The forecasted mean of 80 percent of wild populations and communities shows no change, and why

Claudio Bozzuto[1 *], Anthony R. Ives [2]

[1] Wildlife Analysis GmbH, Oetlisbergstrasse 38, 8053 Zürich, Switzerland. bozzuto@wildlifeanalysis.ch. ORCID 0000-0003-0355-8379

[2] Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI 53706, USA. arives@wisc.edu. ORCID 0000-0001-9375-9523

[*] Corresponding Author: Claudio Bozzuto, Wildlife Analysis GmbH, 8053 Zürich, Switzerland. bozzuto@wildlifeanalysis.ch

Abstract

Broad-scale shifts in biodiversity demand a deeper understanding of both historical and future time trends across the Tree of Life. Retrospective analyses indicate that some groups, such as birds, fishes and insects, have experienced significant declines in recent decades. Broader investigations, on the other hand, reveal a more complex picture, with declines and increases roughly balanced. Here, we address the prospective task to statistically forecast biodiversity trends and assess our ability to do so based on available data. We analyzed approximately 43,000 population-level and 10,000 community-level time series from localities across the globe and covering taxa across the Tree of Life. Using unobserved components models and model selection, we found that only 23.5% of wild populations exhibit forecasted time trends (increasing and decreasing ones roughly balanced), while 76.5% show a forecasted constant mean. For communities (temporal α- and β-diversity), this figure rises to 84.0%. To explore the underlying reasons for the high prevalence of forecasted constant means, we used machine learning techniques (random forest models and SHAP analysis). We identified high variation in the rate of change (RoC) in the time series as the most important predictor of a forecasted constant mean. Further, life-history traits and locality accounted for 39.6 % of the variation in the RoC variance of anmiotes, and sampling unit explained 44.4 % of that variability for vascular plants. These findings contribute to the emerging appreciation of RoC as a key facet of biodiversity responses to environmental change. In summary, despite trends being visible in ecological time-series data, parsimonious statistical models will very often suggest a forecasted constant mean. Our results offer a tempered perspective that can help realistically inform stakeholders and the general public, and guide the assessment of conservation actions.

Keywords: biodiversity; machine learning; predictions; random forest models; SHAP analysis; state-space models; time series analysis.

Introduction

Since E. O. Wilson's seminal book "The Diversity of Life" (1992) – published the same year as the United Nations "Earth Summit" held in Rio de Janeiro – a growing body of research has chronicled an accelerating biodiversity crisis (e.g. Dirzo et al. 2014, and references therein). Numerous studies document alarming shifts in species' abundance and genetic diversity, in community composition, and in extinction rates. However, whether these trends amount to an impending or unfolding "sixth mass extinction" is contested. For example, Wiens and Saban (2025) show that, although contemporary losses are elevated relative to long-term baselines, they lack the catastrophic severity characteristic of past mass-extinction events.

Retrospective temporal analyses of recent biodiversity changes have shown that some taxonomic groups in particular have experienced drastic declines: examples are birds (e.g. Rosenberg et al. 2019), fishes (e.g. UNEP-WCMC 2024) and insects (e.g. Hallmann et al. 2017). Broader studies, on the other hand, have found that the picture of biodiversity trends is more nuanced than a narrative of a pervasive decline implies, with no net change at local scales, and with declining and increasing trends roughly balanced (Dornelas et al. 2023, and references therein). At the same time, dozens of databases have been designed and made publicly available, containing many thousands of time series at the population-level and at the community-level (also referred to as assemblages), spanning a wide range of taxa and localities globally (e.g. Johnson et al. 2024, and references therein). These data collections offer an unprecedented opportunity to investigate biodiversity changes at different temporal and spatial scales. Partly based on these databases, several authors have explored why reported biodiversity trends diverge so widely. To just name three reasons, one general source of confusion lies in the diversity of definitions, methods, and approaches adopted (Boënnec et al. 2024). A second reason is methodological, pointing to the common neglect of correlation structures in the data (Johnson et al. 2024) and the presence of mathematical biases (Toszogyova et al. 2024) when computing summary trends. A third reason concerns the time series data themselves. Although an impressive volume of records is now available for studies of biodiversity change, these data vary greatly in their temporal extent, reliability and the types of biodiversity attributes they capture (e.g. Toszogyova et al. 2024). Nonetheless, Boënnec et al. (2024) rightly state that the

confusion on biodiversity trends is not weakening the knowledge on biodiversity loss; instead, working on uncertainty is essential to consolidate our respective knowledge.

Prospective studies on expected biodiversity trends predominantly focus on capturing the effect of human-caused drivers, including climate change, on future populations and community states (e.g. Urban et al. 2016, Dietze et al. 2024). Furthermore, long-term scenario-based projections dominate, for example based on projected climate change responses over several decades. Studies indicate that if current trends in human-caused stressors continue or worsen in the absence of significant conservation interventions, many systems might experience local biodiversity declines on the order of 10–20% over the next few decades (e.g. Newbold et al. 2015, IPBES 2019). As opposed to long-term projections, several authors have recently stressed the importance of near-term forecasts (i.e. a forecast horizon of at most a decade), not least to match the typical timescale of decision-making in biodiversity conservation management (Wood et al. 2018, Tulloch et al. 2020, Dietze et al. 2024). But despite this change of timescale to better and more actionably understand expected biodiversity trends, the focus on process-based approaches encompassing mechanistic links between stressors, populations and ecosystems remains a formidable task. In fact, Lasky et al. (2020) rightly ask about the usefulness to include processes from a multitude of fields (such as ecophysiology, evolutionary biology, demography): "We argue that answering this critical question is ultimately an empirical exercise requiring a substantial amount of data that have not been integrated for any system to date."

The number of studies making and reporting near-term statistical forecasts is rapidly growing (Lewis et al. 2022). Broad assessments across localities, realms and taxonomic groups, on the other hand, continue being rare. Compared to retrospective biodiversity trend analyses, the above-mentioned publicly available data collections have not yet been used extensively to construct a coarse forecasting baseline, against which more refined and system-specific approaches can be evaluated. Some of these studies have analyzed single taxonomic groups or realms (Harris et al. 2018, Olsson et al. 2025). Other studies have used several data collections to either assess the performance of forecasting models (Ward et al. 2014, Johnson et al. 2024) or to investigate the concept of predictability in ecology and conservation (Pennekamp et al. 2019, Bozzuto and Ives 2024). Predictability-related methods are mainly insightful when analyzing time series spanning at least 20 years (Bozzuto and Ives 2024). A considerable fraction of the mentioned data collections, however, is made up of (very) short time series. For example, in the widely used BioTIME database 50% of the time series span at most 7 years, with a minimum of two years (Dornelas et al. 2025). We can expect that adapting approaches from climate change science will significantly improve our ability to predict expected biodiversity changes at different spatial and temporal scales (Dietze et al. 2024). But for the time being, we need to start answering a simple question: what do

parsimonious, robust statistical models show about expected near-term biodiversity trends, based on typical ecological time-series data?

For this study we analyzed approximately 43,000 population-level and 10,000 community-level time series globally, covering 64 taxonomic classes nested within 6 kingdoms. We found that, for the vast majority of populations and communities, parsimonious statistical models show a forecasted constant mean rather than a directional time trend. We investigate important drivers of this result and ask about the contribution of characteristics of typical ecological time-series data to this result. Our findings do not in any way dismiss further declines or increases. They do, however, challenge the prevailing narrative that past trends are expected to "automatically" continue. We show that, based on data available through high-profile databases, the most frequent parsimonious forecast is for no expected change instead, even though trends may be visible in the data.

Materials and methods

Time series data

We used time-series data from ten different sources/databases (see Appendix S1: Table S1 for additional details, including access information): BioTIME (Dornelas et al. 2025), GBIF (Filippova et al. 2024), GPDD (Prendergast et al. 2010), LPI (Loh et al. 2005), LTER (Richter 2023), MASTREE+ (Hacket-Pain et al. 2022), RivFishTIME (Comte et al. 2021), TREAM (Welti et al. 2024), Swiss long-term phytoplankton data (Pomati et al. 2020), and Swiss long-term macrofungal data (Egli 2006). To avoid the potential repeated inclusion of the same time series, we used each data collection for a specific taxonomic group and/or realm. Moreover, where possible we used a data collection for either population-level or community-level analyses. We achieved an almost global coverage for population-level data (Appendix S1: Figure S1; community-level: Appendix S1: Figure S2), and a broad population-level taxonomic coverage with 64 classes nested within six kingdoms (Appendix S1: Figure S3; community-level: Appendix S1: Figure S4).

We considered time series spanning at least 5 years. We allowed a proportion of missing values of at most 25% (of which at most 50% in consecutive order), and we excluded binary data (e.g. presence-absence time series). We required the first and last time series value a) to be not missing values, and b) to be not zero (Toszogyova et al. 2024). Unless time series had already been transformed and contained negative values, we transformed the data using the inverse hyperbolic sine function. This transformation resembles the log-transformation for positive values, but in contrast to the latter it can handle zeros. In Appendix S2: Section S1 we provide additional details on data preprocessing.

At the community-level we considered two biodiversity metrics covering temporal α-diversity and temporal β-diversity. We only analyzed communities with at least 5 species (populations). To construct α-diversity time series, we computed year-specific Hill numbers (Dornelas et al. 2014). In particular, we used the ${}^{1}D$ metric which is the exponential of the Shannon entropy. To construct β-diversity time series, we compared each pair of temporally consecutive community compositions and quantified their similarity using Ruzicka's metric (Cha 2007). The latter is related to Jaccard's similarity metric, but additionally accounts for population abundances instead of identities alone. We then transformed these time series for further analysis as described above.

Unobserved components models and model selection

Unobserved components models (UCMs) are univariate state-space models which are widely used in many science fields, including ecology (e.g. Harvey et al. 1997, Durbin and Koopman 2012, Auger-Méthé et al. 2021) . They are flexible, explicitly separate measurement error from process error, can handle missing values, and accommodate nonstationary dynamics without the strict assumption of stationarity that is common in other time series frameworks. As a template for our four UCMs (see below), we used the local linear trend model given by

$$y_t = m_t + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \qquad \text{eq. 1a}$$

$$m_{t+1} = m_t + s_t + \xi_t, \qquad \xi_t \sim N\left(0, \sigma_\xi^2\right), \qquad \text{eq. 1b}$$

$$s_{t+1} = s_t + \zeta_t, \qquad \zeta_t \sim N\left(0, \sigma_\zeta^2\right). \qquad \text{eq. 1c}$$

Eq. 1a relates the unobserved true mean (eq. 1b) to the data, y_t, with measurement error. The true mean (eq. 1b) evolves as a random walk and drifts due to a slope which also evolves as a random walk (eq. 1c). Depending on the variances included, many different UCMs can be constructed based on the local linear trend model. Preliminary investigations showed that, due to the typical characteristics of the time series here analyzed, the full local linear trend model (eq. 1a-c) almost never was selected as the best-fitting model. Thus, for our analysis we selected the following four models.

- *Deterministic constant*: This is the baseline model, with only eq. 1a included and where $m_t \equiv m$ is a constant. For this model, a single variance has to be estimated.
- *Deterministic trend*: This model includes both the measurement equation (eq. 1a) and the mean equation (1b), which however evolves deterministically. Therefore, $s_t \equiv s$ and $\sigma_\xi^2 = \sigma_\zeta^2 \equiv 0$, so that only one variance is estimated (eq. 1a).
- *Local level*: This model includes both the measurement equation (eq. 1a) and the mean equation (1b). The mean evolves stochastically as a random walk, but there is no stochastically changing slope. Therefore, $s_t = \sigma_\zeta^2 \equiv 0$, and two variances are estimated (eq. 1a-b).
- *Smooth trend*: This model includes all three equations (eq. 1a-c). However, the mean equation has no stochastic variation of its own. Therefore, $\sigma_\xi^2 \equiv 0$, and two variances are estimated (eq. 1a, 1c).

For each population or community time series, we fitted all four models and selected the best-fitting one using AIC corrected for small sample size (AICc).

Two of our models produce forecasts with no time trend (the deterministic constant model and the local level model), whereas the other two models will allow the forecasted mean to include a linear time trend. But whether the forecasted mean is constant or changes linearly, fitted UCMs can show a stochastic trend over the timespan of the data. To illustrate this, we analyzed the Eastern monarch butterfly population (Figure 1) (e.g. Zylstra et al. 2021). Each row shows a selected period of the whole time series (Figure 1a-b), and for each timespan we show all four model fits and include ΔAICc values, where a value of zero designates the best-fitting model. The left column in Figure 1 shows the two models with a forecasted constant mean, whereas the right column shows the two models with forecasted linear trends. We highlight three aspects. First, for the full time series (Figure 1a-b) and the one up to year 2018 (Figure 1c-d), the local level model is the best-fitting one. Here, the mean evolves stochastically and follows the data, including turns (2013, 2018). But because we can have no knowledge about future stochastic deviations (eq. 1b), the logical and robust consequence is a forecasted constant mean.

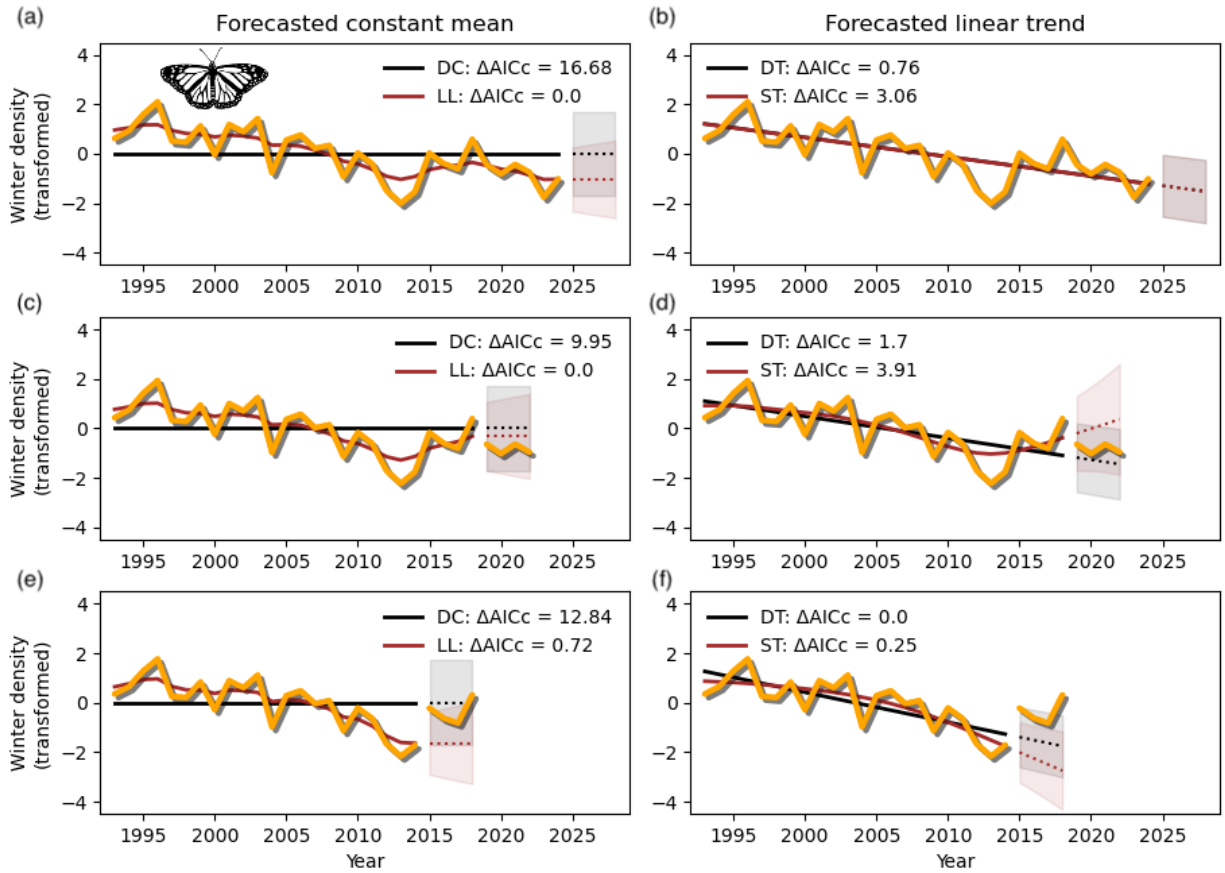

Figure 1. Eastern monarch butterfly population: dynamics and forecasts. Panels (a-b) show the complete time series (up to 2024), along with reconstructed mean trajectories and forecasts based on our four models. Panels (c-d) and (e-f) show the analogous results, however, assuming a shorter timespan (up to 2018 and 2014, respectively). The left column shows the fits of the two models exhibiting a constant forecasted mean: deterministic constant model (DC) and local level model (LL). The right column shows the fits of the two models exhibiting a forecasted linear time trend: deterministic trend model (DT) and smooth trend model (ST). All legends show model name abbreviations along with ΔAICc values, where a value of zero designates the best-fitting model. All forecasts are shown for 4 years ahead, including 90% confidence intervals. Icon: public domain (www.phylopic.org).

Second, AICc as a selection metric covers all aspects of the fit, and it might not always confirm our "visual intuition". Despite the smooth trend model and its increasing forecast seem to confirm that after the monarch population hit its bottom in 2013/14 and increased again afterwards (Figure 1d), the local level model is the better model to explain the data up to 2018, and thus delivers the most parsimonious forecast. In fact (and unfortunately), the increasing period ended in 2018. Third, forecasting is not foreseeing: when only analyzing data up to year 2014, the best-fitting model is the deterministic trend

model with a forecasted mean continuing to decline. Needless to say, nobody could have anticipated the sharp increase in 2015. Here, even if the local level model (as second-best model) suggests a forecasted constant mean, its forecasted variance equally well (or better) captures the unforeseeable turn in dynamics. In sum, parsimonious models can predict a forecasted constant mean, despite stochastic trends being present in the data.

Analyzing the high prevalence of forecasted constant means

To investigate what best explains the unbalanced proportion of populations with forecasted constant mean vs. linear trend (see the *Results* section), we used a random forest classifier (Breiman 2001, Cutler et al. 2007). Two main advantages over similar methods like a logistic regression are that the algorithm automatically considers nonlinear effects and interaction effects among predictors (henceforth called features). We analyzed population-level and community-level results separately. For each set, we split all analyzed instances into a training and a test set, the latter one used for out-of-sample predictions. Because the target data were already imbalanced, we refrained from splitting the time-trend instances into declining vs. increasing categories – doing so would have further skewed the class balance and reduced sample support in each group.

Our selected features cover different aspects of the populations, communities and the analyzed data. To roughly summarize the time series dynamics, we first computed first-differences of the transformed time series; from eq. 1a, these are the differences $y_t - y_{t-1}$. Based on these differences (rates of change, RoC), we computed the mean(RoC) and the var(RoC) and included these two metrics as features. To capture "data quality", we included three features: time series length (henceforth 'length'), proportion of non-missing values (henceforth 'p(data)'), and proportion of unique (non-missing) values (henceforth 'p(unique)'). Furthermore, we included locality (coordinates), and taxonomic information. We matched all taxonomic data with the GBIF backbone taxonomy (https://doi.org/10.15468/39omei). In Appendix S2: Section S2 we provide more details on data preprocessing and feature selection.

Subsequently we used the trained random forest classifier to investigate feature importance in terms of predicting a constand vs. linearly changing forecasted mean. Hereto, we computed SHAP values (Shapley Additive ExPlanations values), a meanwhile widespread approach to explain machine learning models (Lundberg et al. 2020, Molnar 2025). Simply put, for each instance (e.g. a population) the respective prediction is compared to the overall average prediction, and based on game-theoretic methods the influence of each feature is computed to best explain the difference between the instance-specific prediction and overal average prediction. Advantages over alternative ways to compute feature

importance include the additive nature of SHAP values (for example to group features), their consistency (features that add more value to a prediction yield higher SHAP contributions), and their local reliability, that is, providing instance-level insights that aggregate to global interpretations (Molnar 2025). SHAP values have the same units as the target variables: for a classification, they are measured in log-odds. In our analysis, the sign and magnitude of the SHAP values jointly influence the predictions: positive values push predictions towards a forecasted constant mean, and negative values towards a forecasted time trend.

Understanding the drivers of important features

The SHAP analysis allows ranking features based on their importance, and var(RoC) emerged as the most important driver of the type of forecasted mean (see the *Results* section). As a further step, we used random forest regressors to investigate important predictors of var(RoC). A random forest regressor offers the same advantages as a classifier (see preceding subsection), with the target variable now being continuous instead of binary. We fit two different regressors with features we anticipated would affect var(RoC). First, sampling unit: we compared time series of reproductive output (e.g. pine cone counts) versus time series of, for example, counts of individual adults. We expected sampling unit to affect the amount of fluctuations in the data, in turn affecting var(RoC). Based on auxiliary data we could find, we analyzed population-level time series from the taxonomic class Magnoliopsida (dicots; n = 4,268), with reproductive output data stemming from MASTREE+ (Hacket-Pain et al. 2022), and all other time series from BioTIME (Dornelas et al. 2025) (Appendix S1: Table S1). Second, life history characteristics and locality: we expected population fluctuations to be associated with life history traits and locality (Albaladejo-Robles et al. 2023, and references therein), and consequently affecting var(RoC). Based on available auxiliary data, we restricted our analysis to amniote populations (birds, mammals, and reptiles; n = 6,916). In addition to the populations' locality, we included four life-history traits (Myhrvold et al. 2015): age at maturity, maximum lifespan, body mass, and yearly reproductive output. In Appendix S2: Section S3 we provide more details on the auxilliary data and data preprocessing.

Results

Forecasted dynamics of populations and communities

After fitting the unobserved components models and retaining time series with no missing attributes for the subsequent classification (next subsection), we grouped time series according to the type of forecasted mean of the respective best-fitting model: forecasted time trend (deterministic trend model and smooth trend model) vs. forecasted constant mean (deterministic constant model and local level model). We

found that only 23.5% of populations exhibit forecasted time trends (increasing and decreasing ones roughly 12% each), while 76.5% have a forecasted constant mean (Table 1). For communities, this figure rises to 84.0%, with a negligible difference between temporal α- and β-diversity time series. What is more, the simple baseline model (deterministic constant model) with a forecasted constant mean was selected as best-fitting model for 51.2% of the population-level and 69.6% of the community-level time series (Table 1).

Table 1. **Overview of forecasted mean results**. *n*: number of time series. *p*(constant mean): proportion of time series with a forecasted constant mean. *p*(time trend): proportion of time series with a forecasted linear trend. *p*(baseline): proportion of time series where the baseline model (deterministic constant model) is the best-fitting model.

	n	*p*(constant mean)	*p*(time trend)	*p*(baseline)
Populations	43,403	76.5 %	23.5 %	51.2 %
Communities				
total	10,409	84.0 %	16.0 %	69.6 %
α-diversity	5,098	84.6 %	15.4 %	64.6 %
β-diversity	5,311	83.4 %	16.6 %	74.7 %

Explaining why forecasted constant means dominate

To understand what biological and data-related characteristics drive the dominance of a forecasted constant mean, we first fit a random forest classifier with the type of forecasted mean as target, followed by a SHAP analysis. The trained random forest classifier achieved good predictive ability, with a weighted average f1-score of 0.82 (Appendix S1: Table S2). The model, however, was much better at predicting forecasted constant means (f1-score: 0.88) than forecasted time trends (f1-score: 0.65), despite we explicitely accounted for class imbalance (76.5% vs. 23.5%; Table 1). One reason might be the diffuse boundaries of feature values between the two classes (Appendix S1: Figure S5). A second reason might be the limited support for time series with forecasted time trends, due to class imbalance (Appendix S1: Table S1). A third reason might be that the features used here are less suited to explain forecasted time trends; see the section *Discussion* for more details.

The subsequent SHAP analysis at the population-level showed that the most important feature driving a predicted forecasted constant mean vs. time trend is the variation in the rate of change in the data, var(RoC), followed by mean(RoC) and time series length (Figure 2). The results at the community-level are very similar (Figure 2). For these importance rankings, we computed the mean absolute SHAP value per feature. In Appendix S1 with show the detailed SHAP value results: their distribution per feature (Figure S6); as a function of the respective feature values (Figures S7-S8; non-aggregated features); visualizations for aggregated features (taxonomy: Figures S9-S10; locality: Figures S11-S12).

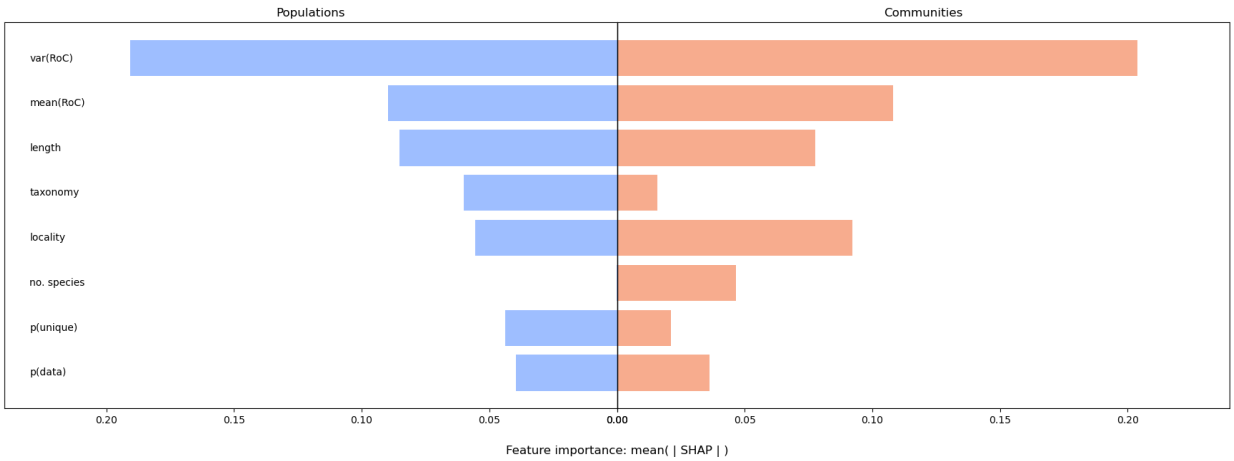

Figure 2. **Feature importance values from the random forest classifiers**. For each feature in the model, the mean absolute SHAP value is shown separately for population-level and community-level predictions. Note that the features 'locality' and 'taxonomy' are aggregated: 'locality' represents coordinates, and 'taxonomy' comprises four PCA components.

Both metrics var(RoC) and mean(RoC) – roughly capturing the dynamics in the time-series data – account for approximately 50% of the total importance. Time series with high yearly fluctuations are more likely to be associated with a forecasted constant mean (Figure 3a), and we find the same association for (very) low-magnitude mean(RoC) values, as well as high-magnitude mean(RoC) values (Figure 3d). As for time series length (Appendix S1: Figure S7-S8), expectedly short time series tend to contain spurious trends in the data, with forecasted means containing time trends too. Longer time series, however, will increase the probability of a forecasted constant mean. Despite other data characteristics (proportion of non-missing values and proportion of unique values), locality and taxonomy playing a less important role on their own (Figure 2), they partly contribute – via interactions – to the importance of var(RoC) and mean(RoC), as shown next.

What drives var(RoC)?

A multitude of factors can influence the extent of yearly fluctuations of populations and communities, including climatic conditions, human pressures like harvesting or land-use change, and community-specific species fluctuations, just to name a few. From a more technical perspective, one source contributing to var(RoC) in driving the type of forecasted trend are interaction effects with other features. In Figure 3 we show population-level SHAP values for var(RoC) and mean(RoC) as a function of the respective feature values, and where SHAP values are colored by a second feature. We highlight two findings. First, short time series increase the likelihood of a forecasted time trend for a given low var(RoC). Moreover, short time series increase the likelihood of a forecasted constant mean for a given high var(RoC). As for mean(RoC), short time series make up the bulk of high-magnitude trends in the data and tendentially push predictions towards a forecasted time trend. Second, the proportion of non-missing data – despite being of secondary importance overall (Figure 2) – tends to affect predictions in two ways via interaction effects. For high var(RoC) values, missing values in the data decrease the likelihood of a forecasted constant mean. For mean(RoC), on the other hand, missing data make up the bulk of high-magnitude trends in the data and tend to push predictions towards a forecasted constant mean.

An additional interaction with var(RoC) stems from taxonomy. In Figure 4 we show how population-level var(RoC) values are associated with all taxonomy-related data, resolved at the taxonomic rank of class; for better average values, we only show classes with at least 10 populations (41 out of 64 classes; *cf*. Appendix S1: Figure S3). The differences at the kingdom level are minimal, but differences build up markedly towards classes. Already at the phylum level var(RoC) differs markedly between invertebrate and vertebrate animals or between mosses (*sensu latu*) and vascular plants.

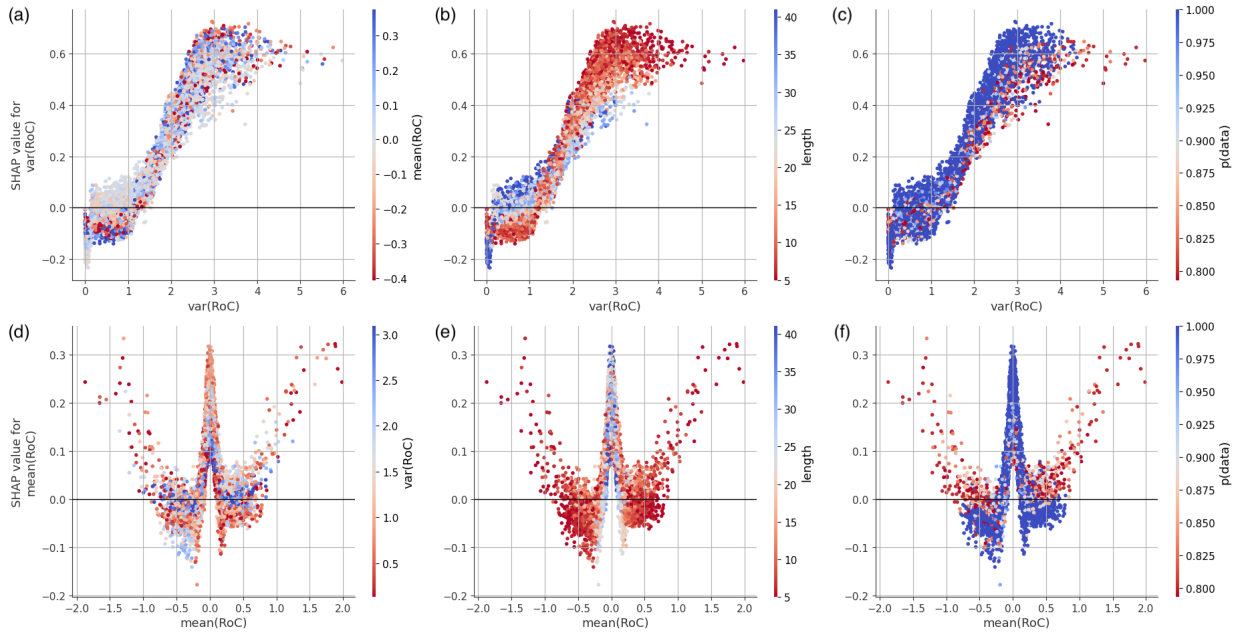

Figure 3. **SHAP values and feature interactions**. For the population-level time series, panels (a-c) show SHAP values of var(RoC) plotted against its values, and panels (d-f) show SHAP values of mean(RoC) plotted against its values. In addition, the data in each panel are colored based on the values of a different feature (colorbar labels) to visualize interactions among features.

At the low end of average var(RoC), vertebrate classes (phylum Chordata) stand out, exhibiting low to medium annual fluctuations. To explore what drives this variability within vertebrate populations, we fitted a random forest regressor using four life-history traits and locality as predictors. The model explained R^2 = 39.6% of the variation in var(RoC) (range-normalized RMSE: 16.1%). Appendix S1: Table S3 reports R^2 values when each predictor group is analyzed separately. Within the phylum Tracheophyta (vascular plants), we found pronounced differences in average var(RoC) values among classes (Figure 4). We therefore fitted a second random forest regressor to assess the effect of sampling unit on var(RoC) of populations in the phylum Magnoliopsida (dicots), that is, comparing time series based on reproductive output to time series of individual plants or similar. The model explained R^2 = 44.4% of the variation in var(RoC) (range-normalized RMSE: 16.3%; Appendix S1: Table S3). This result also clarifies the variation within the vascular plants (Figure 4): at the high end of average var(RoC), all but one Pinopsida (conifers) time series are based on some reproductive output as sampling unit; at the low end of average var(RoC), only about one percent of Liliopsida (monocots) time series is based on reproductive output. Note that the two reported R^2 values are not directly comparable because of differences in data preprocessing (Appendix S2: Section S3).

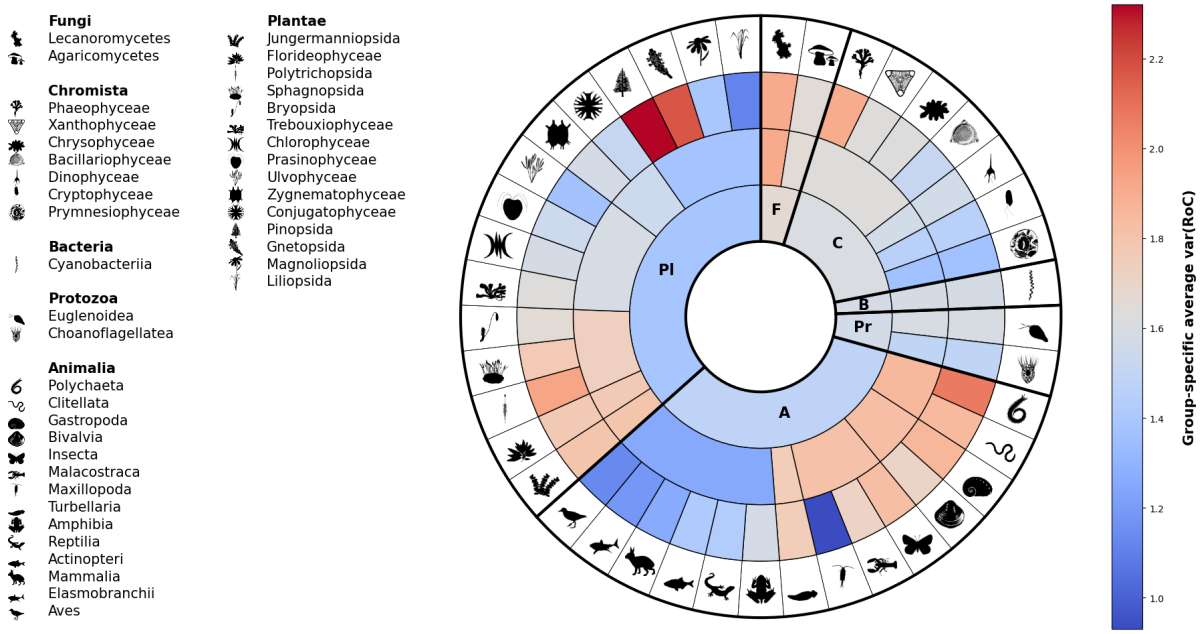

Figure 4. Visualization of the taxonomy dependence of var(RoC). For the population-level time series, the nested pie chart shows average var(RoC) values per taxon (ranks kingdom, phylum, class). Only classes with at least ten populations are shown (41 out of 64 classes: Appendix S1: Figure S3). All taxa are ordered in descending order (clockwise; classes and phyla within higher taxon). Kingdom abbreviations: F[ungi], C[hromista], B[acteria], Pr[otozoa], A[nimalia], Pl[antae]. Icons: public domain (www.phylopic.org).

Discussion

In this study, we asked a simple question: What do parsimonious, robust, and rigorously grounded statistical models reveal about expected near-term biodiversity trends when applied to typical ecological time-series data? Such insights can provide a coarse forecasting baseline against which more targeted, system-specific approaches can be evaluated. We fitted unobserved components models to roughly 43,000 population-level and 10,000 community-level time series – spanning the globe and the Tree of Life – to generate forecasts. We then investigated which biological and data-related characteristics best explain the high prevalence of forecasted constant means (i.e. no expected time trends). Below, we place our findings in context and discuss some implications for biodiversity conservation management.

Compared to the prevailing narrative of a pervasive biodiversity decline (see the section *Introduction*), our result of about 80% of population-level and community-level time series showing a forecasted constant mean appears sobering (Table 1). To make things worse, our baseline model

(deterministic constant model) – that is, a simple horizontal line with fluctuations around it (Figure 1) – was selected as best-fitting model for ~50% of the population-level and ~70% of the community-level time series. As we have illustrated with the Eastern monarch butterfly population (Figure 1), even if stochastic time trends are visible in the data, a parsimonious statistical model might nonetheless suggest a forecasted constant mean. The main point here is not the constant mean per se. Rather, the forecast distribution also accounts for a potential future decline or increase (Figure 1), but the predicted chance of either one is 50%. This, in our opinion, goes somewhat against a common misconception, namely if a population has been declining for a decade, say, then we can "expect" this decline to continue in the near-term future. Parsimonious models take trends *and* variation around a changing mean into account. Only if the signal (e.g. a declining mean) can be detected with confidence within the noise, a forecasted time trend is likely. If the data are too noisy, a visible trend in the data might simply be a random realization of a process with no long-term time trend.

We propose a simple yet comprehensive baseline for forecasting biodiversity trends, covering wide geographic and taxonomic breadth and a substantial time-series dataset. Our use of parsimonious state-space models (e.g. Harvey et al. 1997). and random forest + SHAP analyses (e.g. Molnar 2025) follows established best practices for robustness and interpretability. Extensions that directly model spatial and phylogenetic correlations (Johnson et al. 2024) – for example to compute an overall expected trend for a specific taxonomic group (e.g. WWF 2024) – or incorporate environmental covariates are logical next steps, particularly for well-studied systems where such data are abundant. In a proof-of-concept re-analysis of the Eastern monarch butterfly population (Appendix S1: Figure S13, Appendix S2: Section S4), however, adding an important covariate (spring rain) had minimal impact on the best-fitting model or produced overly pessimistic short-term forecasts (predicting declines after 2014, when the population rebounded). We worry that under current ecological data regimes integrating covariates might yield limited gains (see also Lasky et al. 2020). That said, we very much hope that broader and more data-intensive approaches supported by the initiative Group on Earth Observations Biodiversity Observation Network (GEO BON 2025) will eventually surpass our simple baseline in forecasting accuracy. Given sufficient data, these efforts could refine biodiversity forecasts well beyond the baseline we propose here.

Even within our framework for typical ecological time series there remains ample opportunity for further exploration and improvement. For instance, provided data availability, it would be valuable to perform a taxonomic cross-database forecast comparison to assess whether group-specific forecasts are consistent across different data collections – or whether discrepancies suggest unintended biases in specific datasets. Additionally, computing supplementary biodiversity metrics could prove instructive. Although no single metric will capture all aspects of biodiversity change (e.g. McGill et al. 2015),

comparing different approaches may help identify sensible and unbiased best practices. Finally, the poorer performance of the random forest classifier in predicting populations or communities with a forecasted time trend deserves further analysis. One reason for the poor performance might be that the features included are less suited for forecasted time trend predictions. Despite mean(RoC) and var(RoC) roughly mirroring the time series dynamics – and therefore potentially picking up effects of environmental drivers – the data quality of typical ecological time series without explicitly considering covariate data might pose a challenge at discriminating forecasted time trends. Such insights will certainly help to better understand – and to use in practice – our main result of approximately 80% of populations and communities showing a forecasted constant mean.

Our results highlight important drivers of the type of forecasted mean, but to what degree can these insights inform biodiversity conservation management? We have shown that the variance of yearly fluctuations in the data, var(RoC), is a critical metric. This variance is influenced by factors such as life-history characteristics and climatic variation – factors that are inherently challenging to modify. Additionally, our taxonomic analysis reveals that certain groups, like vertebrates, are more likely to exhibit forecasted time trends (Figure 4). However, since vertebrates represent only one phylum among many in the Tree of Life, relying on their trends might be misleading when generalizing across taxa. Our findings also show that the choice of sampling unit has a significant effect on var(RoC). Biodiversity conservationists should be aware that, for instance, time series of reproductive output often tend to produce a forecasted constant mean. This awareness can inform the design of monitoring schemes that aim to reduce noise in the data, thereby enhancing the detection of meaningful trends—whether a decline or, ideally, a conservation action-related increase. More generally, our RoC-related findings add to the growing interest in investigating how populations react to changing RoC of environmental drivers (Pinek et al. 2020, Albaladejo-Robles et al. 2023). This link might offer additional opportunities when forecasting population or community dynamics.

Finally, we address the relevance of time series length. Many studies call for longer time series (e.g. White 2019, Cusser et al. 2021). Our results at first sight seem to go against it. Time series length *per se* is positively correlated with the type of forecasted trend: time series shorter than 12-15 years are more likely associated with forecasted time trends (Appendix S1: Figures S7-S8), but this is mainly due to spurious trends in the short time series (Figure 3). Beyond 15 years, the likelihood of a forecasted constant mean increases, and mean(RoC) often approaches zero. Longer time series are always preferable, both for narrowing the uncertainty in mean(RoC) and var(RoC) and for ensuring robust analyses. However, this will often result in a forecasted constant mean for the best-fitting models. This observation aligns with results from retrospective trend analyses, indicating that steady means are common, and

detected declining and increasing trends are roughly balanced (Dornelas et al. 2023, and references therein). We should be wary of short time series because their fitted and forecasted dynamics more often than not are not trustworthy.

The ongoing biodiversity crisis demands both retrospective and forward-looking trend analyses. Working with typical ecological time-series data can be challenging, but overall a nuanced picture is emerging with respect to the details and extent of population-level and community-level changes. Near-term statistical forecasts are critical for informing stakeholders, as well as devising and monitoring conservation actions. We hope that our straightforward and robust approach will prove useful as a baseline for biodiversity trend forecasting. This baseline will hopefully soon – data permitting – be improved by more targeted and system-specific approaches. Yet despite their simplicity, we believe parsimonious forecasting approaches continue to offer valuable and actionable insights.

Acknowledgements: We thank Simon Egli for sharing his long-term macrofungal data.

Data availability: Most of the data that support the findings of this study are publicly available online and are referenced in the main text. See Appendix S1: Table S1 for further details.

Code availability: Code (Python), along with raw results files, will be uploaded to the zenodo repository after acceptance.

Competing interests: The authors declare no conflict of interest.

Author contributions: **C.B.** Conceptualization, Methodology, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **A.R.I.** Methodology, Formal analysis, Writing - Review & Editing.

References

Albaladejo-Robles, G., M. Böhm, and T. Newbold. 2023. Species life-history strategies affect population responses to temperature and land-cover changes. Global Change Biology 29:97–109.

Auger-Méthé, M., K. Newman, D. Cole, F. Empacher, R. Gryba, A. A. King, V. Leos-Barajas, J. Mills Flemming, A. Nielsen, G. Petris, and L. Thomas. 2021. A guide to state–space modeling of ecological time series. Ecological Monographs 91:e01470.

Boënnec, M., V. Dakos, and V. Devictor. 2024. Sources of confusion in global biodiversity trends. Oikos 2024:e10732.

Bozzuto, C., and A. R. Ives. 2024. Predictability of ecological and evolutionary dynamics in a changing world. Proceedings of the Royal Society B: Biological Sciences 291:20240980.

Breiman, L. 2001. Random Forests. Machine Learning 45:5–32.

Cha, S.-H. 2007. Comprehensive Survey on Distance/Similarity Measures Between Probability Density Functions. Int. J. Math. Model. Meth. Appl. Sci. 1.

Comte, L., J. Carvajal-Quintero, P. A. Tedesco, X. Giam, U. Brose, T. Erős, A. F. Filipe, et al. 2021. RivFishTIME: A global database of fish time-series to study global change ecology in riverine systems. Global Ecology and Biogeography 30:38–50.

Cusser, S., J. Helms IV, C. A. Bahlai, and N. M. Haddad. 2021. How long do population level field experiments need to be? Utilising data from the 40-year-old LTER network. Ecology Letters 24:1103–1111.

Cutler, D. R., T. C. Edwards Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random Forests for Classification in Ecology. Ecology 88:2783–2792.

Dietze, M., E. P. White, A. Abeyta, C. Boettiger, N. Bueno Watts, C. C. Carey, R. Chaplin-Kramer, et al. 2024. Near-term ecological forecasting for climate change action. Nature Climate Change 14:1236–1244.

Dirzo, R., H. S. Young, M. Galetti, G. Ceballos, N. J. B. Isaac, and B. Collen. 2014. Defaunation in the Anthropocene. Science 345:401–406.

Dornelas, M., L. H. Antão, A. E. Bates, V. Brambilla, J. M. Chase, C. F. Y. Chow, A. Fontrodona-Eslava, et al. 2025. BioTIME 2.0: Expanding and Improving a Database of Biodiversity Time Series. Global Ecology and Biogeography 34:e70003.

Dornelas, M., J. M. Chase, N. J. Gotelli, A. E. Magurran, B. J. McGill, L. H. Antão, S. A. Blowes, et al. 2023. Looking back on biodiversity change: lessons for the road ahead. Philosophical Transactions of the Royal Society B: Biological Sciences 378:20220199.

Dornelas, M., N. J. Gotelli, B. McGill, H. Shimadzu, F. Moyes, C. Sievers, and A. E. Magurran. 2014. Assemblage Time Series Reveal Biodiversity Change but Not Systematic Loss. Science 344:296–299.

Durbin, T. late J., and S. J. Koopman. 2012. Time Series Analysis by State Space Methods. Second Edition. Oxford University Press, Oxford, New York.

Egli, S., M. Peter, C. Buser, W. Stahel, F. Ayer. 2006. Mushroom picking does not impair future harvests – results of a long-term study in Switzerland. Biological Conservation 129:271–276.

Filippova, N., E. Rudykina, E. Zvyagina, A. Dobrynina, V. Lutovinova, and K. Popesku. 2024. Plot-based observations of terrestrial macrofungi in different forest types of boreal zone in West Siberia (2015-2024). https://www.gbif.org/dataset/7648036d-fdac-4ebe-8861-9a28134adae1.

GEO BON. 2025. Group on Earth Observations Biodiversity Observation Network (GEO BON). https://geobon.org.

Hacket-Pain, A., J. J. Foest, I. S. Pearse, J. M. LaMontagne, W. D. Koenig, G. Vacchiano, M. Bogdziewicz, et al. 2022. MASTREE+: Time-series of plant reproductive effort from six continents. Global Change Biology 28:3066–3082.

Hallmann, C. A., M. Sorg, E. Jongejans, H. Siepel, N. Hofland, H. Schwan, W. Stenmans, et al. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLOS ONE 12:e0185809.

Harris, D. J., S. D. Taylor, and E. P. White. 2018. Forecasting biodiversity in breeding birds using best practices. PeerJ 6:e4278.

Harvey, A. C., S. J. Koopman, and J. Penzer. 1997. Messy Time Series: A Unified Approach. STICERD - Econometrics Paper Series.

IPBES. 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Zenodo.

Johnson, T. F., A. P. Beckerman, D. Z. Childs, T. J. Webb, K. L. Evans, C. A. Griffiths, P. Capdevila, et al. 2024. Revealing uncertainty in the status of biodiversity change. Nature:1–7.

Lasky, J. R., M. B. Hooten, and P. B. Adler. 2020. What processes must we understand to forecast regional-scale population dynamics? Proceedings of the Royal Society B: Biological Sciences 287:20202219.

Lewis, A. S. L., W. M. Woelmer, H. L. Wander, D. W. Howard, J. W. Smith, R. P. McClure, M. E. Lofton, et al. 2022. Increased adoption of best practices in ecological forecasting enables comparisons of forecastability. Ecological Applications 32:e2500.

Loh, J., R. E. Green, T. Ricketts, J. Lamoreux, M. Jenkins, V. Kapos, and J. Randers. 2005. The Living Planet Index: using species population time series to track trends in biodiversity. Philosophical Transactions of the Royal Society B: Biological Sciences 360:289–295.

Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee. 2020. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2:56–67.

McGill, B. J., M. Dornelas, N. J. Gotelli, and A. E. Magurran. 2015. Fifteen forms of biodiversity trend in the Anthropocene. Trends in Ecology & Evolution 30:104–113.

Molnar, C. 2025. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Third edition. https://christophm.github.io/interpretable-ml-book/.

Myhrvold, N. P., E. Baldridge, B. Chan, D. Sivam, D. L. Freeman, and S. K. M. Ernest. 2015. An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. Ecology 96:3109–3109.

Newbold, T., L. N. Hudson, S. L. L. Hill, S. Contu, I. Lysenko, R. A. Senior, L. Börger, et al. 2015. Global effects of land use on local terrestrial biodiversity. Nature 520:45–50.

Olsson, F., C. C. Carey, C. Boettiger, G. Harrison, R. Ladwig, M. F. Lapeyrolerie, A. S. L. Lewis, et al. 2025. What can we learn from 100,000 freshwater forecasts? A synthesis from the NEON Ecological Forecasting Challenge. Ecological Applications 35:e70004.

Pennekamp, F., A. C. Iles, J. Garland, G. Brennan, U. Brose, U. Gaedke, U. Jacob, et al. 2019. The intrinsic predictability of ecological time series and its potential to guide forecasting. Ecological Monographs 89:e01359.

Pinek, L., I. Mansour, M. Lakovic, M. Ryo, and M. C. Rillig. 2020. Rate of environmental change across scales in ecology. Biological Reviews 95:1798–1811.

Pomati, F., J. B. Shurin, K. H. Andersen, C. Tellenbach, and A. D. Barton. 2020. Interacting Temperature, Nutrients and Zooplankton Grazing Control Phytoplankton Size-Abundance Relationships in Eight Swiss Lakes. Frontiers in Microbiology 10.

Prendergast, J., E. Bazeley-White, O. Smith, J. Lawton, P. Inchausti, D. Kidd, and S. Knight. 2010. The Global Population Dynamics Database.

Richter, D. 2023. Mycorrhizal Fungi of Native Red Pine Stands in the Forests of the Huron Mountains (1996-2015). Environmental Data Initiative. https://portal.edirepository.org/nis/mapbrowse?packageid=edi.1470.1.

Rosenberg, K. V., A. M. Dokter, P. J. Blancher, J. R. Sauer, A. C. Smith, P. A. Smith, J. C. Stanton, A. Panjabi, L. Helft, M. Parr, and P. P. Marra. 2019. Decline of the North American avifauna. Science 366:120–124.

Toszogyova, A., J. Smyčka, and D. Storch. 2024. Mathematical biases in the calculation of the Living Planet Index lead to overestimation of vertebrate population decline. Nature Communications 15:5295.

Tulloch, A. I. T., V. Hagger, and A. C. Greenville. 2020. Ecological forecasts to inform near-term management of threats to biodiversity. Global Change Biology 26:5816–5828.

UNEP-WCMC. 2024. State of the World’s Migratory Species Report. Cambridge, United Kingdom.

Urban, M. C., G. Bocedi, A. P. Hendry, J.-B. Mihoub, G. Pe’er, A. Singer, J. R. Bridle, et al. 2016. Improving the forecast for biodiversity under climate change. Science 353:aad8466.

Ward, E. J., E. E. Holmes, J. T. Thorson, and B. Collen. 2014. Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. Oikos 123:652–661.

Welti, E. A. R., D. E. Bowler, J. S. Sinclair, F. Altermatt, M. Álvarez-Cabria, G. Amatulli, D. G. Angeler, et al. 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.

Wiens, J. J., and K. E. Saban. 2025. Questioning the sixth mass extinction. Trends in Ecology & Evolution 40:375–384.

Wilson, E. O. 1992. The Diversity of Life. Harvard University Press.

Wood, K. A., R. A. Stillman, and G. M. Hilton. 2018. Conservation in a changing world needs predictive models. Animal Conservation 21:87–88.

WWF. 2024. Living Planet Report 2024 - A System in Peril. WWF, Gland, Switzerland.

Zylstra, E. R., L. Ries, N. Neupane, S. P. Saunders, M. I. Ramírez, E. Rendón-Salinas, K. S. Oberhauser, M. T. Farr, and E. F. Zipkin. 2021. Changes in climate drive recent monarch butterfly dynamics. Nature Ecology & Evolution 5:1441–1452.