Global Ecology and Biogeography

WILEY

Global Ecology and Biogeography A Journal of Macroecology

METHOD OPEN ACCESS

Environmental Predictability in Phylogenetic Comparative Analysis: How to Measure It and Does It Matter?

Ming Liu[1,2] | Louis Bell-Roberts[1] | Carlos A. Botero[3] | Charlie K. Cornwallis[4] | Stuart A. West[1]

[1]Department of Biology, University of Oxford, Oxford, UK | [2]Sir William Dunn School of Pathology, University of Oxford, Oxford, UK | [3]Department of Integrative Biology, University of Texas at Austin, Austin, Texas, USA | [4]Department of Biology, Lund University, Lund, Sweden

Correspondence: Ming Liu (ming.liu@path.ox.ac.uk)

Received: 26 November 2024 | **Revised:** 8 July 2025 | **Accepted:** 29 July 2025

Handling Editor: Andrew Rominger

Funding: This work was supported by the European Research Council (834164; ML, L.B.-R. and S.A.W.), Natural Motion and Ministry of Education of Taiwan (M.L.) and Knut and Alice Wallenberg Foundation (2018.0138; C.K.C.).

Keywords: climate change | climatic time series | cooperation | global climatic dataset | harsh environment | macroecology

ABSTRACT

Aim: Abiotic environmental conditions shape ecological and evolutionary processes, yet quantifying their influence on organisms remains challenging due to variation among metrics and their intercorrelations. This study evaluates the utility of temporal environmental predictability measures and assesses their explanatory power in phylogenetic comparative analyses.

Innovation: We systematically compare widely used metrics of predictability and explore their correlations with environmental means and variances in a global meteorological dataset. Using cooperative breeding birds as a case study, we assess the impact of including predictability metrics in phylogenetic comparative analyses. We demonstrate the consequences of choosing specific metrics and the trade-offs between increased data inclusion and model interpretability.

Main Conclusions: Predictability metrics, though intuitively meaningful, have been conceptualised and quantified with diverse approaches. We found that different measures of predictability can exhibit contrasting global patterns and strong correlations with other environmental quantities. Therefore, our findings caution against overloading statistical analyses with correlated predictors, highlighting the need for a thoughtful selection of environmental metrics to avoid spurious interpretations in ecological and evolutionary studies.

1 | Introduction

Variation in abiotic environmental conditions has been argued to be important for evolutionary biology, macroecology and conservation biology (Parmesan 2006; Wiens et al. 2010). For example, temperature and rainfall are key components of the environment that can drive eco-evolutionary dynamics, community stability and life history evolution (Orzack 1985; Drake 2005; Parmesan 2006; Barabás et al. 2018; Yamamichi et al. 2023). Biologists have increasingly used measures of such climatic variables from meteorological datasets to characterise the temporal dynamics of environments that species inhabit on a global scale, such as dormancy, trophic interactions, species diversity and conservation (Hawkins and Porter 2003; O'Connor et al. 2009; Araújo et al. 2011; Jetz and Rubenstein 2011; Cornwallis et al. 2017; Rubio de Casas et al. 2017). Understanding how climatic variation influences processes of adaptation and species population dynamics has become a more pressing issue with recent climate change (Loarie et al. 2009; Bathiany et al. 2018).

Among the terms commonly used to describe climate, temporal predictability refers to the extent to which climatic variables follow a non-random temporal structure. It has been used in phrases such as predictable environment, unpredictable

---

rainfall and high predictability (Fisher et al. 2013; Cornwallis et al. 2017; Fristoe et al. 2017; Griesser et al. 2017; Ringen et al. 2019; Firman et al. 2020; Martin et al. 2020; Diamant et al. 2021). While formal metrics of predictability often focus on autocorrelation or periodicity within a single variable, many organisms rely on consistent environmental cues, regardless of their specific statistical properties, to anticipate future conditions. For example, UK great tit populations use spring temperature to time breeding in anticipation of peak caterpillar availability (Charmantier et al. 2008). Red deer use rainfall to predict plant growth (Nussey et al. 2005). The termination of the dormant stage in freshwater invertebrate eggs is closely linked to temperature increment or a recent rain event (Forrest and Thomson 2011; Stenert et al. 2017). These examples highlight how organisms use temporal structure in the environment to make adaptive decisions, supporting the biological relevance of studying environmental predictability.

Predictability can be difficult to quantify. Most commonly, predictability refers to whether past observations are autocorrelated with current observations (Colwell 1974; Koenig and Knops 2000; Norden et al. 2007; Botero et al. 2014). In some cases, predictability is characterised by both seasonality (partitioned data variance) and autocorrelation in the residuals (e.g., Marshall and Burgess (2015)). Predictability has also sometimes been used interchangeably with variability (SD), where low predictability corresponds with high variability (Jetz and Rubenstein 2011). In other cases, measures of unpredictability have been defined by life history outcomes, such as offspring mortality (Caro et al. 2016) and rates of development (Brumbach et al. 2009). Consequently, the concept of predictability, while intuitive, has been studied using various metrics, with unclear relationships between them and uncertain impacts on conclusions in comparative studies.

In addition to predictability, the contributions of temperature and rainfall to species adaptations are often summarised across different moments of their temporal distribution. These moments usually include the average (mean) and the variation using standard deviations (SD) or sometimes the variance. These two quantities, average and SD, have been widely used in biological studies to capture differences in environmental conditions, such as temperature and rainfall and how much they fluctuate over time (Lawing and Polly 2011; Diamant et al. 2021). For example, it has been argued that environments with a high average temperature and a low/variable rainfall are relatively harsher environments for breeding birds and mammals (Cornwallis et al. 2017; Firman et al. 2020; Martin et al. 2020). To capture such harsh environments, do we need all climatic variables and moments of their temporal distributions, or is it potentially compromised by the correlation between each quantity? And more generally, is it problematic to include multiple climatic quantities into these phylogenetic comparative studies?

Here, we address these questions in three sections. First, we begin with a systematic comparison of the most frequently used predictability measures (Section 2). This section aims to illustrate what the conceptual similarities are between different predictability measures and how they differ from each other in their calculations. Second, we apply these measures to a standard global meteorological dataset (Section 3.1) to ask: how are climatic quantities distributed across the globe, and to what extent each measure correlates to each other on the scale of species distribution ranges. Third, we examine the explanatory power gained from adding multiple climatic quantities—the mean, standard deviation and predictability of both temperature and precipitation—in the context of phylogenetic comparative analysis (Sections 3.2–4.2). We reanalysed a published dataset on the evolution of cooperative breeding in birds as a case study (Cornwallis et al. 2017). We then return to the biological meaning of predictability in the discussion, and especially how it depends upon the perception of organisms and not just climatic variables.

2 | How to Measure Predictability?

Before comparing the difference between the commonly used predictabilities, we introduce their definition and underlying concepts. We compared three different methods for quantifying predictability: Colwell's predictability with fixed binning, Colwell's predictability with dynamic binning, and normalised spectral entropy. The first two methods were developed by Colwell (1974) and have been commonly used in phylogenetic comparative analysis over the last decade (Fisher et al. 2013; Cornwallis et al. 2017; Fristoe et al. 2017; Griesser et al. 2017; Ringen et al. 2019; Firman et al. 2020; Martin et al. 2020; Diamant et al. 2021). Both of Colwell's predictability measures are calculated by putting observations into data categories and using the Shannon information statistic, also known as information entropy, to calculate uncertainty across time and data categories. The third method is from spatial and movement ecology, where some recent papers have suggested using spectral entropy to calculate the periodicity of time series data (Zaccarelli et al. 2013; Riotte-Lambert and Matthiopoulos 2020; Morrison et al. 2021). We chose these three methods because of their popularity (for Colwell's measurements) and because of their potential methodological benefits (for normalised spectral entropy). As all methods apply the concept of information entropy to measure the degree of uncertainty, we first explain the concept of entropy, then describe the three methods in more detail, and finally use two example time series to illustrate the similarities and differences between each method.

2.1 | The Concept of Entropy in Information Theory

Entropy is one of the fundamental concepts in thermodynamics and it measures the degree of uncertainty. The most commonly used formulation was developed by Shannon (1948),

$$H(X) = \sum_{i}^{s} - p_i ln(p_i) / ln(s) \qquad (1)$$

where X is the data, i is the index in observation categories, p is the proportion of observations belonging to the focal category, and s is the number of categories. $H(X)$ is called normalised information entropy in information theory, and normalisation is added to facilitate comparison between different data. $H(X)$ is maximised when all p have equal values, and minimised when one p is 1 and all others are 0. In other words, larger values of H indicate greater uncertainty.

While all three methods use the same formulation of $H(X)$, the key difference between them is how categories are defined. For fixed and dynamic binning measures of Colwell's predictability, the continuous observational data are 'binned' or grouped into discrete quantity categories, so that two observed X values with small differences are viewed as being in the same category. In contrast, normalised spectral entropy uses frequencies of waves to define categories.

2.2 | Method 1: Colwell's Predictability With Fixed Binning

Colwell's predictability focuses on the yearly cycles by processing data from the same time of the year (e.g., month) into categories that capture different levels of intensity in the variable of interest (e.g., 1°C bins in mean temperature). Here, X (columns) refers to the time in the cycle and Y (rows) refers to the number of sampled cycles in which a given month exhibited environmental values of a certain level. We call this table the processed occurrence table and visualise it in the form of heat maps in Figure 1. For example, if we have monthly temperature data for 5 years with 60 observations, the processed occurrence table would have 12 columns, each represents a month (X) and a number of rows, each stands for a data category (Y). The sum of each column is therefore 5 (i.e., 5 observations per month). Subsequently, Colwell's predictability is defined as,

$$P_{\text{Colwell}} = 1 - (H(XY) - H(X)) \tag{2}$$

where $H(X)$ is the uncertainty over time to account for missing observations, $H(XY)$ is the uncertainty over all cells in the table. Note that there is a slight difference in Equation (2) versus Colwell's original paper, because we have already included $ln(s)$ within the entropy function.

Importantly, the final component is how each data category is defined in the processed occurrence table. For instance, one could define it as 1°C, or any arbitrary value, because they are interested in a certain thermal resolution of temperature time series data. For any given time series, we could maintain the same resolution of the data because the bin width of each data category does not change with the range of the input data. We refer to this binning method as 'fixed binning method'. The rationale behind Colwell's method was to provide a generic way to calculate predictability for both continuous and discrete variables, where the latter one cannot be treated with traditional time series analysis (i.e., nominal measures; Siegel 1956; Stevens 1958; Colwell 1974).

We generated two example time series to illustrate how methods vary (Figure 1a). Time series 1 was generated from a sine wave of yearly cycle with the addition of some white noise. The sine wave fluctuates between 1 and −1, while the white noise ranges between 0.25 and −0.25. Each datapoint represents a monthly average temperature for a 5-year period. For simplicity, we did not shift this temperature time series to more common values (e.g., 15°C) because shifting does not change the results. Time series 2 is time series 1 multiplied by 5 so the only difference between them is the scale of variation. The processed occurrence table for both time series, with 1°C width for each data category on the rows, ranging from −6°C to +6°C is given in Figure 1. This shows that time series 1 is concentrated on a few rows, whereas time series 2 is much more spread out. Consequently, Colwell's predictability with fixed binning method results in time series 1 being more predictable than time series 2 (0.924 vs. 0.803).

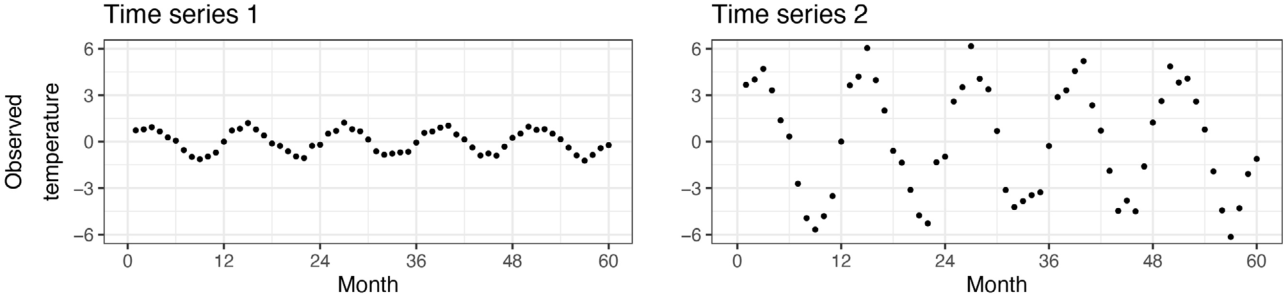
2.3 | Method 2: Colwell's Predictability With Dynamic Binning

An alternative approach for creating data categories is keeping the number of valid rows constant between input data, and letting the bin widths vary. That is, the width of a data category changes with the difference of maximal and minimal observation in a time series. The benefit of this approach is that it provides a cleaner measure of periodicity, that is independent of the average or variability (SD). The cost of this approach is that the calculation unit is not consistent across the entire globe—a difference in one degree Celsius will be a bigger jump between categories when the range of the input time series is smaller. Historically, the binning details have been suggested to cause differences in Colwell's indices and were discussed in the context of fix-state and scaled-state calculation (Beissinger 1986; Gan et al. 1991; Beissinger and Gibbs 1993). An example is shown in Figure 1c, where both time series are processed into tables of 25 rows. We call this method 'dynamic binning' because the definition of a single bin (i.e., data category) is dynamic and depends on the scale of the input data. Because data are compressed at the same level, both example time series have equal predictabilities of 0.691.

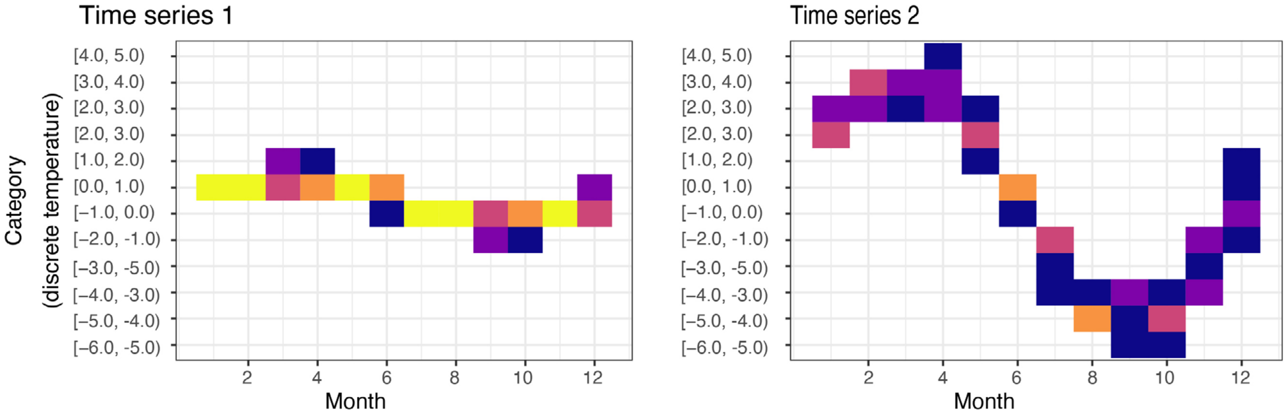
2.4 | Fundamental Components of Colwell's Predictability

As pointed out in Colwell's original paper (Colwell 1974), Colwell's predictability contains two fundamental components: constancy and contingency, where each can be referred to as a predictability measure. Constancy focuses on whether observations are always in the same data category, $C = 1 - H(Y)$, where $H(Y)$ is the uncertainty over data categories. On the other hand, contingency focuses on whether observations from the same month of different years are in the same data category while being distinct from other months, $M = H(X) + H(Y) - H(XY)$. Constancy and contingency have been analysed extensively in movement ecology because of their distinctive meaning, where contingency in particular has been closely linked to temporal structure and spectral entropy (see a recent review in Riotte-Lambert and Matthiopoulos (2020)). Despite this, most comparative studies have focused on the composite predictability metric, rather than examining individual components. We will examine these distinctions in Sections 3.1.2 and 5.1 and in the Supporting Information S1: Section 2.2.

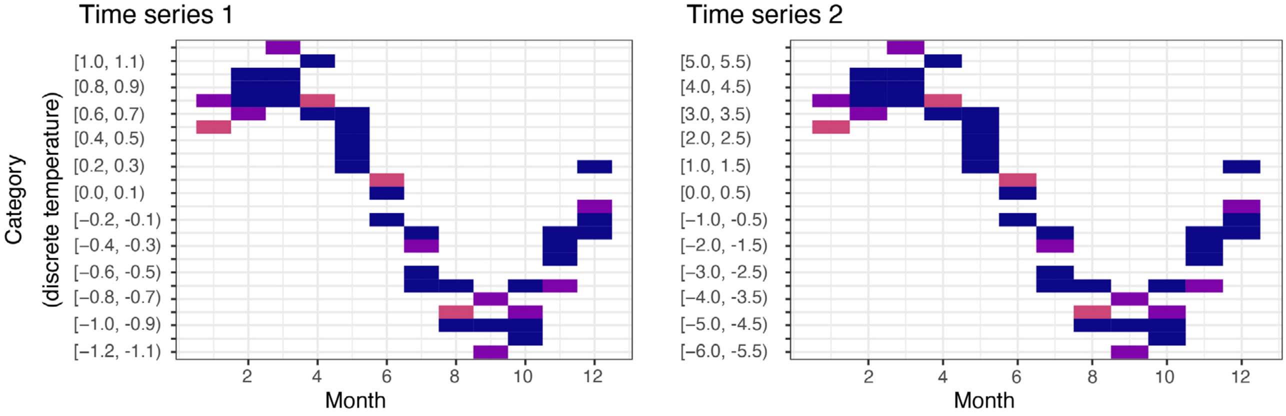
(a) Original data



(b) Colwell's P with fixed binning: constant resolution



(c) Colwell's P with dynamic binning: constant compression



(d) Normalised spectral entropy: no compression

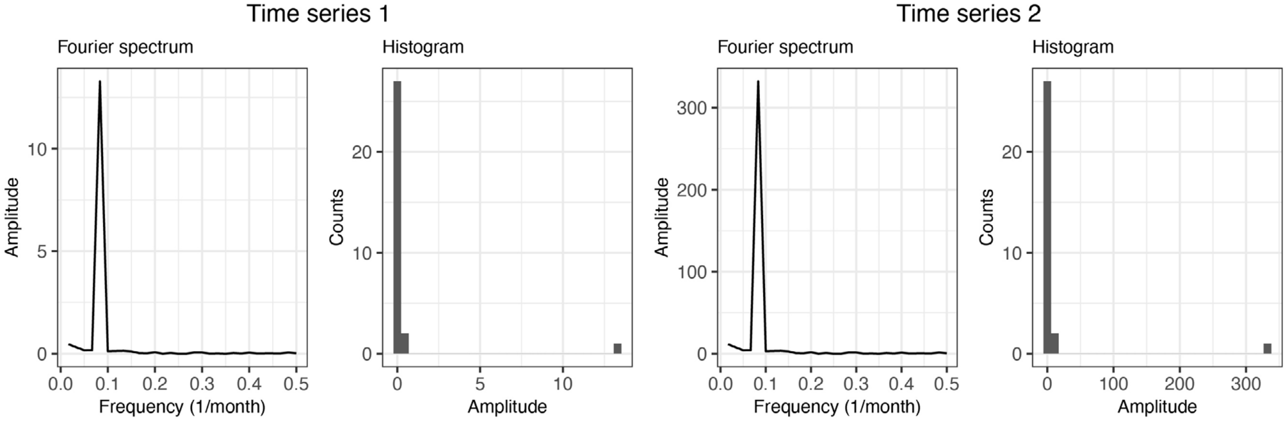


FIGURE 1 | Three different types of predictabilities estimated for two example time series. (a) The raw data from the two time series. Time series 1 is generated through a sine wave plus random numbers bounded between −0.25 and 0.25. Time series 2 is time series 1 multiplied by a factor of 5. (b–d) The processed data, or intermediate step, used in calculating three types of predictability. We set the bin width to 1 for fixed binning, and set 25 categories for dynamic binning. Brighter colour indicates there is a higher occurrence.

2.5 | Method 3: Normalised Spectral Entropy

Normalised spectral entropy is calculated from a Fourier spectrum produced by discrete Fourier transformation compared to the occurrence table in Colwell's methods. Discrete Fourier transformation is one of the standard methods in time series analysis (Shumway et al. 2000); it uses a discrete set of sine waves to describe the input data set: if the data are very noisy, such as white noise, then the resulting spectrum would have roughly equal amplitudes across all frequencies; if the data is very periodic, then the spectrum would have a few amplitude peaks and almost zero amplitudes in other frequencies. Normalised spectral entropy analyses the distribution of these amplitudes and treats each frequency as a category to calculate uncertainty (histogram in Figure 1d). Specifically,

$$P_{\text{SpcEntropy}} = 1 - H(X) = 1 - \frac{\sum_i^s - p_i \ln(p_i)}{\ln(s)} \tag{3}$$

where X is the Fourier spectrum, p_i is the amplitude of focal frequency divided by the sum of amplitudes, and s is the number of frequencies in the spectrum (equal to half the length of the time series; Zaccarelli et al. (2013)). Thus, predictability with normalised spectral entropy is quantifying the periodicity of input data: a white noise would produce a low predictability, and a pure sine wave would produce the highest predictability. See Supporting Information S1: Section 7.1 for graphic examples.

Normalised spectral entropy focuses on the uncertainty among frequencies, which is more independent of the variance of the data (compared to Method 2). One potential benefit of this is that spectral entropy looks into autocorrelation across multiple timescales such as days, months and years, while Colwell's calculation focuses on the correlation between years (12 months lag). In other words, the periodicity in multiple-year events like El Nino and La Nina can be captured by spectral entropy but not Colwell's measures. Another potential benefit is almost no information is lost during the transformation process in spectral entropy (except the phase of each sine wave) and it could be more sensitive to capture temporal signals (Stearns 1981; Beissinger and Gibbs 1993). Nevertheless, the downside of spectral entropy is the same as Method 2 that the calculation unit is not constant and varies with the scale of input data, and so far it has not been used in phylogenetic comparative analysis. Going back to our two example time series (Figure 1d), because the amplitude scales up evenly across all frequencies (at a factor of squared scaling so that time series 2 has 25 times larger amplitudes than time series 1), the normalised spectral entropy is identical between both time series: 0.796. We provide additional discussion between each method and the source code of Figure 1 in Supporting Information S1: Section 7.

3 | Using the Measures of Predictability

Next, we estimated the three measures of predictability for temperature and precipitation across the Globe from a meteorological dataset to ask: do certain places on earth consistently have a higher degree of predictability in temperature (Section 3.1); and how much does each climatic quantity correlate to each other in (Section 3.2).

3.1 | How Does Climatic Predictability Vary Across the World?

3.1.1 | Data Collection

We collected climatic data from the climate research unit (CRUTS v4.06; Harris et al. (2020)), by taking monthly measurements of temperature and precipitation (rainfall). We collected the data from January 1901 to December 2020, at a spatial resolution of 0.5° × 0.5° longitude-latitude grids (~55.5 km north-to-south, east-to-west varies with latitude). For each grid, we processed the climatic data into average, SD and the three predictabilities over the 1440 months ($N = 67{,}420$ grids).

3.1.2 | The Global Pattern

The three methods for measuring temperature predictability generated different global patterns (Figure 2a–c). Colwell's predictability with fixed binning (Method 1) is highest around the equator, but the other two methods (2&3) find that predictability is lowest in those regions (Figure 2a vs. Figure 2b,c). Because Method 1 is influenced by variability and periodicity, higher predictability measures can occur in less seasonal places because they show less variation. In contrast, Colwell's predictability with dynamic binning and normalised spectral entropy is focused on measuring periodicity (Methods 2 and 3). This is greatest in areas which show more consistent seasonality, which is in more temperate areas where there are warmer summers and colder winters. The opposite is also true for tropical places where temperature is less periodic because of less seasonality (see Supporting Information S1: Section 2.1 and Figure S1 for the quantitative correlation results).

The three methods applied to precipitation data also showed different global patterns (Figure 2d–f). All methods suggest rainfall is highly predictable in the Amazon rainforest, Arabian Desert and some regions of the Himalayas, but they disagree regarding Africa, Northern America and Eurasia. This inconsistency is likely because Colwell's predictability uses log-2-based binning in both Methods 1 and 2 to reduce the correlation between SD and mean, whereas Method 3 does not (Colwell 1974). We also provide the global maps for averages and SDs of temperature and precipitation in Figure 3 for comparison. In Supporting Information S1: Section 2.2, we provide further analysis on the global patterns of constancy and contingency, the two finer components of Colwell's predictability. We found that the differences between Methods 1 and 2 versus Method 3 are mostly due to constancy (Figures S2 and S3). In contrast, there is great similarity between the patterns of contingency and spectral entropy, but the quantities still vary with binning methods (Figures S2 and S3).

3.1.3 | Conclusion 1: Different Predictabilities Are Quantifying Different Properties

With temperature, we found that Colwell's predictability with fixed binning (Method 1) creates contrastingly different results than the other two methods, which appear to be because it is

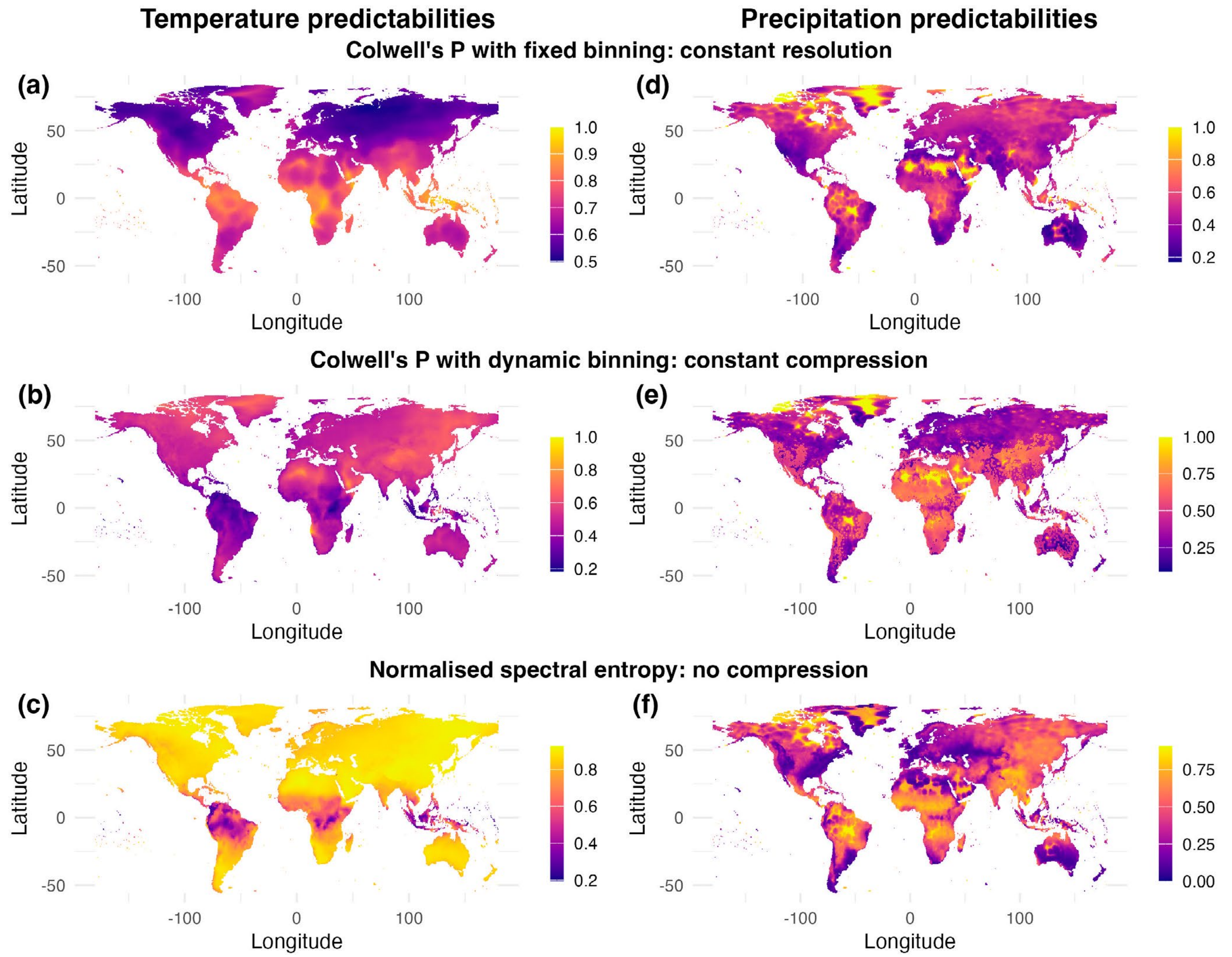


FIGURE 2 | Measuring climate predictability on a global scale. (a–c) The three predictabilities for temperature. Methodological details for binning in Colwell's predictability are described in Figure 1. (d–f) The three predictabilities for precipitation. Following previous literature and Colwell's original paper, the bins for precipitation data when calculating Colwell's predictability are on a log2 scale. Normalised spectral entropy, on the other hand, does not have this log2 transformation pre-process (i.e., calculation process is identical for temperature and precipitation in entropy method). For temperature, Method 1 (a: Fixed binning) shows higher predictabilities in tropical, seasonal places, whereas Methods 2 and 3 (b and c) show higher predictabilities in temperate, seasonal places. For precipitation, the different methods disagree with each other on where predictability is high or low (d–f).

capturing both the repeatability of oscillations and the overall tendency to oscillate at the same time. With precipitation, normalised spectral entropy had the opposite correlation to Colwell's predictabilities.

3.2 | Do Different Climatic Quantities Correlate at the Scale of Species Ranges?

With the knowledge that methods in quantifying predictability can generate a variety of global patterns, we analysed the correlation between predictability measures as well as mean and SD of climatic variables. This is a crucial step before moving to phylogenetic regression analysis; because high correlation leads to problems of collinearity in model interpretation.

3.2.1 | Data Collection: A Reanalysis

We used a global bird data set published in 2017 as a case study to test the usefulness of different measures of environmental variables (Cornwallis et al. 2017). The data contains 286 bird species for which there were estimates of multiple paternity, climatic variation and the breeding system (cooperative breeder or not). We re-extracted climate data for the study sites for each species from the global CRU dataset presented in the previous section. For simplicity, we assume species abundance is uniform within study sites, and the average was taken with equal weighting for each spatial grid at the unit of months. We examined how each method of predictability correlated with the average and SD. To increase the normality of data, we applied transformations to some of the variables before the analysis (also applied to Figure S2 for consistency).

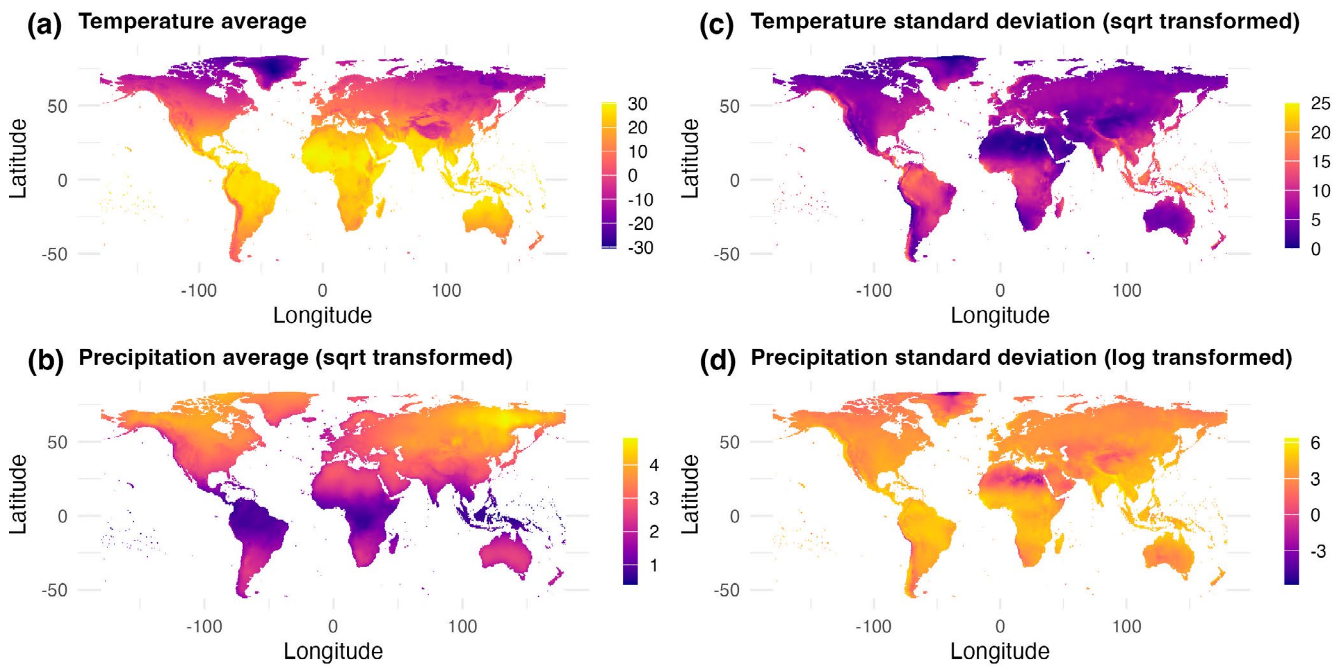


FIGURE 3 | Climate averages and SDs across the globe. (a) Average temperature. (b) Temperature SD. (c) Average precipitation. (d) Precipitation SD. All data are averaged across 1901–2020, and transformations were made to improve normality.

Supporting Information S1: Section 6 contains the Q-Q plots for various common transformations to show the transformations we applied are appropriate.

3.2.2 | Correlation Between Predictabilities and Other Quantities

We found very strong correlations between all temperature predictabilities, the average and the SD (Figure 4a–f). In highly seasonal regions (low average and high SD of temperature), Colwell's predictability with fixed binning (Method 1) has low predictability values, while the other two methods have high predictability values compared to less seasonal regions. These findings are consistent with the global pattern in Figure S4. Taken together, we found that predictabilities have overlapping properties with the average and the SD, providing limited extra information (see Section 5.3 for more explanation).

There were much weaker correlations for the precipitation predictability measures and the average and the SD (Figure 4g–l). The correlation coefficients were lower in the precipitation analysis than in the temperature analysis, which is consistent with the global analysis (Figure S4). However, unlike the global dataset (Figure S4) where the three predictabilities had inconsistent relationships with the average and SD of precipitation, we found positive correlations between all precipitation predictabilities in the cooperative breeding bird dataset. These results suggest that the predictabilities are not independent of the average or the SD of precipitation and highlight that the relationship in precipitation can vary across datasets that include different regions. In Supporting Information S1: Section 2.4, we analysed the correlation between means and SDs and found even stronger correlations than between predictabilities (Figure S5). We also conducted a similar analysis on precipitation coefficient of variation, an alternative variance measure (Figure S6). Together, these results highlight the potential problem of including both averages and SDs as independent covariates in regression models (due to collinearity) and justify the use of phylogenetic principal component analysis (phyloPCA) in the original paper as one common practice.

3.2.3 | Conclusion 2: Predictability Metrics Are Highly Correlated With Other Climatic Quantities

We found strong correlations between the average and the SD of climatic variables. Similarly, we found correlations between predictabilities and other quantities, regardless of the calculation method. However, comparing patterns in the global and regional datasets showed that only results with temperature predictability (Figure 4a–f, Figure S4a–f) were consistent, with precipitation results varying across datasets (Figure 4g–l; Figure S4g–l). This result highlights that climatic quantities are correlated, but that the direction of correlation is dataset dependent.

4 | Predictability in Phylogenetic Regression Analyses

The correlations between climatic quantities have been pointed out by many meteorologists and biologists (Walker and Cocks 1991; Kriticos et al. 2014; Dinnage 2023). This raises the question of how much extra statistical explanatory power is provided by including multiple climatic quantities in an analysis? We used phylogenetic logistic regression models to analyse the presence of cooperative breeding (response variable)

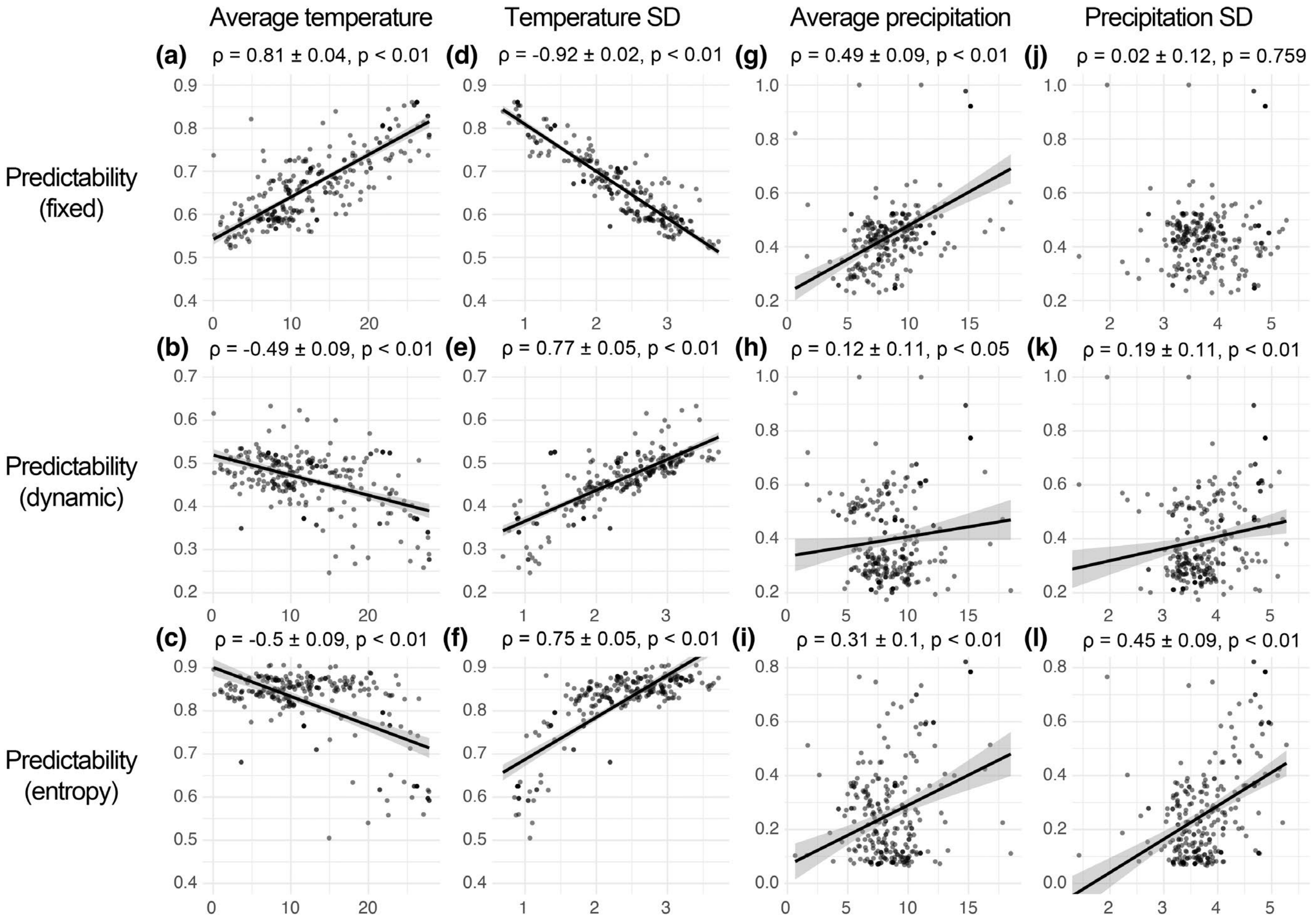


FIGURE 4 | Correlations between predictabilities and other climatic quantities across species. Against the (a–c) average temperature, (d–f) standard deviation of temperature, (g–i) average precipitation and (j–l) standard deviation of precipitation. In each panel, we report Pearson's correlation coefficient, ρ, with 95% confidence interval and p-value.

in relation to climatic variation (explanatory variables). We accounted for the non-independence of data that arises due to species phylogenetic relationships by modelling the phylogenetic correlation between residuals (Felsenstein 1985; Harvey and Pagel 1991).

In Section 4.1, we sequentially added raw climatic quantities as explanatory variables to quantify changes in explanatory power in phylogenetic regression models. In Section 4.2, we used the same approach, but included principal component scores as explanatory variables, which were constructed using different combinations of climatic quantities in a phylogenetic principal components analysis. These two different methods are both commonly used and each offers different advantages: biological interpretation can be clearer with separate estimates of each explanatory variable in multiple regressions, but results may be unreliable due to high multicollinearity, and vice versa for principal components. In addition to these two established methods, we also carried out a series of random forest analyses in Supporting Information S1: Section 5, but these results should be interpreted with caution as there is no established way to include phylogenetic information in such analyses.

4.1 | Do Predictability and SD Measures Increase the Explanatory Power of Regression Models (Without phyloPCA)?

4.1.1 | Statistical Methods and Interpretations

We used phylogenetic regressions to directly compare models with different predictors, using model selection procedures to evaluate competing models, and estimate the percentage of variance explained by different predictors (partial R^2). Specifically, we used maximum penalised likelihood estimation, a well-established algorithm in frequentist statistics, to find the best model (Anderson and Blair 1982; Tung Ho and Ané 2014). To evaluate model fitting, we calculated the Akaike information criteria (AIC) and overall R^2 (R^2_{lik}; Ives (2018)). AIC balances model fit and model complexity with lower AIC values indicating favoured models, while higher R^2 indicates that more variance in the input data is explained.

To examine the extent of multicollinearity in each regression model, we used variance inflation factors (VIF) from generalised linear models (GLM) with binary error distributions (logistic regression) that were equivalent to our phylogenetic models (because there is no readily available VIF tool for phylogenetic logistic regressions). VIF gives an indication as to how much the standard error in the estimated coefficient is inflated (Miles 2014). If a VIF is 4, it means the standard error is 2 times (square root of 4) larger than if that predictor had no correlation with other predictors. The degree of multicollinearity is subjective. Here we use the term moderate collinearity to describe cases where the VIF ranges between 2 and 4, and high collinearity where the VIF is larger than 4. We carried out the analyses using R v4.3.0, phylolm package v2.6.2, rr2 package v1.1.0 and car package v3.1.2 (Tung Ho and Ané 2014; Fox and Weisberg 2019; Ives 2018; R Core Team 2024).

Our first approach was to analyse the data with just the averages of temperature and precipitation, and then examine the extent to which our model is improved by adding standard deviations and predictabilities sequentially. To do this, we analysed five models: the **mean model** where cooperative breeding was predicted by average temperature and average precipitation (2 predictors); the **SD model** where SDs were added to the mean model (4 predictors); the **fixed predictability model** where Colwell's predictability with fixed binning was added to model 2 (6 predictors); the **dynamic predictability model** where Colwell's predictability with dynamic binning was added to model 2 (6 predictors); and the **entropy predictability model** where normalised spectral entropy was added to the SD model (6 predictors). We discuss the limitation of this model design in Section 5.3.

4.1.2 | Results

We found that adding predictability to regression models that already include average and SD increased AIC and reduced overall R^2 values (SD model compared to predictability models in Table 1). These results support our previous correlation analyses (Figure 4) and suggest that including predictability does not increase the power to explain cooperative breeding (over that provided by the mean). We also found that adding SDs improved the model, but only by a small ΔAIC (slightly greater than 2) and only by a 3% increase to R^2. This small improvement aligns with the correlation analysis of averages and SDs (Figure S5). Thus, there is a limited gain in explaining cooperative breeding from adding variability (SD) measurements to a model that already contains the mean. The summaries of all five regression models are in Supporting Information S1: Section 3.1. We also performed an alternative analysis, replacing precipitation SD with the coefficient of variation (CV) in Supporting Information S1: Section 3.3, because CV is less correlated with average precipitation than SD. We found adding CV also leads to a marginal improvement but adding predictability on top of CV leads to mixed results.

Multicollinearity was moderate to strong in all models apart from the mean model (Table 2). In particular, the SD model had moderate levels of collinearity between all predictors; the fixed predictability model had four out of six predictors with strong collinearities, and the dynamic and entropy predictability models both had one predictor with strong collinearity. These results suggest that the evaluated model estimates between climate and cooperative breeding, especially all predictability models, may be jeopardised by including multiple correlated climate summary statistics.

4.1.3 | Robustness Check Through Variable Importance Analysis

Our above approach examined the explanatory power of gradually adding in higher moments of the climatic variables (first

TABLE 1 | Summary of model fitting for each phylogenetic regression model analysing the cooperative breeding bird data from Cornwallis et al. (2017).

Model type	Mean Model (2 predictors)	SD Model (4 predictors)	Predictability (fixed) Model (6 predictors)	Predictability (dynamic) Model (6 predictors)	Predictability (entropy) Model (6 predictors)
AIC	192.321	190.065	200.722	204.988	202.598
Overall R^2	0.348	0.378	0.346	0.325	0.337

TABLE 2 | Variance inflation factor analysis of each phylogenetic regression model.

Predictor	Mean Model	SD Model	Predictability (fixed)	Predictability (dynamic)	Predictability (entropy)
Temperature average	1.221	2.441	2.691	2.969	3.07
Precipitation average	1.221	3.211	7.113	3.251	3.367
Temperature SD	NA	2.98	5.917	4.754	5.338
Precipitation SD	NA	3.271	6.014	3.399	3.715
Temperature predictability	NA	NA	6.453	2.439	3.202
Precipitation predictability	NA	NA	2.321	1.489	1.789

SD and then predictability). As an alternative, we performed a relative variable importance analysis. In this approach, a series of regression models are built with each possible combination of climatic quantities, and the differences in the model AICs are used to calculate the sum of the weights for each variable. The advantage of this approach is that there are no prior assumptions about the hierarchy between climatic quantities (we don't start with the mean and work up to SD and then predictability). However, the disadvantage is that multicollinearity could be high in many models, as shown in Table 2.

We repeated three equivalent relative variable importance analyses, one for each predictability, and found the most important climatic quantities are very volatile (Supporting Information S1: Section 3.2). Averages and SDs are the most important factors in dynamic binning and spectral entropy, but predictabilities are the most important factors in the fixed binning method (Tables S2–S4). Furthermore, many of the weights are similar to each other, probably due to multicollinearity, which makes it difficult to estimate the importance of climatic quantities. Overall, it is hard to disentangle from these analyses which is the ‘best’ single variable to include in phylogenetic multiple regression models.

4.2 | Does Including Predictabilities and SDs in Phylogenetic PCA Analyses Increase the Explanatory Power of Principal Components in Regression Models?

4.2.1 | Statistical Methods and Interpretations

We analysed the same dataset for cooperative breeding in birds, but with principal components (PCs) as explanatory variables. We used phylogenetic principal component analysis (phyloPCA; phytools package v2.0; Revell (2024)) to find the best linear combination of climatic quantities. Following previous work (Cornwallis et al. 2017), we z-transformed all quantities to standardise their scales and avoid biased estimations on PC weights (mean = 0 and SD = 1). We used the first two PCs as explanatory variables in phylogenetic logistic regression models: each PC axis captures a distinct aspect of climatic data, with PC1 representing the greatest variance and PC2 the next. As in the previous section, we analysed five regression models where PC1 and PC2 were generated with the same sets of explanatory variables: the averages, the averages and SDs and the three different predictability estimates. We call these models the **mean-PCA model**, **SD-PCA model** and **predictability-PCA models**. We ask if the PCs with more quantities better explain the occurrence of cooperative breeding in birds, by comparing the model AICs and overall R^2s.

4.2.2 | Results

We found that adding predictabilities to PCAs resulted in a marginally better or worse regression model fits (SD-PCA model compared to predictability-PCA models in Table 3). Only the fixed predictability-PCA model had a reduced AIC and increased overall R^2, whereas the dynamic predictability-PCA model is much worse than all other models. This result highlights that the binning method can alter results and supports conclusion 3 that adding predictabilities to the analysis does not improve models. Furthermore, by comparing the mean-PCA model and SD-PCA model, we found that AIC was only marginally significantly reduced and the R^2 was only improved by 1%. Thus, adding SD had a very limited effect on explaining cooperative breeding (see Figure S7 for the PC composition in biplots).

4.2.3 | Conclusion 3: Adding Some Predictability Measures Can Reduce Model Fit; Adding SDs Can Lead to Small Improvements

When examining the occurrence of cooperative breeding, we found that relative to a model that already contained the mean, adding predictabilities does not improve the model in most cases and can make it worse; and SDs only slightly improved models and increased multicollinearity in multiple regression analyses. Together, the results here suggest there is a high overlap in the amount of variation explained in cooperative breeding by mean, SD and predictability, which agrees with the correlation analysis in conclusion 2. In line with the findings above, our random forest analysis suggests that all climatic quantities have similar importance and are clustered into a small number of groups (Supporting Information S1: Section 5.3; Figure S10). These results further indicate the difficulty in discerning the single most important climatic property that associates with the evolution of cooperative breeding.

5 | Discussion

We found that predictability can be measured in different ways, but that these were correlated with other climate variables; and so did not significantly increase the explanatory power of analyses. First, different predictability measures give different and

TABLE 3 | Summary of model fitting for all PCA-processed regression models analysing the cooperative breeding bird data from Cornwallis et al. (2017).

Model type	Mean-PCA Model (2 predictors)	SD-PCA Model (2 predictors)	Predictability-PCA (fixed) Model (2 predictors)	Predictability-PCA (dynamic) Model (2 predictors)	Predictability-PCA (entropy) Model (2 predictors)
AIC	192.377	190.368	189.459	209.132	191.978
Overall R^2	0.348	0.357	0.362	0.265	0.350

potentially opposite results (Figure 2). This finding highlights the importance of considering binning methods (fixed, dynamic, or other ways) when interpreting results, an issue that has been largely neglected in the application of Colwell's metrics. Second, the average, SD and predictabilities were all highly correlated for temperature and to a lesser extent for precipitation (Figure 4, Figure S4), indicating potential issues in interpreting statistical models with multiple climatic quantities. Third, there is limited gain from adding multiple climatic quantities to phylogenetic multiple regression or PC analyses.

5.1 | Different Methodologies of Predictability Quantify Different Things

We found that different measures of predictability can give very different results, and that the best measure to use could depend upon the aims of the study. Colwell's predictability with fixed binning (Method 1) suggests that more tropical climates are more predictable, because they show smaller variation. Colwell's predictability with dynamic binning and normalised spectral entropy (Method 2 and 3) suggest that more temperate climates are more predictable because they show stronger seasonality (larger variation, but seasonally predictable). The predictability terminology can therefore be confusing, and the choice of method will likely depend on whether periodicity, variability or both are important to investigate. In particular, Colwell's predictability with dynamic binning is measuring periodicity across years, and normalised spectral entropy is measuring periodicity across a broad range of time scales (Figure 1c,d). Consequently, it might be simpler to just refer to these two metrics as ‘periodicity’ or ‘cycle repeatability’. On the contrary, Colwell's predictability with fixed binning is measuring periodicity rescaled by variability (Figure 1b), thus could be called as ‘rescaled periodicity’ or ‘rescaled cycle repeatability’.

Continuing on the difference between fixed and dynamic binning, past literature from the 1980s and 1990s has pointed out Colwell's predictability is sensitive to how data is categorised (Stearns 1981; Beissinger 1986; Gan et al. 1991; Beissinger and Gibbs 1993). Among this literature, Gan et al. (1991) created two binning methods that are very similar to ours, fixed-states and scaled-states predictabilities. However, there are key differences: our fixed binning uses uniform bin widths, unlike their arbitrary ranges; our dynamic binning creates a constant number of bins based on the range of data, while theirs scales with the mean. This difference, as well as the difference in sample size, might lead to an opposite correlation being found between our study and their study (Supporting Information S1: Section 2.3 vs. Beissinger and Gibbs (1993)). Our binning methods were adopted from the original study (fixed; Cornwallis et al. 2017), and the default method in the hydrostat package (dynamic). Unfortunately, the binning method is not described in other comparative studies analysing climatic quantities and biological traits. Our finding aligns with previous literature, suggesting this is a decisive detail when using Colwell's predictability and should be carefully treated to avoid accidentally making misleading predictability measures.

Colwell's method can also be broken down into two components that capture different aspects of predictability: constancy and contingency. This examines how environments can be predictable because they never change (high constancy), or because they are easy to anticipate (high contingency; Supporting Information S1: Section 2.2, Figure S3). In particular, we found contingencies of both binning methods match well with normalised spectral entropy, suggesting they might have similar biological interpretations. The utility of different measures could also depend upon the geographical area examined. For example, if examining tropical organisms, then constancy could be the most useful metric because it captures differences among relatively stable habitats (low seasonality). In contrast, if temperate organisms are examined, then contingency could be more useful because it captures autocorrelations among relatively variable habitats (high seasonality). Finally, if examining across the entire globe, then it could be useful to examine predictability due to both constancy and contingency.

5.2 | Predictability for Real Organisms

An important biological question is: what is predictability from an organism's perspective? It is likely that predictability is a species-explicit and time-explicit property as organisms observe their surrounding environments in different durations or different periods of the same time frame, and as different organisms could live on drastically different time scales with different life histories (Salguero-Gómez et al. 2016; Van de Walle et al. 2023). Moreover, it is possible that meteorological measures of predictability relate very poorly to an organism's ability to predict the environment because of the time resolution of these climatic data. At one extreme, consider a case where the meteorological data suggest that environmental conditions show low predictability or periodicity, but the organisms can use finer environmental cues to predict future environmental conditions. In this case, the environment would be highly predictable to that organism (Botero et al. 2015). At the other extreme, if organisms cannot predict environmental variation on the scale of their lifetime, the environment could be less predictable than suggested by meteorological measures. Altogether, the term' predictability' in terms of how it matters, is a term that implies intentionality from an organismal perspective—this can be something very different from the entropy measures of periodicity or other aspects of predictability (Ghoul et al. 2014). Empirical studies could address this by carefully considering the temporal and spatial resolutions that are biologically meaningful for their focal species at the earliest stages of study design.

5.3 | Explanatory Power of Climatic Quantities

Our results raise the issue of how much additional explanatory power different quantities of temperature and precipitation may add to phylogenetic studies. The high correlations between means, SDs and predictabilities we found coincide with a recent study on a similar global meteorological database (WorldClim), which showed that 5 synthetic variables are sufficient to capture most of the variation in the entire dataset of 19 variables (Dinnage 2023). Similarly, these correlations have been noticed for a long time in the biogeography and other literature (Birks 1996; Heikkinen 1996; Guisan and Zimmermann 2000; Pendergrass et al. 2017; Harp and Horton 2023). Indeed, high levels of interdependence between explanatory variables affect the interpretability of multiple regression models and do not violate the assumptions behind multiple regression (Morrissey and Ruxton 2018). Nevertheless, in our single case study with reanalysing published dataset, we found VIF coefficients quickly increase with including more climatic quantities. These results suggest using multiple regression to figure out the key contributing climatic factor to biological adaptation can be challenging.

It is important to note that we are not discouraging the use of predictability, but rather highlighting the complex nature of different predictability measures, and the potential gaps between predictability and biological processes. In our statistical analyses (Section 4), we employed the perspective that mean is a more fundamental property than SD, and both are more fundamental than predictability. This perspective was derived from the moments of distribution (as well as the ease of measurement). Another reason is the mean is less versatile than variability, which can be measured in SD, CV and variance; and predictability is an even more versatile metric (Figure 2). Indeed, if a biologist can justify that a certain type of predictability should be the primary contributor to a biological process, they might take a different plan to set up the statistical models and reach a slightly different conclusion. However, changing the sequence of adding different properties of a single climatic variable would not change the conclusion: they have a high overlap to which they can explain the variance of cooperative breeding (Conclusion 3). In addition, although our analyses employed three distinct entropy metrics and found these entropies are insufficient in adding more information to the statistical analyses, it is possible that other entropy metrics would work, or the ones we used can be important in other datasets.

5.4 | How Should We Move Forward?

The most useful methodology can depend on the hypothesis being tested, or if an analysis is more exploratory. With a specific hypothesis, we suggest constructing phylogenetic regression models with a minimal number of explanatory variables, and then compare the hypothesis model, which contains the variable(s) of interest, against alternative models. For example, if a study aims to test whether the periodicity effects of yearly environmental cycles are fundamental for a behavioural strategy, then the variable of interest could be Colwell's predictabilities (including constancy and contingency; Section 5.1). In contrast, if a macroecology study is designed to test the climatic variability hypothesis, then the variable of interest could be the SD or maximal difference in temperature (Janzen 1967; Quintero and Wiens 2013; Gutiérrez-Pesquera et al. 2016; Polato et al. 2018; Chiono and Paul 2023). In addition, if variation in precipitation is more biologically relevant than temperature variability for cooperative breeding, then comparing models that include only precipitation SD may help clarify its independent effect (Rubenstein and Lovette 2007; Jetz and Rubenstein 2011; Cornwallis et al. 2017). Alternatively, if a study aims to test the temporal scale effects, such as diurnal thermal range and seasonal thermal range being hypothesised to have opposite effects on natural selection—because the duration of environmental variation can change the ecological dynamics—then different aspects of thermal variability could be crucial (Chan et al. 2016; Liu et al. 2021). If the regression model with the variable(s) of interest is significantly better than alternative models, which should have similar numbers of comparable explanatory variables, then it would support the focal hypothesis.

If there is no specific hypothesis to test, regression models with PCs could be a better option; though it is important to be cautious about the wording when referring to PCs. Phylogenetic PCA retains great interpretability even when the model contains numerous climatic quantities (Botero et al. 2014; Fristoe et al. 2017; Griesser et al. 2017; Firman et al. 2020). For instance, it could tell how much variance is explained by each PC, and with acknowledging the background assumptions, one could use the PCA loadings to figure out the key factors (Chong et al. 2018). However, one should be careful when inferring the biological meaning of PCs and avoid labelling these PCs with terms like ‘predictable environment’. This is because the process of making PCs does not contain any biological insight. Another way to deal with multicollinearity is through machine learning tools, such as random forest models. These methods usually do not make strong prior assumptions and are excellent in coping with non-linear relationships, as well as correlations between predictors (sometimes referred to as features). Nevertheless, there are currently no available tools or established pipelines to construct a phylogenetic random forest model. Our attempts in Supporting Information S1: Section 5 used one categorical trait to represent phylogenetic groups, while alternative methods include but are not limited to the eigenvectors from the phylogenetic variance–covariance matrix and the PCs of the phylogenetic distance matrix. Yet, these methods all create additional predictors. In conclusion, while giving explicit and clear definitions of the meaning of predictable climatic conditions in each study is the best way to facilitate communication, our findings also highlight the need of developing techniques to examine predictability from the organisms' perspective.

Author Contributions

M.L. and S.A.W. conceived the ideas and designed methodology. M.L. and C.K.C. collected the data. M.L. and L.B.-R. analysed the data. M.L. and S.A.W. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Acknowledgements

We thank, Samuel Gascoigne, Andrew Wood, James McCulloch, Juliet Turner and Robert Colwell for their input at various stages of this project; the European Research Council (834164; ML, L.B.-R. & S.A.W.); NaturalMotion and Ministry of Education of Taiwan (M.L.); Knut and Alice Wallenberg Foundation (2018.0138; C.K.C.).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The codes and data used in this study are available at: https://doi.org/10.5061/dryad.gtht76j02.

References

Anderson, J. A., and V. Blair. 1982. “Penalized Maximum Likelihood Estimation in Logistic Regression and Discrimination.” *Biometrika* 69: 123–136.

Araújo, M. B., D. Alagador, M. Cabeza, D. Nogués-Bravo, and W. Thuiller. 2011. “Climate Change Threatens European Conservation Areas.” *Ecology Letters* 14: 484–492.

Barabás, G., R. D'Andrea, and S. M. Stump. 2018. “Chesson's Coexistence Theory.” *Ecological Monographs* 88: 277–303.

Bathiany, S., V. Dakos, M. Scheffer, and T. M. Lenton. 2018. “Climate Models Predict Increasing Temperature Variability in Poor Countries.” *Science Advances* 4: eaar5809.

Beissinger, S. 1986. “Demography, Environmental Uncertainty, and the Evolution of Mate Desertion in the Snail Kite.” *Ecology* 67: 1445–1459.

Beissinger, S. R., and J. P. Gibbs. 1993. *Are Variable Environments Stochastic? A Review of Methods to Quantify Environmental Predictability*, 132–146. Springer.

Birks, H. J. B. 1996. “Statistical Approaches to Interpreting Diversity Patterns in the Norwegian Mountain Flora.” *Ecography* 19: 332–340.

Botero, C. A., R. Dor, C. M. McCain, and R. J. Safran. 2014. “Environmental Harshness Is Positively Correlated With Intraspecific Divergence in Mammals and Birds.” *Molecular Ecology* 23: 259–268.

Botero, C. A., F. J. Weissing, J. Wright, and D. R. Rubenstein. 2015. “Evolutionary Tipping Points in the Capacity to Adapt to Environmental Change.” *Proceedings of the National Academy of Sciences* 112: 184–189.

Brumbach, B. H., A. J. Figueredo, and B. J. Ellis. 2009. “Effects of Harsh and Unpredictable Environments in Adolescence on Development of Life History Strategies.” *Human Nature* 20: 25–51.

Caro, S. M., A. S. Griffin, C. A. Hinde, and S. A. West. 2016. “Unpredictable Environments Lead to the Evolution of Parental Neglect in Birds.” *Nature Communications* 7: 10985.

Chan, W.-P., I.-C. Chen, R. K. Colwell, W.-C. Liu, C.-y. Huang, and S.-F. Shen. 2016. “Seasonal and Daily Climate Variation Have Opposite Effects on Species Elevational Range Size.” *Science* 351: 1437–1439.

Charmantier, A., R. H. McCleery, L. R. Cole, C. Perrins, L. E. B. Kruuk, and B. C. Sheldon. 2008. “Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population.” *Science* 320: 800–803.

Chiono, A., and J. R. Paul. 2023. “The Climatic Variability Hypothesis and Trade-Offs in Thermal Performance in Coastal and Inland Populations of *Mimulus guttatus*.” *Evolution* 77: 870–880.

Chong, V. K., H. F. Fung, and J. R. Stinchcombe. 2018. “A Note on Measuring Natural Selection on Principal Component Scores.” *Evolution Letters* 2: 272–280.

Colwell, R. K. 1974. “Predictability, Constancy, and Contingency of Periodic Phenomena.” *Ecology* 55: 1148–1153.

Cornwallis, C. K., C. A. Botero, D. R. Rubenstein, P. A. Downing, S. A. West, and A. S. Griffin. 2017. “Cooperation Facilitates the Colonization of Harsh Environments.” *Nature Ecology & Evolution* 1: 0057.

Diamant, E. S., J. J. Falk, and D. R. Rubenstein. 2021. “Male-Like Female Morphs in Hummingbirds: The Evolution of a Widespread Sex-Limited Plumage Polymorphism.” *Proceedings of the Royal Society B: Biological Sciences* 288: 20203004.

Dinnage, R. 2023. “How Many Variables Does Wordclim Have, Really? Generative A.I. Unravels the Intrinsic Dimension of Bioclimatic Variables.” bioRxiv. 2023.2006.2012.544623.

Drake, J. M. 2005. “Population Effects of Increased Climate Variation.” *Proceedings of the Royal Society B: Biological Sciences* 272: 1823–1827.

Felsenstein, J. 1985. “Phylogenies and the Comparative Method.” *American Naturalist* 125: 1–15.

Firman, R. C., D. R. Rubenstein, J. M. Moran, K. C. Rowe, and B. A. Buzatto. 2020. “Extreme and Variable Climatic Conditions Drive the Evolution of Sociality in Australian Rodents.” *Current Biology* 30: 691–697.

Fisher, D. O., C. R. Dickman, M. E. Jones, and S. P. Blomberg. 2013. “Sperm Competition Drives the Evolution of Suicidal Reproduction in Mammals.” *Proceedings of the National Academy of Sciences* 110: 17910–17914.

Forrest, J. R. K., and J. D. Thomson. 2011. “An Examination of Synchrony Between Insect Emergence and Flowering in Rocky Mountain Meadows.” *Ecological Monographs* 81: 469–491.

Fox, J., and S. Weisberg. 2019. *An R Companion to Applied Regression*. 3rd ed. Sage.

Fristoe, T. S., A. N. Iwaniuk, and C. A. Botero. 2017. “Big Brains Stabilize Populations and Facilitate Colonization of Variable Habitats in Birds.” *Nature Ecology & Evolution* 1: 1706–1715.

Gan, K. C., T. A. McMahon, and B. L. Finlayson. 1991. “Analysis of Periodicity in Streamflow and Rainfall Data by Colwell's Indices.” *Journal of Hydrology* 123: 105–118.

Ghoul, M., A. S. Griffin, and S. A. West. 2014. “Toward an Evolutionary Definition of Cheating.” *Evolution* 68: 318–331.

Griesser, M., S. M. Drobniak, S. Nakagawa, and C. A. Botero. 2017. “Family Living Sets the Stage for Cooperative Breeding and Ecological Resilience in Birds.” *PLoS Biology* 15: e2000483.

Guisan, A., and N. E. Zimmermann. 2000. “Predictive Habitat Distribution Models in Ecology.” *Ecological Modelling* 135: 147–186.

Gutiérrez-Pesquera, L. M., M. Tejedo, M. Á. Olalla-Tárraga, H. Duarte, A. Nicieza, and M. Solé. 2016. “Testing the Climate Variability Hypothesis in Thermal Tolerance Limits of Tropical and Temperate Tadpoles.” *Journal of Biogeography* 43: 1166–1178.

Harp, R. D., and D. E. Horton. 2023. “Observed Changes in Interannual Precipitation Variability in the United States.” *Geophysical Research Letters* 50: e2023GL104533.

Harris, I., T. J. Osborn, P. Jones, and D. Lister. 2020. “Version 4 of the CRU TS Monthly High-Resolution Gridded Multivariate Climate Dataset.” *Scientific Data* 7: 1–18.

Harvey, P. H., and M. D. Pagel. 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press.

Hawkins, B. A., and E. E. Porter. 2003. “Water–Energy Balance and the Geographic Pattern of Species Richness of Western Palearctic Butterflies.” *Ecological Entomology* 28: 678–686.

Heikkinen, R. K. 1996. “Predicting Patterns of Vascular Plant Species Richness With Composite Variables: A Meso-Scale Study in Finnish Lapland.” *Vegetatio* 126: 151–165.

Ives, A. R. 2018. “R2s for Correlated Data: Phylogenetic Models, LMMs, and GLMMs.” *Systematic Biology* 68: 234–251.

Janzen, D. H. 1967. “Why Mountain Passes Are Higher in the Tropics.” *American Naturalist* 101: 233–249.

Jetz, W., and D. R. Rubenstein. 2011. “Environmental Uncertainty and the Global Biogeography of Cooperative Breeding in Birds.” *Current Biology* 21: 72–78.

Koenig, W. D., and J. M. H. Knops. 2000. “Patterns of Annual Seed Production by Northern Hemisphere Trees: A Global Perspective.” *American Naturalist* 155: 59–69.

Kriticos, D. J., V. Jarošik, and N. Ota. 2014. “Extending the Suite of Bioclim Variables: A Proposed Registry System and Case Study Using Principal Components Analysis.” *Methods in Ecology and Evolution* 5: 956–960.

Lawing, A. M., and P. D. Polly. 2011. “Pleistocene Climate, Phylogeny, and Climate Envelope Models: An Integrative Approach to Better Understand Species' Response to Climate Change.” *PLoS One* 6: e28554.

Liu, M., D. R. Rubenstein, S. A. Cheong, and S.-F. Shen. 2021. “Antagonistic Effects of Long- and Short-Term Environmental Variation on Species Coexistence." *Proceedings of the Royal Society B: Biological Sciences* 288: 20211491.

Loarie, S. R., P. B. Duffy, H. Hamilton, G. P. Asner, C. B. Field, and D. D. Ackerly. 2009. "The Velocity of Climate Change." *Nature* 462: 1052–1055.

Marshall, D. J., and S. C. Burgess. 2015. "Deconstructing Environmental Predictability: Seasonality, Environmental Colour and the Biogeography of Marine Life Histories." *Ecology Letters* 18: 174–181.

Martin, J., E. Ringen, P. Duda, and A. Jaeggi. 2020. "Harsh Environments Promote Alloparental Care Across Human Societies." *Proceedings of the Royal Society B* 287: 20200758.

Miles, J. 2014. "Tolerance and Variance Inflation Factor." In *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd.

Morrison, T. A., J. A. Merkle, J. G. C. Hopcraft, et al. 2021. "Drivers of Site Fidelity in Ungulates." *Journal of Animal Ecology* 90: 955–966.

Morrissey, M. B., and G. D. Ruxton. 2018. "Multiple Regression Is Not Multiple Regressions: The Meaning of Multiple Regression and the Non-Problem of Collinearity." *Philosophy, Theory, and Practice in Biology* 10, no. 3.

Norden, N., J. Chave, P. Belbenoit, et al. 2007. "Mast Fruiting Is a Frequent Strategy in Woody Species of Eastern South America." *PLoS One* 2: e1079.

Nussey, D. H., T. H. Clutton-Brock, D. A. Elston, S. D. Albon, and E. B. K. Loeske. 2005. "Phenotypic Plasticity in a Maternal Trait in Red Deer." *Journal of Animal Ecology* 74: 387–396.

O'Connor, M. I., M. F. Piehler, D. M. Leech, A. Anton, and J. F. Bruno. 2009. "Warming and Resource Availability Shift Food Web Structure and Metabolism." *PLoS Biology* 7: e1000178.

Orzack, S. H. 1985. "Population Dynamics in Variable Environments. V. The Genetics of Homeostasis Revisited." *American Naturalist* 125: 550–572.

Parmesan, C. 2006. "Ecological and Evolutionary Responses to Recent Climate Change." *Annual Review of Ecology, Evolution, and Systematics* 37: 637–669.

Pendergrass, A. G., R. Knutti, F. Lehner, C. Deser, and B. M. Sanderson. 2017. "Precipitation Variability Increases in a Warmer Climate." *Scientific Reports* 7: 17966.

Polato, N. R., B. A. Gill, A. A. Shah, et al. 2018. "Narrow Thermal Tolerance and Low Dispersal Drive Higher Speciation in Tropical Mountains." *Proceedings of the National Academy of Sciences* 115: 12471–12476.

Quintero, I., and J. J. Wiens. 2013. "What Determines the Climatic Niche Width of Species? The Role of Spatial and Temporal Climatic Variation in Three Vertebrate Clades." *Global Ecology and Biogeography* 22: 422–432.

R Core Team. 2024. "R: A Language and Environment for Statistical Computing."

Revell, L. J. 2024. "Phytools 2.0: An Updated R Ecosystem for Phylogenetic Comparative Methods (And Other Things)." *PeerJ* 12: e16505.

Ringen, E. J., P. Duda, and A. V. Jaeggi. 2019. "The Evolution of Daily Food Sharing: A Bayesian Phylogenetic Analysis." *Evolution and Human Behavior* 40: 375–384.

Riotte-Lambert, L., and J. Matthiopoulos. 2020. "Environmental Predictability as a Cause and Consequence of Animal Movement." *Trends in Ecology & Evolution* 35: 163–174.

Rubenstein, D. R., and I. J. Lovette. 2007. "Temporal Environmental Variability Drives the Evolution of Cooperative Breeding in Birds." *Current Biology* 17: 1414–1419.

Rubio de Casas, R., C. G. Willis, W. D. Pearse, C. C. Baskin, J. M. Baskin, and J. Cavender-Bares. 2017. "Global Biogeography of Seed Dormancy Is Determined by Seasonality and Seed Size: A Case Study in the Legumes." *New Phytologist* 214: 1527–1536.

Salguero-Gómez, R., O. R. Jones, E. Jongejans, et al. 2016. "Fast–Slow Continuum and Reproductive Strategies Structure Plant Life-History Variation Worldwide." *Proceedings of the National Academy of Sciences* 113: 230–235.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27: 379–423.

Shumway, R. H., D. S. Stoffer, and D. S. Stoffer. 2000. *Time Series Analysis and Its Applications*. Springer.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill.

Stearns, S. C. 1981. "On Measuring Fluctuating Environments: Predictability, Constancy, and Contingency." *Ecology* 62: 185–199.

Stenert, C., R. Wüsth, M. M. Pires, R. F. Freiry, D. Nielsen, and L. Maltchik. 2017. "Composition of Cladoceran Dormant Stages in Intermittent Ponds With Different Hydroperiod Lengths." *Ecological Research* 32: 921–930.

Stevens, S. S. 1958. "Measurement and Man." *Science* 127: 383–389.

Tung Ho, L., and C. Ané. 2014. "A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models." *Systematic Biology* 63: 397–408.

Van de Walle, J., R. Fay, J.-M. Gaillard, et al. 2023. "Individual Life Histories: Neither Slow nor Fast, Just Diverse." *Proceedings. Biological Sciences* 290: 20230511.

Walker, P. A., and K. D. Cocks. 1991. "HABITAT: A Procedure for Modelling a Disjoint Environmental Envelope for a Plant or Animal Species." *Global Ecology and Biogeography Letters* 1: 108–118.

Wiens, J. J., D. D. Ackerly, A. P. Allen, et al. 2010. "Niche Conservatism as an Emerging Principle in Ecology and Conservation Biology." *Ecology Letters* 13: 1310–1324.

Yamamichi, M., A. D. Letten, and S. J. Schreiber. 2023. "Eco-Evolutionary Maintenance of Diversity in Fluctuating Environments." *Ecology Letters* 26: S152–S167.

Zaccarelli, N., B.-L. Li, I. Petrosillo, and G. Zurlini. 2013. "Order and Disorder in Ecological Time-Series: Introducing Normalized Spectral Entropy." *Ecological Indicators* 28: 22–30.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** geb70108-sup-0001-Supinfo.pdf.