years as the predictor variable, assessing its effect on the estimated slope of trends. Additionally, the variance associated with each estimated trend slope was incorporated to weight the meta-regression model, ensuring that estimates were adjusted for variability in trend calculations. Meta-regression models were applied using the *rma* function of `the metafor` R package (Viechtbauer, 2010).

3 | RESULTS

3.1 | Displaying raw data from source populations

The analysis of species richness of the three original macroinvertebrate time series, depicted using a loess function (i.e. the smoothed curve), revealed distinct patterns. In Denmark, species richness was markedly non-linear, as the trend exhibited a peak in 2005 (Figure 1a). Conversely, in the Netherlands, time series displayed an overall positive trend, characterised by an increase in species richness until 2000, followed by a period of almost linear stability (Figure 1b). Lastly, in Sweden, time series demonstrated a high variability, with a decrease until 2005 followed by a steady increase until 2011 and lastly a plateau phase of the community.

3.2 | Influence of gap filling on conventional modelling methods

Random deletion of annual data led to notable deviations from the original trend, with the extent of variation with the number of deleted years (Figures 2 and 3; left). As the number of deleted years increases, the extent of variation and randomness within the time series becomes more pronounced. Deleting 1 or 2 years has a minimal impact, allowing for the retention of general trends. However, with 5 or 10 years deleted, the data gaps become more significant, possibly diminishing the ability to capture nuanced patterns and weakening the overall representation of macroinvertebrate population dynamics over time. Subsequently, filling these gaps yielded further deviation, without conforming to the initial level of observed variability (Figures 2 and 3; right panel).

When using linear models to display trends over time, the deletion of 5 or 10 years resulted in two of three cases having a deviation from the (albeit non-significant) slightly positive trend into both more positive and more negative trends. In Denmark, the removal of data cases resulted in an inversion of the trend, which linear imputation methods exacerbated (Figure 2a–c). Upon filling in the missing data, a positive incline in trends was observed, with this effect becoming more prominent as the number of filled gaps increased. In the Netherlands, which exhibited an overall positive raw data trend, gap filling introduced substantial deviation, but served to amplify the positive slope of the trend. Conversely, in Sweden, where the raw data demonstrated a flat trend, the filling of gaps led to a decline in the deviation from original trends, intensifying the negative trajectory with missing data. Notably, when trends were analysed using generalised additive models, the estimated variability tended to be more pronounced than when analysed using linear models (Figure 3a). Pertaining to the ability of generalised additive models to display the non-linear nature of time-series data, trends (especially in Denmark and the Netherlands; Figure 3a,b) deviated substantially from the raw data trend. Filling data gaps resulted in a further increase in deviation from the raw data trend, scaled by the number of filled data gaps.

As the number of introduced data gaps increased, AIC values exhibited a declining trend, indicating that the model found a more straightforward or 'plausible' fit (Figure 4; Supporting

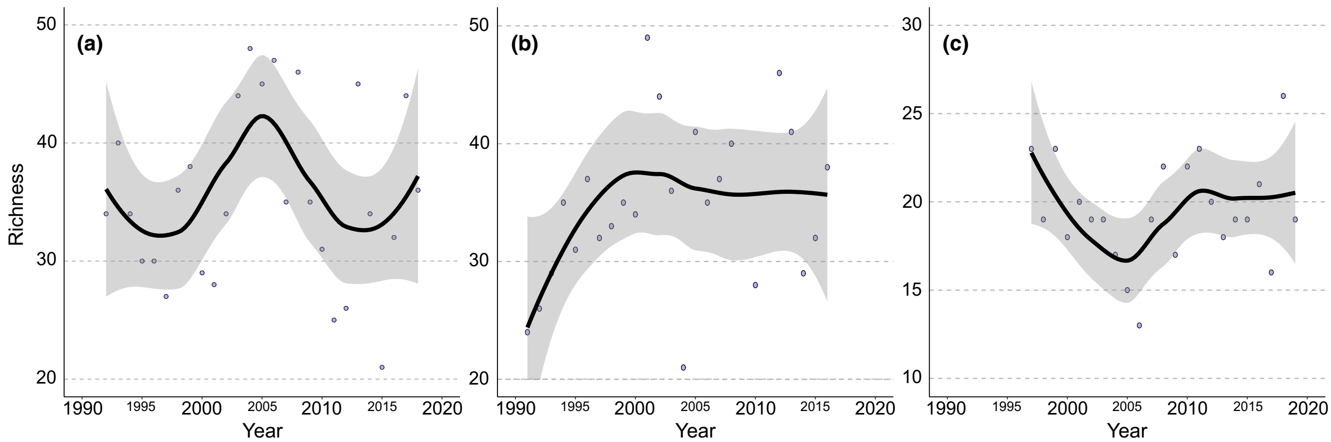

FIGURE 1 Original trends (black lines, i.e. raw data) of macroinvertebrate richness sampled in (a) Denmark, (b) the Netherlands and (c) Sweden. Differences in visualisation are due to smoothing functions applied in specific figures for trend clarity using a loess function implemented in the R package `ggplot`. The grey shaded area indicates the 95% confidence interval.

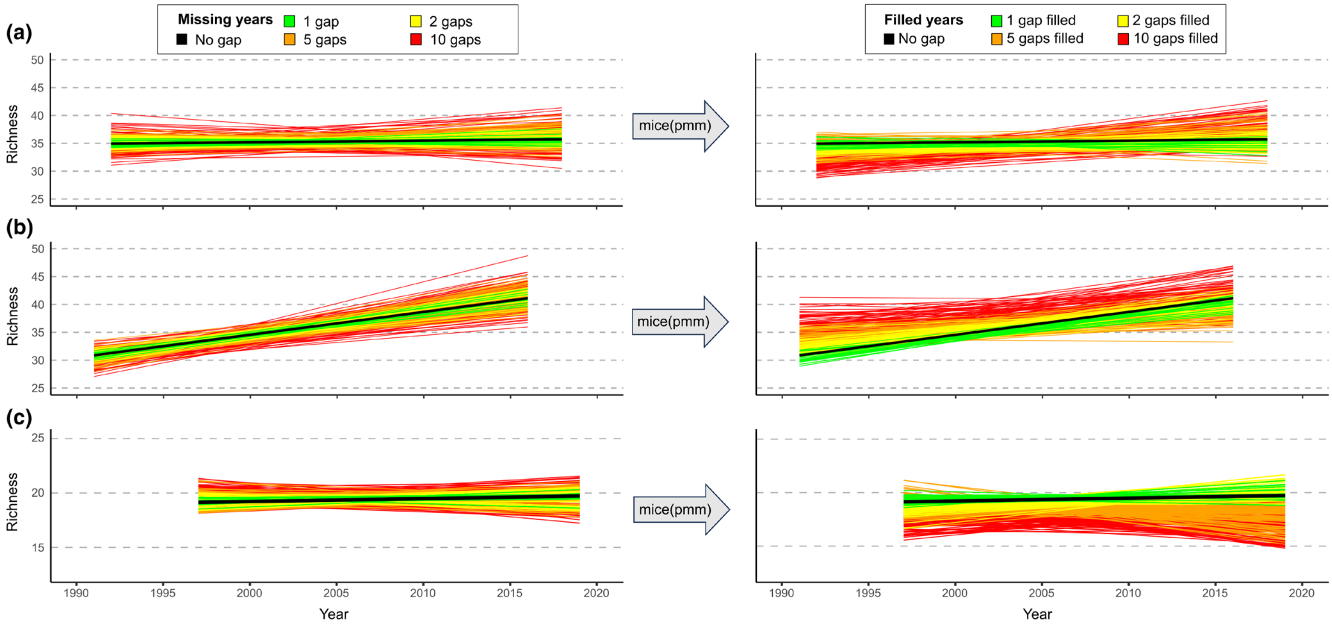

FIGURE 2 Variability in time series from (a) Denmark, (b) the Netherlands and (c) Sweden analysed using linear models with randomly introduced gaps (left; black: original trend; green: 1 gap; yellow: 2 gaps; orange: 5 gaps; red: 10 gaps) and gaps filled using the *mice* function of the `mice` R package. Each line represents one of 100 random iterations per gap. Please see Supporting Information S1 for a detailed breakdown of each time series by the number of missing or, respectively, filled years.

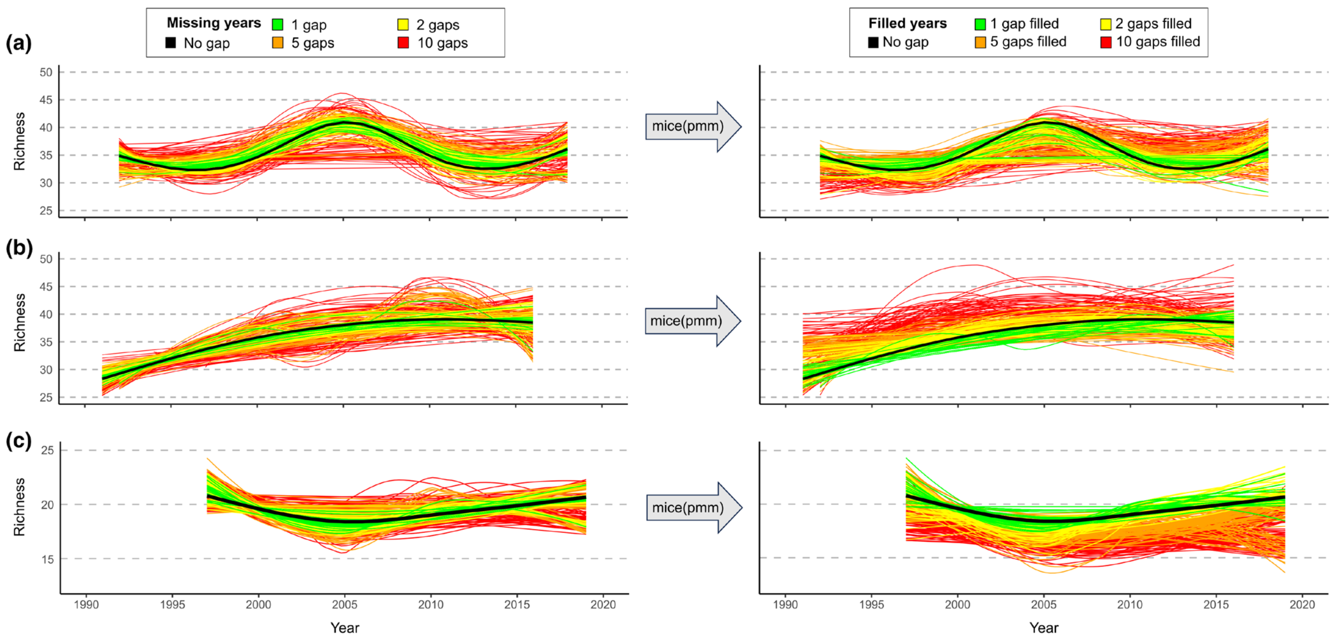

FIGURE 3 Variability in time series from (a) Denmark, (b) the Netherlands and (c) Sweden analysed using generalised additive models with randomly introduced gaps (left; black: original trend; green: 1 gap; yellow: 2 gaps; orange: 5 gaps; red: 10 gaps) and gaps filled using the *mice* function of the `mice` R package. Each coloured line represents one of 100 random iterations per gap. Please see Supporting Information S2 for a detailed breakdown of each time series by the number of missing or, respectively, filled years.

Information S3 and S4). A similar decrease was notably visible when these introduced gaps were subsequently imputed, though to a lesser extent. In Sweden, AIC values displayed an opposite trend, increasing with a greater number of artificially filled gaps. Deviance explained did not follow a common trend across the three time series, but the variability among imputed trends increased with the number of introduced and, respectively, filled gaps (Supporting Information S5 and S6). In Denmark, the explained deviance showed a reduction with an increasing number of gaps or equivalently filled gaps. Conversely, in the Netherlands

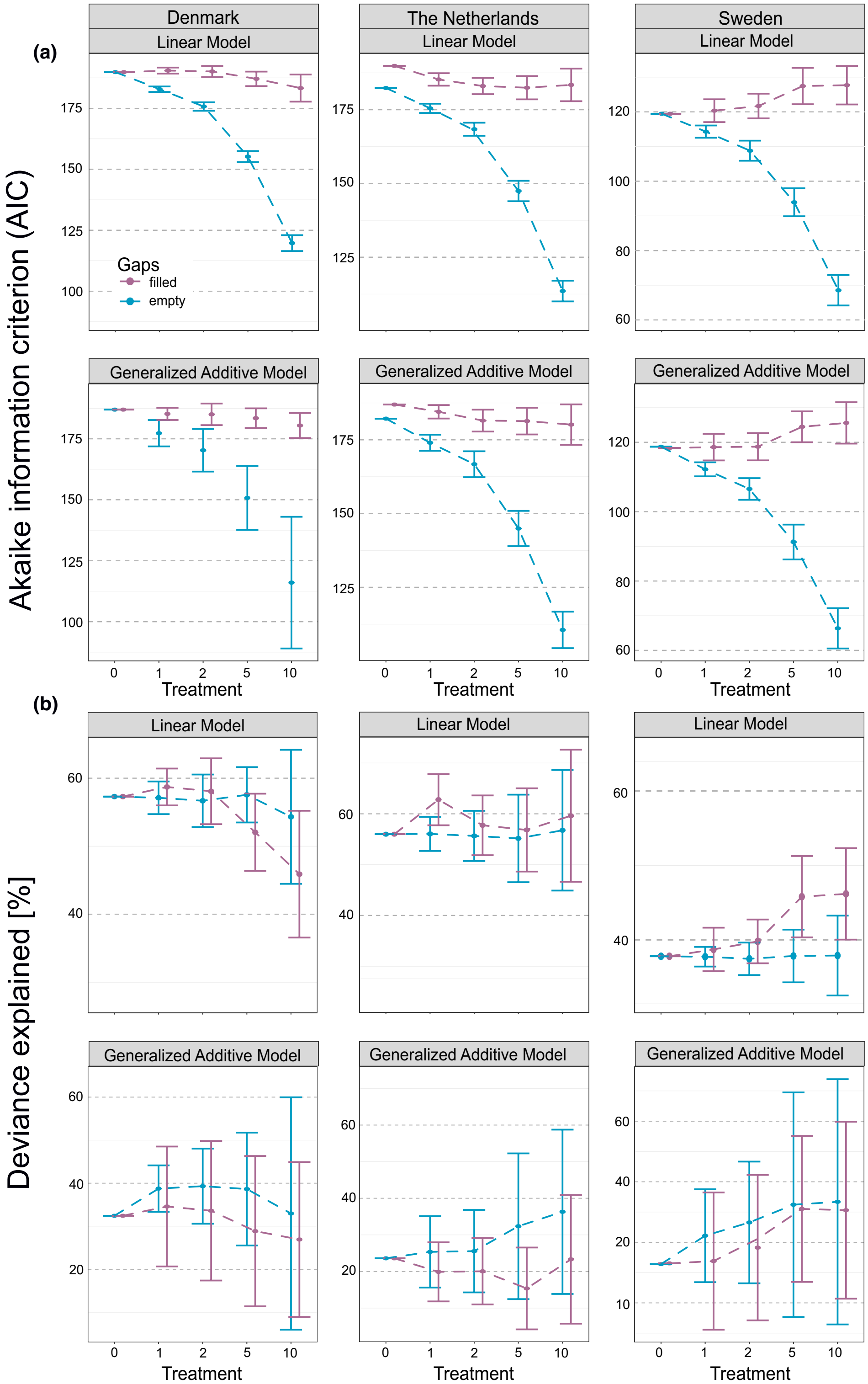

FIGURE 4 Akaike information criterion (a) and explained deviance (b), including their respective standard deviation across the randomised time series with gaps (blue) and after gaps were filled (purple) to explore model fit for the time series with randomly increased gaps (blue) and filled gaps (purple) for the time series from Denmark (left), the Netherlands (middle) and Sweden (right).

and Sweden, there was an observed increase in deviance explained, albeit with substantial variability, suggesting a more complex pattern in these regions. Generally, the standard deviation of the explained deviance was larger in generalised additive models, pertaining to its non-linear nature. In contrast to both AIC and deviance explained, the RMSE showed almost no change with an increasing number of gaps and, respectively, filled gaps (Supporting Information S7).

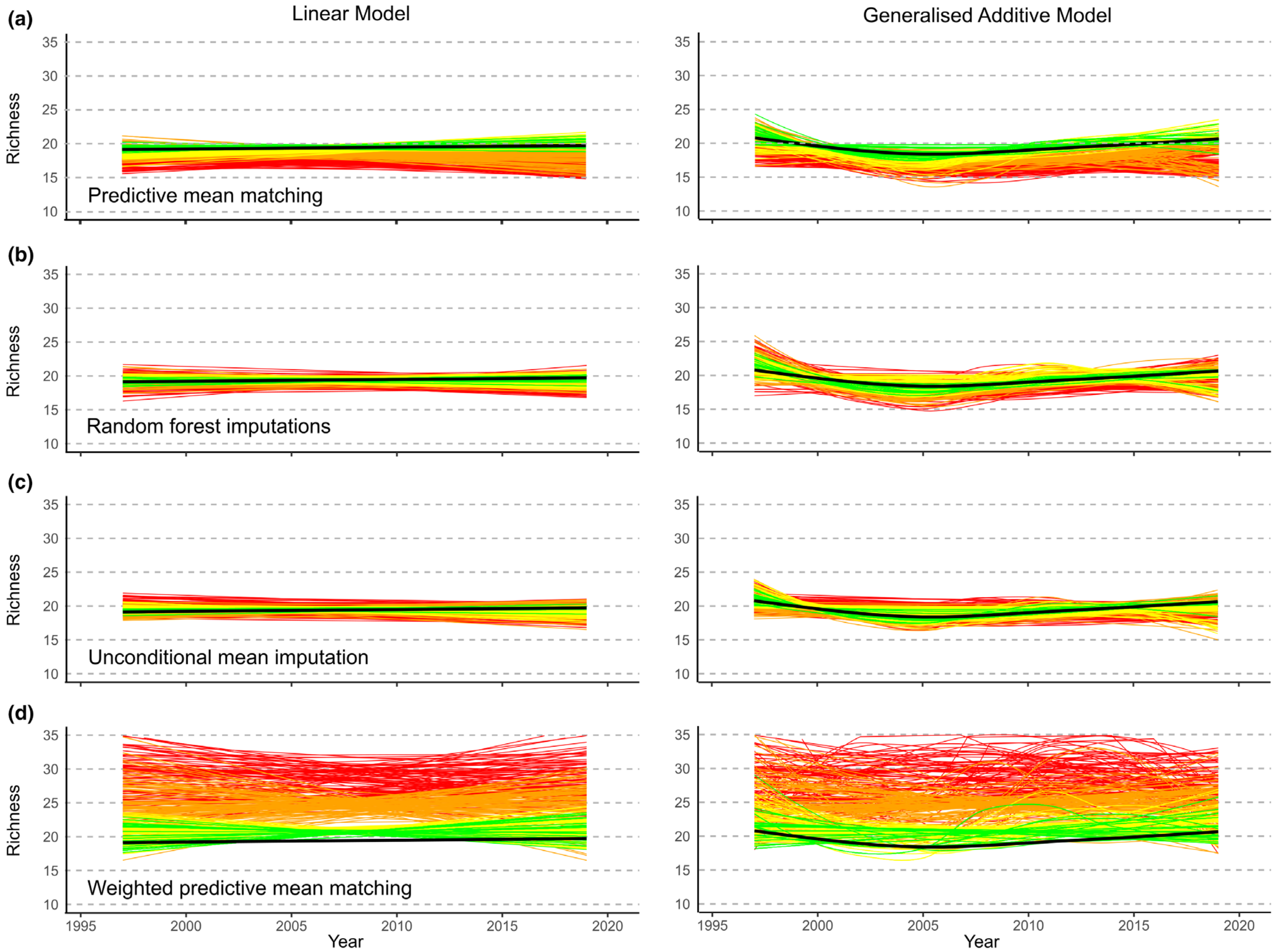

FIGURE 5 Variability in the randomised Swedish time series (chosen due to its length and richness for comparative purposes) analysed using linear (left) and generalised additive models (right) after filling 1 (green), 2 (yellow), 5 (orange) or, respectively, 10 (red) gaps using *Predictive Mean Matching* (a), *Random Forest Imputations* (b), *Random Sample from Observed Values* (c), and *Weighted Predictive Mean Matching* (d) using the *mice* function of the `mice` R package compared with the respective original trend (black). Each coloured line represents one of 100 random iterations per gap.

Comparing different gap-filling algorithms, we found substantial differences in performance among the four tested gap-filling algorithms (Figure 5). Using *Random Forest Imputations* and *Random Sample from Observed Values* resulted in considerably less variation with the increasing number of filled gaps in both linear and generalised additive models. *Weighted Predictive Mean Matching*, however, resulted in substantial deviation from the raw data trend with increasing gap numbers, with species richness becoming increasingly positive (Supporting Information S8).

3.3 | Influence of gap filling on meta-regression modelling

In the case of the time series from Denmark and Sweden, the meta-regression trend derived from the monotonic trends exhibited a slight positive inclination (Figure 6), which remained consistent even as the number of gaps and, respectively, filled gaps increased. The initially positive trends of time series from the Netherlands, however, experienced a reduction in the positive direction of trends, decreasing from +100 without gaps to +50 with 10 gaps (Supporting Information S9). In all three cases, the variance of each trend decreased with increasing gaps but increased when these were filled (Figure 6a–c).

4 | DISCUSSION

When investigating the impact of missing years and subsequent data gap filling using different methods (i.e. imputation algorithms) on the accuracy of riverine macroinvertebrate time series from Norway, Sweden and Denmark, we found that both the number of gaps (hypothesis i) and the increase in filled gaps (hypothesis ii) caused richness trends to deviate progressively from the original over time. This deviation, causing fluctuations in both directions, affects subsequent inferences. While imputing missing values

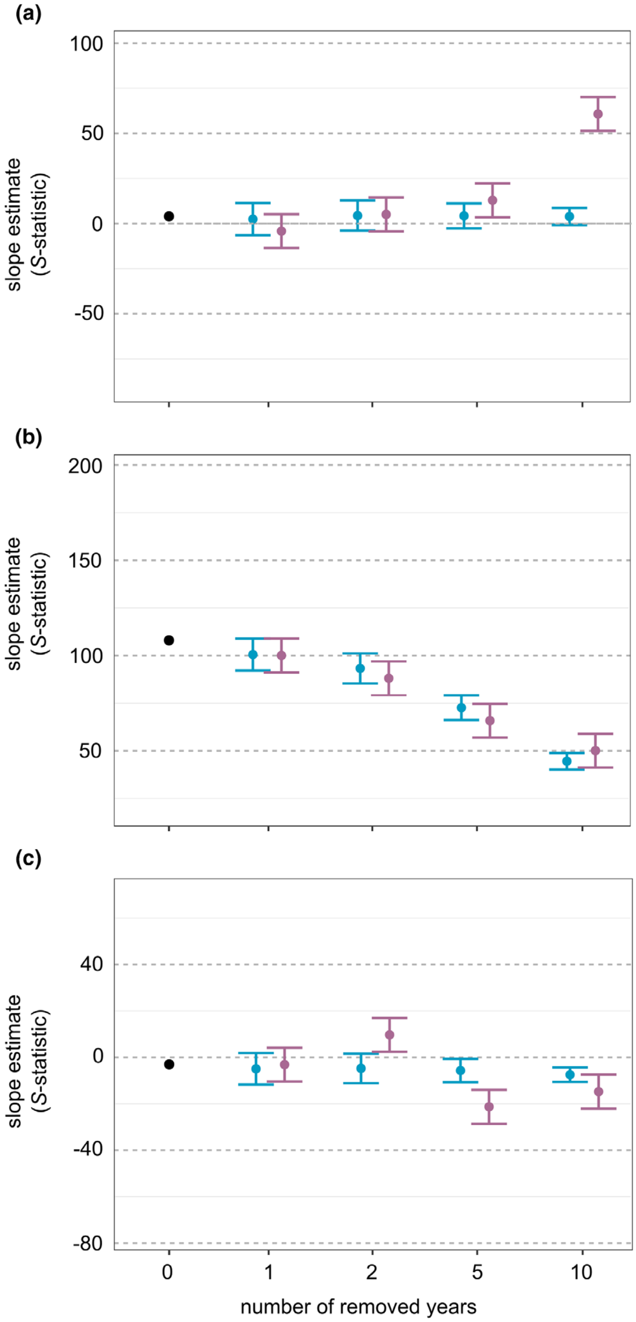

FIGURE 6 Slope (S-Statistic) averaged across the 100 randomised introductions of 1, 2, 5 and 10 gaps (blue) and, respectively, after being filled with the *mice* function (purple) for the time series from (a) Denmark, (b) the Netherlands, (c) Sweden. The black dot indicates the slope of the original time series.

may appear to address data gaps, relying on predictive models, such as Predictive Mean Matching and its weighted variants to infer biodiversity patterns from datasets with over 20% missing values can raise critical concerns about the interpretability of the outcome from time series. Such approaches, although statistically sophisticated, risk amplifying biases and misrepresenting ecological dynamics. This becomes particularly problematic when predictor variables or covariates (in addition to temporal variables like the respective year used in this study) included in models build upon time series with imputed data that may ultimately fail to adequately capture—and explain—temporal and spatial variability and introduce spurious trends or erroneous interpretations. Here, we further discuss the importance of adequate selection of statistical models and addressing associated uncertainties for the scientific process and management interfaces of riverine biodiversity (Lottig & Carpenter, 2012; Larson et al., 2023). Only by carefully considering these intricacies can long-term data effectively assess biodiversity trends and support the conservation of riverine ecosystems.

4.1 | Data gaps and their implications on biodiversity trends

Data gaps within large spatial or temporal datasets are commonly caused by irregular sampling, technological issues, financial constraints (often fuelled by the perception that funding agencies are reluctant to support biomonitoring projects; Geijzendorffer et al., 2016) or social and logistic restrictions (Carvalho et al., 2023; Hossie et al., 2021). In our analysis, such gaps increased both the variability of the data and the uncertainty in trend estimates, particularly for linear and generalised additive models, with the level of uncertainty escalating as more gaps were introduced. This challenge has been widely recognised in biodiversity monitoring, where gaps are increasingly treated as a missing data problem, requiring statistical corrections to improve inference (Bowler et al., 2025).

Our results also demonstrate that gaps can distort trend interpretation—especially when the data are non-linear or contain temporal fluctuations—which remain ecologically meaningful and reflect important dynamics, such as disturbance-recovery cycles or successional changes (Collen et al., 2008; Conde et al., 2019; Strayer et al., 2017). In the time series from Denmark and the Netherlands, for example, post-2005 fluctuations may indicate regime shifts or structural changes in macroinvertebrate communities rather than random variation, emphasising the importance of retaining and analysing such variability in biodiversity assessments.

The necessity to interpret incomplete datasets often leads to the use of imputation methods, which, while helpful, can obscure underlying patterns, such as boom-bust dynamics (Strayer et al., 2017). Additionally, gaps in biodiversity time series are not limited to structured monitoring programmes but also extend to community science data, where biases in observer effort and spatial distribution further complicate trend assessments. While community science datasets are often scrutinised for their biases, professionally collected datasets are also subject to similar limitations, including observer variability, spatial biases and taxonomic inconsistencies (Binley & Bennett, 2023). Such limitations can hamper management and conservation efforts and impede progress towards sustainable ecosystem management (Dietze et al., 2018; Henry et al., 2023; James et al., 2007; Power, 2008). Misinterpretation of trends due to incomplete data can have far-reaching implications, particularly for ecosystem policy and management decisions (Hering et al., 2023) as policymakers and other relevant stakeholders heavily rely on accurate trend analyses to formulate effective conservation strategies and policies. For example, past misinterpretations of biodiversity decline due to data gaps have influenced conservation priorities, such as in the case of amphibian population assessments where incomplete monitoring led to either overestimation or underestimation of extinction risks, affecting conservation funding allocation (Adams et al., 2013; Hefley et al., 2013). When decision making is however derived from incomplete data, there is a risk of allocating resources inaccurately, implementing ineffective conservation measures (e.g. for floodplain and river restorations; Sinclair et al., 2023; Stoltefaut et al., 2024) and fostering an overall misunderstanding of ecosystem health (Lackey, 2001). In short, the repercussions of data gaps extend beyond statistical limitations, influencing conservation priorities, ecosystem policy and sustainable management practices.

While the focus of this study is on time-series data, as true long-term biodiversity time series of annually collected data extending over a decade or more are increasingly rare (but see Haase et al., 2023), other approaches to handling missing data often incorporate spatial correlations in addition to temporal dependencies. Methods such as geostatistical modelling (e.g. kriging; Montero et al., 2015) and hierarchical Bayesian approaches (Schrodt et al., 2015) leverage spatial autocorrelation to estimate missing values. However, such spatially driven imputations may not always be feasible in long-term monitoring datasets where sampling is irregular, geographic coverage is limited, or environmental variability disrupts spatial autocorrelation. While MICE does not inherently include spatial information, its framework can be adapted to incorporate spatial covariates, potentially enhancing its applicability for datasets where both temporal and spatial gaps exist. Exploring such adaptations could be a valuable avenue for future research.

4.2 | Filling algorithm performance

The performance of filling missing data varied among algorithms. Missing data imputation using *Predictive Mean Matching* has been tested before and heralded for its versatility (Maziero et al., 2023), although acknowledging potential pitfalls that may lead to biases (Allison, 2015; Kleinke, 2017). While several studies (e.g. in medical fields) have tested the applicability of numerous imputation approaches (Chhabra et al., 2017), and ecological applications exist for specific taxa, to our knowledge no study has systematically evaluated these methods in the context of long-term, community-level biodiversity trends using real-world macroinvertebrate monitoring data (but see Vink et al., 2015; Yang & Kim, 2017). Here, we found that filling gaps in time series using *Predictive Mean Matching* did not necessarily reduce the confidence in the observed trends; instead, the variability appeared to increase. Interestingly, time series characterised by previously flat linear trends exhibited an increased negative or positive variance post-imputation. This intensification of directionality in trends and variance after filling the gaps indicates that the imputation of data may amplify existing trends, suggesting a 'cascading effect' For instance, time series with previously flat trends developed stronger positive or negative slopes, while those with pre-existing trends saw these trends become steeper. This compounding of trend direction can be seen as a 'snowball effect'—where once a directional trend begins, imputation tends to strengthen it further. In the case of generalised additive models, increasing numbers of imputed data using *Predictive Mean Matching* resulted in substantial deviation from the original trend, exceeding the variability observed when gaps were left unfilled. However, when complete and consistent annual data are crucial to compute reliable temporal biodiversity trends (Sridevi et al., 2018; Toutain et al., 2024), it is important to consider the choice of different gap-filling algorithms, because the performance of gap-filling success is also influenced by the extent of missing data.

Predictive Mean Matching is often the preferred method of choice for imputing missing data due to its ability to incorporate predictive modelling which facilitates refined estimates using missing values (Chhabra et al., 2017; Shah et al., 2014). By adjusting predicted values based on the similarity of means between observed and predicted values, *Predictive Mean Matching* aims to help maintain the statistical properties of the original dataset, thereby contributing to the accuracy and reliability of imputations. However, *Predictive Mean Matching* introduced considerable variation into the imputed time series as the number of missing gaps increases. Our findings suggest that *Predictive Mean Matching* relies on a more rigid predictive model and may struggle to maintain accuracy as the number of gaps increases. The observed deviation in *Predictive Mean Matching* could stem from a sensitivity to missing values and the challenge of preserving the original statistical properties of the data under higher rates of missingness. An adaptation of *Predictive Mean Matching that has been argued to be more robust*, like *Weighted Predictive Mean Matching* (Maziero et al., 2023), performed poorly, potentially due to the weighting of observed values, leading to higher species richness predictions. Hence, filling gaps using *Predictive Mean Matching* as well as *Weighted Predictive Mean Matching* may not be appropriate for time series with a larger number of gaps (i.e. more than ~25% of available data, which in our case typically corresponded to 5 or more missing years out of 20), while it may work relatively well for single gaps. This threshold emerged empirically from our comparisons of model behaviour and deviation patterns across increasing numbers of gaps and should be seen as a preliminary guide rather than a universal rule. The observed limitations in the percentage of missing data that will hinder the performance of *Predictive Mean Matching* should therefore be subject to further investigation.

Random Forest Imputations and *Random Sample from Observed Values* outperformed *Predictive Mean Matching*, showcasing resilience and flexibility in capturing the underlying temporal patterns even in the presence of substantial missing data, remaining closer to the original time series' trend. The use of linear models and generalised additive models was intentional to account for both simple linear trends and non-linear patterns in ecological time series of limited length, aligning with standard monitoring frameworks (Didham et al., 2020). While smaller datasets can lead to greater sensitivity in imputation methods, using year as a predictor in both the imputation process and subsequent regression modelling may raise concerns about potential bias but is methodologically justified for preserving temporal continuity in ecological time-series data. Given the inherent temporal autocorrelation in such datasets, excluding year could produce implausible imputations that distort ecological patterns rather than reflecting realistic trends. Validation metrics (e.g. RMSE, explained deviance) further confirm that the superior performance of Random Forest imputations stems from their ability to capture complex temporal dependencies, not from artificially inflated accuracy due to shared predictors. As such, the greater methodological fit and superior performance of *Random Forest Imputations* and *Random Sample from Observed Values* compared with *Predictive Mean Matching* in handling the increased gaps in the time series can be attributed to several factors. Random Forest's strength lies in its ability to capture complex patterns in data, making it more resilient to the challenges posed by missing values (Ali et al., 2012; Shah et al., 2014). The ensemble nature of Random Forest, which combines predictions from multiple decision trees and is less likely to be disrupted by incomplete data. Additionally, the *Random Sample from Observed Values* method introduces low additional diversity by randomly selecting values from the existing dataset, mitigating the impact of biased data (i.e. outliers) on the overall trend.

4.3 | Gap and gap filling on model fit complexity

While AIC values are comparable across different model structures as long as the underlying input data remains the same, we observed that an increase in the number of gaps within time series corresponded to lower AIC values (Nakagawa & Freckleton, 2011). This may be misinterpreted as indicating a more accurate model fit overall and even affect model selection procedures (Huque et al., 2018), while in fact suggesting that the models may fit easier to the data (i.e. time series) with an increasing number of gaps, likely due to the reduced data information, and thus fewer penalties for the loss of degrees of freedom (Chaurasia & Harel, 2012). However, when filling these gaps, we noted a decrease in AIC values in two out of three time series (i.e. the time series from the Netherlands and Denmark). While lower AIC values typically indicate improved model fit, in this context, the decrease likely reflects an escalation in model complexity and potential overfitting to the imputed data rather than a genuine improvement in explanatory power. This increase in AIC following the imputation thus indicates an escalation in complexity and, potentially, overfitting of the model to the 'complete' data. The imputation introduces additional data points along with associated variability, challenging the model to account for these new elements (van Buuren, 2018). These results therefore emphasise that lower AIC values in the presence of numerous data gaps may create misleading confidence in the model's adequacy. The ease of fitting due to data gaps does not necessarily enhance accuracy, necessitating a careful interpretation of AIC to ensure it genuinely reflects the model's explanatory power rather than merely its simplicity (Dennis et al., 2019). It is important to recognise that AIC values in such scenarios should not be viewed as absolute indicators of model quality but rather as reflective of the changing dataset properties. This underscores the need for additional metrics and careful consideration of the data context when interpreting AIC in studies involving missing data and analyses relying on imputed data.

A higher variability in the root mean square error (RMSE) when dealing with increasing data gaps or after filling or imputing missing values suggests that the model's predictive accuracy is more sensitive to the handling of missing data. This increased variability may indicate challenges in accurately capturing the underlying time-series trend. Emphasising the importance of robust imputation methods to minimise the impact of gaps on model performance, the average RMSE remained constant over time, indicating that the model's overall predictive accuracy was resilient to the challenges posed by missing data. This suggests that, while individual predictions might exhibit greater variability in the presence of data gaps, the model maintained a consistent level of performance on average, reinforcing its reliability in handling temporal variations and gaps in the dataset. Deviance explained, on the contrary, rescales all values relative to the respective null model, thus is more comparable among models than AIC value. Contrasting to what we expected, we found that the explained deviance varied among models and methods used for gap filling. We observed that while the average deviance explained remained almost constant, the standard deviation of the deviance explained increased as the number of gaps in the time series increased, and this pattern remained consistent whether the gaps were filled. The rise in standard deviation with more numerous gaps suggests an elevated level of unexplained variability introduced by the gaps. Therefore, we deduce that both the presence of data gaps and the imputation process contribute to the generation of more complex models, displaying diverse behaviours to accommodate the introduced variability (Junninen et al., 2004).

However, conventional metrics such as AIC values, the RMSE or the deviance explained are arguably not ideal for quantifying the impacts of missing data and, respectively, filled data gaps on trend analysis, as they may oversimplify the complex interplay of factors influencing trends. Consequently, we recommend adopting a more nuanced approach to handling missing data in time-series analysis. This could involve developing more sophisticated algorithms, including those incorporating machine learning methods and comprehensive metric evaluations, to provide a more accurate depiction of the complex dynamics involved. In this evolving landscape of data analysis, it is essential to adopt a holistic framework that not only considers AIC values and deviance explained but also accounts for the quality and completeness of the underlying data. Recognising the impact of data gaps and potential confounding variables will lead to a more robust understanding of ecological trends and ensure that any revealed or filled gaps are accurately interpreted.

4.4 | Meta-regressions with gaps

In evaluating each time series, the impact of introducing and filling data gaps varies, highlighting the challenges of managing incomplete data and the unique dynamics of each series. For instance, the Mann-Kendall trend analysis of the time series from the Netherlands shows consistent, nearly linear trends. In contrast, the Swedish and Danish time series exhibit more complex non-linear trends, as identified by the generalised additive model. Here, the removal of years leads to notable deviations in the monotonic direction, resulting in less positive overall trends. Introducing gaps results in estimated trends with narrower confidence intervals, suggesting an apparent robustness in meta-regressions. This suggests that meta-regressions may be less sensitive to individual data gaps or imputation artefacts compared with single time-series analyses, especially when trends are aggregated across multiple series. The time series from the Netherlands shows a substantial positive trend, likely due to an increase in richness in the initial years. However, as the number of gaps increases, there is a reduction in the richness trend. Conversely, the Swedish time series demonstrates higher resilience to gap introduction, implying a more stable community or reduced variability in external factors over time, reinforcing the value of considering both the structure of the individual time series and the broader analytical context when applying meta-regression methods.

Notably, these fluctuations were increased when the gaps were filled using the *Predictive Mean Matching* algorithm from the `mice` R package (Van Buuren & Groothuis-Oudshoorn, 2011), suggesting an impact from the imputation on the overall trend (Menéndez García et al., 2022). This could be attributed to the *Predictive Mean Matching* algorithm's ability to leverage predictive modelling and matching, maintaining statistical properties during imputation. The algorithm likely contributes to the alteration of the underlying trend, making the impact of gap filling more conspicuous, particularly when considering the assumed monotonicity in meta-regressions. This nuance underscores not just the accuracy of current trend estimations using meta-regression but also the reliability of associated ecological inferences, which can deteriorate when time series are altered, for example, by imputing missing data.

4.5 | Implications

The utilisation of diverse methodologies and algorithms can yield significantly different outcomes, carrying profound implications for water quality management and biodiversity policy formulation. (Breznau et al., 2022; Kim et al., 2023). Methodological choices and heterogeneous data compilation can influence the interpretation of long-term biodiversity trends, as seen in large-scale databases, such as BioTime (Dornelas et al., 2018). While these datasets provide invaluable resources for biodiversity assessments, previous studies have highlighted how analytical approaches can shape the resulting conclusions (Choi et al., 2023; Dornelas et al., 2014), underscoring the importance of careful methodological consideration in biodiversity trend analyses. The imputation of missing data becomes crucial in addressing gaps that hinder comprehensive assessments, particularly in estimating parameters like beta diversity that require relating different (e.g. local and regional) information to capture the full spectrum of biodiversity dynamics (Toutain et al., 2024). However, imputation approaches may cause a substantial deviation from the original or 'true' trend when not applied with substantial care (Allison, 2015; White et al., 2011). Missing years or the absence of target species observations can have severe implications, especially in environmental and biodiversity monitoring data, such as for endangered or rare species, non-native species or harmful algae blooms, within different aquatic ecosystems over time. Also, specific social conjectures such as economic recessions or the lockdown caused by the recent COVID pandemic can jeopardise data acquisition due to sampling and access restrictions (Hossie et al., 2021).

Aside from considering varying performances among the different imputation approaches (Shah et al., 2014), our results underline the many intricacies of time-series analyses, including the importance of carefully considering methodological adaptability in addressing missing data and interpreting ecological trends. Moreover, our study highlights the resilience of meta-regression methods in mitigating the effect of missing data, offering a robust approach for generating reliable inferences. While our study focuses on three distinct countries—Denmark, the Netherlands and Sweden——the selected long-term monitoring datasets originate from temperate lowland rivers that share common ecological and climatic characteristics with many other river systems across Europe. As such, they provide insights into biodiversity trends at a broader geographical scale, particularly within temperate freshwater ecosystems of Europe. The approach used here can thus be extended to larger datasets across multiple regions, demonstrating its applicability at a continental scale. While our study focused on species richness due to its widespread use in ecological monitoring and reporting, the analytical framework we apply could be extended to other biodiversity metrics, such as abundance, turnover or functional diversity. Future applications of this approach across a wider range of diversity measures could offer further insight into the robustness of ecological inferences under different types and levels of data loss. This comprehensive insight is essential for making informed decisions in water quality management. Therefore, our study emphasises the importance of thoughtful gap filling 'whether to fill (or not)' and careful algorithm selection. These practices play a crucial role in guiding evidence-based policy decisions and improving riverine ecosystem management practices. Consequently, we must exercise caution when deciding whether to fill (or not) the gaps in our dataset and in choosing the appropriate algorithm for doing so. As demonstrated here, these decisions can have important implications for policymakers seeking to develop new management strategies.

Based on our findings, we strongly recommend that the extent and nature of missing data from biodiversity monitoring must be carefully assessed before applying any imputation methods. These recommendations are based on analyses of riverine macroinvertebrate time series from three European monitoring sites and may not directly generalise to other taxa or regions. When data gaps are minimal, simple methods such as *Random Sample from Observed Values* may be sufficient. However, for datasets with more substantial gaps, *Random Forest Imputation* may be the most robust, as it preserves temporal patterns while minimising artificial trend inflation. In contrast, methods such as *Predictive Mean Matching* and its weighted variant should be used with caution, as observed trends tended to amplify deviations from the original trends, particularly as the number of missing years increased. In cases where biodiversity trends are to inform conservation policy and resource allocation, we advise against relying on imputed data alone and recommend supplementing analyses with sensitivity tests, assessing the stability of trends across different imputation approaches (Christopher Frey & Patil, 2002; van der Laan et al., 2018).

AUTHOR CONTRIBUTIONS

Phillip J. Haubrock: Conceptualisation; methodology; supervision; validation; visualisation; writing—original draft preparation; writing—review and editing. Ismael Soto and Rafael L. Macêdo: Writing—review and editing.

ACKNOWLEDGEMENTS

R.L.M. thanks the CAPES-Humboldt Research Fellowship for hosting his postdoctoral research in Germany. We thank Peter Haase for access to the data and acknowledge the help from Emma Hudgins, Ross Cuthbert and Melina Kourantidou, who provided comments on earlier versions of this manuscript. Open access publishing facilitated by Jihoceska Univerzita v Ceskych Budejovicich, as part of the Wiley - CzechELib agreement.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

The explicit data and R scripts used in this study are both available on Zenodo https://doi.org/10.5281/zenodo.15786677 (Soto et al., 2025). This Zenodo archive serves as a permanent, citable record of the precise data and code underlying this research. Ongoing development versions of the R scripts are maintained on GitHub: https://github.com/IsmaSA/data_gaps/ (repository created by Ismael Soto; isma-sa@hotmail.com). The complete dataset, from which the subset used in this work was derived, can be accessed via Haase et al. (2023).

ETHICAL APPROVAL

No ethics approval was required for this research.

ORCID

Phillip J. Haubrock https://orcid.org/0000-0003-2154-4341
Ismael Soto https://orcid.org/0000-0002-7288-6336

REFERENCES

Adams, M. J., Miller, D. A., Muths, E., Corn, P. S., Grant, E. H. C., Bailey, L. L., Grant, E. H., Fellers, G. M., Fisher, R. N., Sadinski, W. J., Waddle, H., & Walls, S. C. (2013). Trends in amphibian occupancy in the United States. *PLoS One*, *8*, e64347.

Akande, O., Li, F., & Reiter, J. (2017). An empirical comparison of multiple imputation methods for categorical data. *The American Statistician*, *71*, 162–170.

Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues*, *9*, 272.

Allison, P. (2015). *Imputation by predictive mean matching: Promise & peril*. Statistical Horizons.

Arnold, B. C., & Press, S. J. (1989). Compatible conditional distributions. *Journal of the American Statistical Association*, *84*, 152–156.

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, *20*, 40–49.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Green, P. (2009). Package 'lme4'. http://lme4.r-forge.r-project.org

Bellard, C., Genovesi, P., & Jeschke, J. M. (2016). Global patterns in threats to vertebrates by biological invasions. *Proceedings of the Royal Society B: Biological Sciences*, *283*, 20152454.

Benton, T. G., Bryant, D. M., Cole, L., & Crick, H. Q. (2002). Linking agricultural practice to insect and bird populations: A historical study over three decades. *Journal of Applied Ecology*, *39*, 673–687.

Binley, A. D., & Bennett, J. R. (2023). The data double standard. *Methods in Ecology and Evolution*, *14*, 1389–1397.

Blowes, S. A., Supp, S. R., Antão, L. H., Bates, A., Bruelheide, H., Chase, J. M., Moyes, F., Magurran, A., McGill, B., Myers-Smith, I. H., Winter, M., Bjorkman, A. D., Bowler, D. E., Byrnes, J. E. K., Gonzalez, A., Hines, J., Isbell, F., Jones, H. P., Navarro, L. M., … Dornelas, M. (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science*, *366*, 339–345.

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*, 50–67.

Breznau, N., Rinke, E. M., Wuttke, A., Nguyen, H. H., Adem, M., Adriaans, J., Alvarez-Benjumea, A., Andersen, H. K., Auer, D., Azevedo, F., Bahnsen, O., Bai, L., Balzer, D., Bauer, P. C., Bauer, G., Baumann, M., Baute, S., Benoit, V., Bernauer, J., … Van Assche, J. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, *119*, e2203150119.

Carvalho, R. L., Resende, A. F., Barlow, J., França, F. M., Moura, M. R., Maciel, R., Alves-Martins, F., Shutt, J., Nunes, C. A., Elias, F., Silveira, J. M., Stegmann, L., Baccaro, F. B., Juen, L., Schietti, J., Aragão, L., Berenguer, E., Castello, L., Costa, F. R. C., … Daly, D. (2023). Pervasive gaps in Amazonian ecological research. *Current Biology*, *33*, 3495–3504.

Chaurasia, A., & Harel, O. (2012). Using AIC in multiple linear regression framework with multiply imputed data. *Health Services and Outcomes Research Methodology*, *12*, 219–233.

Chhabra, G., Vashisht, V., & Ranjan, J. (2017). A comparison of multiple imputation methods for data with missing values. *Indian Journal of Science and Technology*, *10*(19), 1–7.

Choi, J., Lim, K. J., & Ji, B. (2023). Robust imputation method with context-aware voting ensemble model for management of water-quality data. *Water Research*, *243*, 120369.

Christopher Frey, H., & Patil, S. R. (2002). Identification and review of sensitivity analysis methods. *Risk Analysis*, *22*, 553–578.

Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: Addressing disparity in global monitoring. *Tropical Conservation Science*, *1*, 75–88.

Conde, D. A., Staerk, J., Colchero, F., da Silva, R., Schöley, J., Baden, H. M., Jouvet, L., Fa, J. E., Syed, H., Jongejans, E., Meiri, S., Gaillard, J. M., Chamberlain, S., Wilcken, J., Jones, O. R., Dahlgren, J. P., Steiner, U. K., Bland, L. M., Gomez-Mestre, I., … Vaupel, J. W. (2019). Data gaps and opportunities for comparative and conservation biology. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 9658–9664.

Dennis, B., Ponciano, J. M., Taper, M. L., & Lele, S. R. (2019). Errors in statistical inference under model misspecification: Evidence, hypothesis testing, and AIC. *Frontiers in Ecology and Evolution*, *7*, 372.

Didham, R. K., Basset, Y., Collins, C. M., Leather, S. R., Littlewood, N. A., Menz, M. H., Menz, M. H. M., Müller, J., Packer, L., Saunders, M. E., Schönrogge, K., Stewart, A. J. A., Yanoviak, S. P., & Hassall, C. (2020). Interpreting insect declines: Seven challenges and a way forward. *Insect Conservation and Diversity*, *13*, 103–114.

Dietze, M. C., Fox, A., Beck-Johnson, L. M., Betancourt, J. L., Hooten, M. B., Jarnevich, C. S., Keitt, T. H., Kenney, M. A., Laney, C. M., Larsen, L. G., Loescher, H. W., Lunch, C. K., Pijanowski, B. C., Randerson, J. T., Read, E. K., Tredennick, A. T., Vargas, R., Weathers, K. C., & White, E. P. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 1424–1432.

Dornelas, M., Antao, L. H., Moyes, F., Bates, A. E., Magurran, A. E., Adam, D., Akhmetzhanova, A. A., Appeltans, W., Arcos, J. M., Arnold, H., Ayyappan, N., Badihi, G., Baird, A. H., Barbosa, M., Barreto, T. E., Bässler, C., Bellgrove, A., Belmaker, J., Benedetti-Cecchi, L., … Murphy, G. (2018). BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography*, *27*, 760–786.

Dornelas, M., Gotelli, N. J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., & Magurran, A. E. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science*, *344*, 296–299.

Ellison, J. (2010). *Climate change impacts on, and vulnerability and adaptation of mangrove ecosystems*. https://figshare.utas.edu.au/articles/conference_contribution/Climate_change_impacts_on_and_vulnerability_and_adaptation_of_mangrove_ecosystems/23214626?file=40914233

Farley, S. S., Dawson, A., Goring, S. J., & Williams, J. W. (2018). Situating ecology as a big-data science: Current advances, challenges, and solutions. *Bioscience*, *68*, 563–576.

Fernández-Palacios, J. M., Kreft, H., Irl, S. D., Norder, S., Ah-Peng, C., Borges, P. A., Irl, S. D. H., Borges, P. A. V., Burns, K. C., de Nascimento, L., Meyer, J. Y., Montes, E., & Drake, D. R. (2021). Scientists' warning–the outstanding biodiversity of islands is in peril. *Global Ecology and Conservation*, *31*, e01847.

Forister, M. L., Jahner, J. P., Casner, K. L., Wilson, J. S., & Shapiro, A. M. (2011). The race is not to the swift: Long-term data reveal pervasive declines in California's low-elevation butterfly fauna. *Ecology*, *92*, 2222–2235.

Gaston, K. J., & Lawton, J. H. (1988). Patterns in the distribution and abundance of insect populations. *Nature*, *331*, 709–712.

Geijzendorffer, I. R., Regan, E. C., Pereira, H. M., Brotons, L., Brummitt, N., Gavish, Y., Haase, P., Martin, C. S., Mihoub, J.-. B., Secades, C., Schmeller, D. S., Stoll, S., Wetzel, F. T., & Walters, M. (2016). Bridging the gap between biodiversity data and policy reporting needs: An essential biodiversity variables perspective. *Journal of Applied Ecology*, *53*, 1341–1350.

Gelman, A., & Speed, T. P. (1993). Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *55*, 185–188.

Goulson, D. (2019). The insect apocalypse, and why it matters. *Current Biology*, *29*, R967–R971.

Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs shift towards increased linkage and suitability overlap under climate change. *Communications Biology*, *3*, 233.

Gutwinski, S., Schreiter, S., Deutscher, K., & Fazel, S. (2021). The prevalence of mental disorders among homeless people in high-income countries: An updated systematic review and meta-regression analysis. *PLoS Medicine*, *18*, e1003750.

Haase, P., Bowler, D. E., Baker, N. J., Bonada, N., Domisch, S., Garcia Marquez, J. R., Heino, J., Hering, D., Jähnig, S. C., Schmidt-Kloiber, A., Stubbington, R., Altermatt, F., Álvarez-Cabria, M., Amatulli, G., Angeler, D. G., Archambaud-Suard, G., Jorrín, I. A., Aspin, T., Azpiroz, I., … Welti, E. A. (2023). The recovery of European freshwater biodiversity has come to a halt. *Nature*, *620*, 582–588.

Haase, P., Pauls, S., Sundermann, A., & Zenker, A. (2004). Testing different sorting techniques in macroinvertebrate samples from running waters. *Limnologica*, *34*(4), 366–378.

Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., Goulson, D., & De Kroon, H. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS One*, *12*, e0185809.

Hallmann, C. A., Zeegers, T., van Klink, R., Vermeulen, R., van Wielink, P., Spijkers, H., van Deijk, J., van Steenis, W., & Jongejans, E. (2020). Declining abundance of beetles, moths and caddisflies in The Netherlands. *Insect Conservation and Diversity*, *13*, 127–139.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S., & Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*, 156–162.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*, 281–291.

Hastie, T., & Hastie, M. T. (2015). Package 'gam'. R Package Version, 90124–3.

Haubrock, P. J., Pilotto, F., & Haase, P. (2020). Do changes in temperature affect EU water framework directive compliant assessment results of central European streams? *Environmental Sciences Europe*, *32*, 1–13.

Haubrock, P. J., Pilotto, F., & Haase, P. (2023). Multidecadal data indicate increase of aquatic insects in Central European streams. *Science of the Total Environment*, *879*, 163017.

Hefley, T. J., Tyre, A. J., Baasch, D. M., & Blankenship, E. E. (2013). Nondetection sampling bias in marked presence-only data. *Ecology and Evolution*, *3*, 5225–5236.

Henry, M., Leung, B., Cuthbert, R. N., Bodey, T. W., Ahmed, D. A., Angulo, E., Balzani, P., Briski, E., Courchamp, F., Hulme, P. E., Kouba, A., Kourantidou, M., Liu, C., Macêdo, R. L., Oficialdegui, F. J., Renault, D., Soto, I., Tarkan, A. S., Turbelin, A. J., … Haubrock, P. J. (2023). Unveiling the hidden economic toll of biological invasions in the European Union. *Environmental Sciences Europe*, *35*, 1–16.

Hering, D., Schürings, C., Wenskus, F., Blackstock, K., Borja, A., Birk, S., Bullock, C., Carvalho, L., Dagher-Kharrat, M. B., Lakner, S., Lovrić, N., McGuinness, S., Nabuurs, G. J., Sánchez-Arcilla, A., Settele, J., & Pe'er, G. (2023). Securing success for the nature restoration law. *Science*, *382*, 1248–1250.

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*, 669477.

Huque, M. H., Carlin, J. B., Simpson, J. A., & Lee, K. J. (2018). A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Medical Research Methodology*, *18*, 1–16.

Jäger, S., Allhorn, A., & Bießmann, F. (2021). A benchmark for data imputation methods. *Frontiers in Big Data*, *4*, 693674.

James, E. A., Milenkiewicz, M. T., & Bucknam, A. (2007). *Participatory action research for educational leadership: Using data-driven decision making to improve schools*. Sage Publications.

Jolani, S., Debray, T. P., Koffijberg, H., van Buuren, S., & Moons, K. G. (2015). Imputation of systematically missing predictors in an individual participant data meta-analysis: A generalized approach using MICE. *Statistics in Medicine*, *34*, 1841–1863.

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., & Kolehmainen, M. (2004). Methods for imputation of missing values in air quality data sets. *Atmospheric Environment*, *38*, 2895–2907.

Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., & Hooker, G. (2009). Data-intensive science: A new paradigm for biodiversity studies. *Bioscience*, *59*, 613–620.

Kéry, M., & Royle, J. A. (2016). Modeling static occurrence and species distributions using site-occupancy models. *Applied Hierarchical Modelling in Ecology*, *1*, 551–630.

Kim, H. G., Cho, K. H., & Recknagel, F. (2023). Time-series modelling of harmful cyanobacteria blooms by convolutional neural networks and wavelet generated time-frequency images of environmental driving variables. *Water Research*, *246*, 120662.

Kleinke, K. (2017). Multiple imputation under violated distributional assumptions: A systematic evaluation of the assumed robustness of predictive mean matching. *Journal of Educational and Behavioral Statistics*, *42*, 371–404.

Lackey, R. T. (2001). Values, policy, and ecosystem health: Options for resolving the many ecological policy issues we face depend on the concept of ecosystem health, but ecosystem health is based on controversial, value-based assumptions that masquerade as science. *Bioscience*, *51*, 437–443.

Larson, D. M., Bungula, W., Lee, A., McKean, C., Miller, F., Davis, K., Erickson, R., & Hlavecek, E. (2023). Reconstructing missing data by comparing common interpolation techniques: Applications for long-term water quality data. *Limnology and Oceanography*, *21*, 435–449.

Lavorel, S., Grigulis, K., Leitinger, G., Kohler, M., Schirpke, U., & Tappeiner, U. (2017). Historical trajectories in land use pattern and grassland ecosystem services in two European alpine landscapes. *Regional Environmental Change*, *17*(8), 2251–2264.

Lei, T., & Bae, K. T. (2013). Deviance statistics in model fit and selection in ROC studies. In *Medical imaging 2013: Image perception, observer performance, and technology assessment* (Vol. *8673*, pp. 335–342). SPIE.

Li, G., Fang, C., Li, Y., Wang, Z., Sun, S., He, S., Qi, W., Bao, C., Ma, H., Fan, Y., Feng, Y., & Liu, X. (2022). Global impacts of future urban expansion on terrestrial vertebrate diversity. *Nature Communications*, *13*, 1628.

Li, J., Guo, S., Ma, R., He, J., Zhang, X., Rui, D., Ding, Y., Li, Y., Jian, L., Cheng, J., & Guo, H. (2024). Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. *BMC Medical Research Methodology*, *24*, 41.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.

Lottig, N. R., & Carpenter, S. R. (2012). Interpolating and forecasting lake characteristics using long-term monitoring data. *Limnology and Oceanography*, *57*, 1113–1125. https://doi.org/10.4319/lo.2012.57.4.1113

Luo, Y. (2022). Evaluating the state of the art in missing data imputation for clinical data. *Briefings in Bioinformatics*, *23*, bbab489.

MacGregor, C. J., Williams, J., Bell, J., & Thomas, C. (2019). Moth biomass increases and decreases over 50 years in Britain. *Nature Ecology & Evolution*, *3*, 1645–1649.

Maziero, L. P., Rodrigues, S. A., Pai, A. D., Cremasco, C. P., & Gabriel Filho, L. R. A. (2023). Comparison of the performance of multiple imputation models in filling gaps in hourly and daily meteorological series from two locations in the state of São Paulo-Brazil. *Modeling Earth Systems and Environment*, *10*, 1815–1823.

Menéndez García, L. A., Menéndez Fernández, M., Sokoła-Szewioła, V., Álvarez de Prado, L., Ortiz Marqués, A., Fernández López, D., & Bernardo Sánchez, A. (2022). A method of pruning and random replacing of known values for comparing missing data imputation models for incomplete air quality time series. *Applied Sciences*, *12*, 6465.

Montero, J. M., Fernández-Avilés, G., & Mateu, J. (2015). *Spatial and spatio-temporal geostatistical modeling and kriging*. John Wiley & Sons.

Muñiz Pinto, O. B., Rodríguez Patiño, J. G., Revilla Rodríguez, M. I., Laza Martínez, A., Seoane Parra, S., & Franco, J. (2020). Inhomogeneity detection in phytoplankton time series using multivariate analyses. *Oceanologia*, *62*, 243–254.

Nakagawa, S., & Freckleton, R. P. (2008). Missing inaction: The dangers of ignoring missing data. *Trends in Ecology & Evolution*, *23*, 592–596.

Nakagawa, S., & Freckleton, R. P. (2011). Model averaging, missing data and multiple imputation: A case study for behavioural ecology. *Behavioral Ecology and Sociobiology*, *65*, 103–116.

Pavón-Jordán, D., Clausen, P., Dagys, M., Devos, K., Encarnaçao, V., Fox, A. D., Frost, T., Gaudard, C., Hornman, M., Keller, V., Langendoen, T., Ławicki, Ł., Lewis, L. J., Lorentsen, S.-. H., Luigujoe, L., Meissner, W., Molina, B., Musil, P., Musilova, Z., … Lehikoinen, A. (2019). Habitat-and species-mediated short-and long-term distributional changes in waterbird abundance linked to variation in European winter weather. *Diversity and Distributions*, *25*, 225–239.

Picapedra, P. H., Fernandes, C., Taborda, J., Baumgartner, G., & Sanches, P. V. (2020). A long-term study on zooplankton in two contrasting cascade reservoirs (Iguacu River, Brazil): Effects of inter-annual, seasonal, and environmental factors. *PeerJ*, *8*, e8979.

Pili, A. N., Leroy, B., Measey, J. G., Farquhar, J. E., Toomes, A., Cassey, P., Chekunov, S., Grenié, M., van Winkel, D., Maria, L., Diesmos, M. L. L., Diesmos, A. C., Zurell, D., Courchamp, F., & Chapple, D. G. (2024). Forecasting potential invaders to prevent future biological invasions worldwide. *Global Change Biology*, *30*, e17399.

Pilotto, F., Kühn, I., Adrian, R., Alber, R., Alignier, A., Andrews, C., Bäck, J., Barbaro, L., Beaumont, D., Beenaerts, N., Benham, S., Boukal, D. S., Bretagnolle, V., Camatti, E., Canullo, R., Cardoso, P. G., Ens, B. J., Everaert, G., Evtimova, V., … Haase, P. (2020). Meta-analysis of multidecadal biodiversity trends in Europe. *Nature Communications*, *11*, 3486.

Power, D. J. (2008). Understanding data-driven decision support systems. *Information Systems Management*, *25*, 149–154.

Prendergast, J. R., & Eversham, B. C. (1997). Species richness covariance in higher taxa: Empirical tests of the biodiversity indicator concept. *Ecography*, *20*, 210–216.

Pyšek, P., Hulme, P. E., Simberloff, D., Bacher, S., Blackburn, T. M., Carlton, J. T., Dawson, W., Essl, F., Foxcroft, L. C., Genovesi, P., Jeschke, J. M., Kühn, I., Liebhold, A. M., Mandrak, N. E., Meyerson, L. A., Pauchard, A., Pergl, J., Roy, H. E., Seebens, H., … Richardson, D. M. (2020). Scientists' warning on invasive alien species. *Biological Reviews*, *95*, 1511–1534.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, *27*, 85–96.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, F. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.

Royston, P., & White, I. R. (2011). Multiple imputation by chained equations (MICE): Implementation in Stata. *Journal of Statistical Software*, *45*, 1–20.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.

Schrodt, F., Kattge, J., Shan, H., Fazayeli, F., Joswig, J., Banerjee, A., Reichstein, M., Bönisch, G., Díaz, S., Dickie, J., Gillison, A., Karpatne, A., Lavorel, S., Leadley, P., Wirth, C. B., Wright, I. J., Wright, S. J.,

& Reich, P. B. (2015). BHPMF–a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Global Ecology and Biogeography*, *24*, 1510–1521.
Seibold, S., Gossner, M. M., Simons, N. K., Blüthgen, N., Müller, J., Ambarlı, D., Ammer, C., Bauhus, J., Fischer, M., Habel, J. C., Linsenmair, K. E., Nauss, T., Penone, C., Prati, D., Schall, P., Schulze, E.-D., Vogt, J., Wöllauer, S., & Weisser, W. W. (2019). Arthropod decline in grasslands and forests is associated with landscape-level drivers. *Nature*, *574*, 671–674.
Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *American Journal of Epidemiology*, *179*, 764–774.
Sinclair, J. S., Mademann, J. A., Haubrock, P. J., & Haase, P. (2023). Primarily neutral effects of river restoration on macroinvertebrates, macrophytes, and fishes after a decade of monitoring. *Restoration Ecology*, *31*, e13840.
Singh, J. S. (2002). The biodiversity crisis: A multifaceted review. *Current Science*, *82*, 638–647.
Soto, I., Cuthbert, R. N., Ahmed, D. A., Kouba, A., Domisch, S., Marquez, J. R., Marquez, J. R. G., Beidas, A., Amatulli, G., Kiesel, J., Shen, L. Q., Florencio, M., Lima, H., Briski, E., Altermatt, F., Archambaud-Suard, G., Borza, P., Csabai, Z., Datry, T., … Haubrock, P. J. (2023). Tracking a killer shrimp: *Dikerogammarus villosus* invasion dynamics across Europe. *Diversity and Distributions*, *29*, 157–172.
Soto, I., Haubrock, P. J., & Macêdo, R. L. (2025). Data from: To fill or not to fill: Comparing imputation methods for improved riverine long-term biodiversity monitoring. *Zenodo Repository*. https://doi.org/10.5281/zenodo.15786677
Spengler, S. J. (2000). Bioinformatics in the information age. *Science*, *287*, 1221–1223.
Sridevi, S., Parthasarathy, S., & Rajaram, S. (2018). An effective prediction system for time series data using pattern matching algorithms. *International Journal of Industrial Engineering*, *25*, 123–136.
Stoltefaut, T., Haubrock, P. J., Welti, E. A., Baker, N. J., & Haase, P. (2024). A long-term case study indicates improvements in floodplain biodiversity after river restoration. *Ecological Engineering*, *198*, 107143.
Strayer, D. L., D'Antonio, C. M., Essl, F., Fowler, M. S., Geist, J., Hilt, S., Jarić, I., Jöhnk, K., Jones, C. G., Lambin, X., Latzka, A. W., Pergl, J., Pyšek, P., Robertson, P., von Schmalensee, M., Stefansson, R. A., Wright, J., & Jeschke, J. M. (2017). Boom-bust dynamics in biological invasions: Towards an improved application of the concept. *Ecology Letters*, *20*, 1337–1350.
Stubbington, R., Chadd, R., Cid, N., Csabai, Z., Miliša, M., Morais, M., Munné, A., Pařil, P., Pešić, V., Tziortzis, I., Verdonschot, R. C. M., & Datry, T. (2018). Biomonitoring of intermittent rivers and ephemeral streams in Europe: Current practice and priorities to enhance ecological status assessments. *Science of the Total Environment*, *618*, 1096–1113.
Su, Y. S., Gelman, A., Hill, J., & Yajima, M. (2011). Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. *Journal of Statistical Software*, *45*, 1–31.
Touron-Poncet, H., Bernadet, C., Compin, A., Bargier, N., & Céréghino, R. (2014). Implementing the water framework directive in overseas Europe: A multimetric macroinvertebrate index for river bioassessment in Caribbean islands. *Limnologica*, *47*, 34–43.
Toutain, M., Soto, I., Rasmussen, J. J., Csabai, Z., Várbíró, G., Murphy, J. F., Balzani, P., Kouba, A., Renault, D., & Haubrock, P. J. (2024). Tracking long-term shifts in non-native freshwater macroinvertebrates across three European countries. *Science of the Total Environment*, *906*, 167402.
Van Buuren, S. (2011). Multiple imputation of multilevel data. In *Handbook of advanced multilevel analysis* (pp. 173–196). Routledge.
van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman & Hall/CRC.
Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1–67.
van Buuren, S., Groothuis-Oudshoorn, K., Robitzsch, A., Vink, G., Doove, L., & Jolani, S. (2015). Package 'mice'. Computer software.
van der Laan, M. J., Rose, S., Díaz, I., Luedtke, A. R., & van der Laan, M. J. (2018). Sensitivity analysis. In *Targeted learning in data science: Causal inference for complex longitudinal studies* (pp. 511–522). Springer.
Van Klink, R., Bowler, D. E., Gongalsky, K. B., Swengel, A. B., Gentile, A., & Chase, J. M. (2020). Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*, *368*, 417–420.
Variyath, A. M., Chen, J., & Abraham, B. (2010). Empirical likelihood based variable selection. *Journal of Statistical Planning and Inference*, *140*, 971–981.
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48.
Vink, G., Lazendic, G., & van Buuren, S. (2015). Partioned predictive mean matching as a large data multilevel imputation technique. *Psychological Test and Assessment Modeling*, *57*, 577–594.
Wallner, W. E. (1987). Factors affecting insect population dynamics: Differences between outbreak and non-outbreak species. *Annual Review of Entomology*, *32*, 317–340.
Wauchope, H. S., Amano, T., Geldmann, J., Johnston, A., Simmons, B. I., Sutherland, W. J., & Jones, J. P. (2021). Evaluating impact using time-series data. *Trends in Ecology & Evolution*, *36*(3), 196–205.
Wauchope, H. S., Shaw, J. D., & Terauds, A. (2019). A snapshot of biodiversity protection in Antarctica. *Nature Communications*, *10*(1), 946.
Wei, H., Fan, W., Wang, X., Lu, N., Dong, X., Zhao, Y., Ya, X., & Zhao, Y. (2017). Integrating supply and social demand in ecosystem services assessment: A review. *Ecosystem Services*, *25*, 15–27.
Welti, E. A., Bowler, D. E., Sinclair, J. S., Altermatt, F., Álvarez-Cabria, M., Amatulli, G., Angeler, D. G., Archambaud, G., Arrate Jorrín, I., Aspin, T., Azpiroz, I., Baker, N. J., Bañares, I., Barquín Ortiz, J., Bodin, C. L., Bonacina, L., Bonada, N., Bottarin, R., Cañedo-Argüelles, M., … Haase, P. (2024). Time series of freshwater macroinvertebrate abundances and site characteristics of European streams and rivers. *Scientific Data*, *11*, 601.
White, E. R. (2019). Minimum time required to detect population trends: The need for long-term monitoring programs. *Bioscience*, *69*, 40–46.
White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*, 377–399.
Wickham, H. (2006). An introduction to ggplot: An implementation of the grammar of graphics in R. Statistics, 1.
Woźnica, K., & Biecek, P. (2020). Does imputation matter? Benchmark for predictive models. *arXiv preprint* arXiv:2007.02837.
Yang, S., & Kim, J. K. (2017). Predictive mean matching imputation in survey sampling. *arXiv preprint* arXiv:1703.10256.
Zhang, Z. (2016). Missing data imputation: Focusing on single imputation. *Annals of Translational Medicine*, *4*, 9.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Supporting Information S1. Variability in time series from (a) Denmark, (b) the Netherlands, and (c) Sweden analysed using linear models with randomly introduced gaps (left; black: original trend; blue: 1 gap; green: 2 gaps; yellow: 5 gaps; purple: 10 gaps) and gaps filled using the *mice* function of the `mice` R package. Each line represents one of 100 random iterations per gap.

Supporting Information S2. Variability in time series from (a) Denmark, (b) the Netherlands, and (c) Sweden analysed using generalised additive models with randomly introduced gaps (left; black: original trend; blue: 1 gap; green: 2 gaps; yellow: 5 gaps; purple: 10 gaps) and gaps filled using the *mice* function of the `mice` R package. Each coloured line represents one of 100 random iterations.

Supporting Information S3. AIC averaged across the 100 replicates for each 1, 2, 5, and 10 annual gaps for the time series from (a) Denmark, (b) the Netherlands, and (c) Sweden.

Supporting Information S4. AIC averaged across the 100 replicates for each 1, 2, 5, and 10 annual gaps filled with *mice* for the time series from (a) Denmark, (b) the Netherlands, and (c) Sweden.

Supporting Information S5. Deviance explained averaged across the 100 replicates for each 1, 2, 5, and 10 annual gaps for the time series from (a) Denmark, (b) the Netherlands, and (c) Sweden.

Supporting Information S6. Deviance explained averaged across the 100 replicates for each 1, 2, 5, and 10 annual gaps filled with *mice* for the time series from (a) Denmark, (b) the Netherlands, and (c) Sweden.

Supporting Information S7. Root Mean Square Error and the respective standard deviation across the randomised time series with gaps (blue) and after gaps were filled (purple) to explore model fit for the time series with randomly increased gaps (blue) and filled gaps (purple) for the time series from (left) Denmark, (middle) the Netherlands, (right) Sweden.

Supporting Information S8. Comparative analysis of imputation algorithms on filled gaps evaluated for its performance within the context of both statistical models (linear and generalised additive model).

Supporting Information S9. Meta-regression model results for (a) Denmark, (b) the Netherlands, and (c) Sweden.

How to cite this article: Haubrock, P. J., Soto, I., & Macêdo, R. L. (2025). To fill or not to fill: Comparing imputation methods for improved riverine long-term biodiversity monitoring. *Journal of Applied Ecology*, *62*, 2421–2438. https://doi.org/10.1111/1365-2664.70110