Received: 11 April 2023 | Revised: 26 September 2023 | Accepted: 30 October 2023

DOI: 10.1002/ecy.4231

STATISTICAL REPORT

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Sparse modeling for climate variable selection across trophic levels

Eliza M. Grames[1,2] | **Matthew L. Forister**[1]

[1]Biology Department, University of Nevada, Reno, Reno, Nevada, USA

[2]Department of Biological Sciences, Binghamton University, Binghamton, New York, USA

Correspondence
Eliza M. Grames
Email: egrames@binghamton.edu

Funding information
Division of Environmental Biology, Grant/Award Numbers: DEB-2114793, DEB-2225092; Office of Integrative Activities, Grant/Award Number: OIA-2019528

Handling Editor: Elise F. Zipkin

Abstract

Understanding how populations respond to climate is fundamentally important to many questions in ecology, evolution, and conservation biology. Climate is complex and multifaceted, with aspects affecting populations in different and sometimes unexpected ways. Thus, when measuring the changing climate it is important to consider the complexity of the phenomenon and the number of ways it can be characterized through different metrics. We used a Bayesian sparse modeling approach to select among 80 metrics of climate and applied the approach to 19 datasets of bird, insect, and plant population responses to abiotic conditions as case studies of how the method can be applied for climate variable selection in a time series context. For phenological datasets, mean spring temperature was frequently selected as an important climate driver, while selected predictors were more diverse for population metrics such as abundance or reproductive success. The climate variable selection approach presented here can help to identify potential climate metrics when there is limited physiological or mechanistic information to make an *a priori* variable selection, and is broadly applicable across studies on population responses to climate.

KEYWORDS
climate metrics, global change, impacts of climate change, phenology shifts, sparsity, variable selection

INTRODUCTION

Over the past century, the mean global temperature has risen rapidly as a result of anthropogenic climate forcing. Global climate change has a multitude of impacts on biodiversity, including extinctions and extirpations (Cahill et al., 2013; Urban, 2015); population declines as well as population increases, especially for introduced species and pests (Jactel et al., 2019); range expansions, contractions, and shifts (McCarty, 2001; Thuiller, 2004); microevolution and adaptation to novel climates or expression of phenotypic plasticity (Charmantier et al., 2008; Nicotra et al., 2010); and phenological shifts in which life history events are advanced or delayed (Renner & Zohner, 2018; Visser & Both, 2005).

Understanding how climate affects populations is a central question in ecology, however it is also challenging to address because climate is inherently multifaceted, partly stochastic, and measured with diverse technologies and approaches (Garcia et al., 2014). Different aspects of climate may have effects on populations that are positive, detrimental, synergistic, or have opposing direct and indirect effects, which can result in species responses that are difficult to predict and asynchronous even between closely related or interacting species. For example, some insects exhibit enhanced population growth under warming conditions due to increased rates of development, resulting in more generations per year, but the same warming conditions also shift the phenology of their host plants that can have negative effects, ultimately resulting in population declines (Jactel et al., 2019; Pelini et al., 2009). Without thorough *a priori* knowledge of the mechanisms that lead from shifts in abiotic conditions to population response, which is rarely available for nonmodel species, decisions about what aspects of climate to include in statistical models can be arbitrary.

In addition to lacking sufficient biological information with which to fully inform variable selection, there are also methodological challenges. Researchers have proposed many alternative ways of quantifying climate and climate change: analyzing the difference in means between recent decades and a historical reference period (Kling et al., 2020), characterizing the overlap between the distributions of variables in univariate or multivariate space (Nadeau & Fuller, 2015); accounting for spatial and temporal shifts using the velocity of climate change (Brito-Morales et al., 2018) and minimum cumulative exposure to unfavorable climates (Dobrowski & Parks, 2016); or measuring departure from baseline conditions (Abatzoglou et al., 2020). Given the number of publicly available downscaled global climate models and the data layers they contain, the number of ways researchers have proposed to measure climate change (e.g., see Garcia et al., 2014), the issue of seasonality, and whether extremes, means, or frequency of events are considered, there are essentially infinite ways to measure changes in the climate over recent decades. Depending on which approaches are used and which variables are included, researchers can arrive at vastly different conclusions about the effects of climate change on biodiversity (Baker et al., 2016).

Despite recognizing the complexity of climate change and how it is quantified, ecologists frequently use simplistic measures when modeling the effects of climate change on populations. In terrestrial systems, researchers frequently use trends in mean annual temperature and precipitation to characterize climate change even when climate is expected to affect populations differently throughout the year (Garcia et al., 2014; van de Pol et al., 2016) including through extremes and accumulated degrees above developmental thresholds.

Considering the multitude of ways in which climate change could be characterized, there is an open question not just of which variables should be considered, but how to select among any suite of potentially important climate variables for analysis. Variable selection can be thought of as a subset of model selection, in which researchers develop a candidate model set and use information criteria (Anderson & Burnham, 2002) to reduce the number of potential variables and avoid overfitting. Given the number of potential climate variables, it is practically impossible to develop a tractable candidate set of models for systems with limited physiological or mechanistic information that could inform *a priori* variable selection. For example, consider a study in which researchers expect that temperature and precipitation both affect some population parameter, but have no *a priori* reason to know during which seasons and whether extreme conditions (e.g., minimum and maximum temperature) or mean conditions are more important. Considering two variables, four seasons (in a temperate system), and minima, maxima, and means, resulting in 24 potential climate variables. Building a candidate model set in which each model contains one to four combinations of those variables leads to 49,152 candidate models; including combinations of five variables increases the candidate model set to 261,672 models. Thus, traditional approaches to variable selection in ecology based on information criteria, significance, stepwise selection, and other methods are not practical for climate variable selection.

The case of selecting climate variables as covariates for long-term ecological data can be viewed as a sparse modeling problem, for which the number of predictors (i.e., the climate variables) is greater than the number of observations (Hastie et al., 2009). In sparse modeling, only a few of many covariates are assumed to have true effects, while other variables are unimportant and the analyst seeks to constrain them to have zero coefficients or otherwise exclude them. This is often done using a global shrinkage parameter or penalty, typically denoted λ, in which larger values of λ impose more shrinkage on the model parameters. This is the approach taken in penalized regressions (e.g., in the ridge or LASSO models), although there are many sparse modeling approaches (O'Hara & Sillanpää, 2009). One method that has recently been introduced in ecology is Bayesian regression with horseshoe priors for estimated effects of covariates (Sen et al., 2023; Weiss-Lehman et al., 2022). The horseshoe prior is useful for variable selection because the flat tails allow for large parameter estimates, thus avoiding imposing shrinkage on parameters that truly have large effects, while most estimates are pulled toward zero.

Here, we advocate for using sparse modeling for climate variable selection because it has the advantage of allowing researchers to fit a single model for variable selection rather than building many candidate models and is flexible enough to allow some variables with known or expected effects to not be subjected to shrinkage. We use "climate variables" to mean summary metrics from a location over seasonal or annual periods, as opposed to proximate weather conditions. We illustrate the approach with a set of case studies documenting bird, plant, and insect populations and community responses to climate. We only consider a suite of 80 univariate measures of climate divided seasonally. However, the approach can be expanded to encompass any number of climate variables, including the methods described above.

MATERIALS AND METHODS

To identify long-term datasets documenting population responses to climate, we searched Dryad and the Environmental Data Initiative Portal for data sets with at least 10 years of data collected on plants, insects, or birds. We only included data sets from studies in which there was some expectation that the population or community was responding to climate and for which the response could reasonably be summarized at the population level on an annual basis; we did not include studies on morphological or genetic changes. To be included, studies had to report latitudes and longitudes, or use recognizable site names for which we could easily identify study site coordinates. Based on these criteria and attempting to balance taxonomic representation, we selected 19 data sets (Table 1) for the analysis; for studies with multiple outcomes (e.g., multiple species responses) we selected the outcome with the most coverage (e.g., the species with the most observations or site with the longest time series); if there was no single outcome that met this criterion, an arbitrary outcome was selected. There are hundreds of long-term data sets on this topic that have been included in previous meta-analyses (e.g., Halupka & Halupka, 2017; Massad & Dyer, 2010), however, we required access to the raw data and coordinates for analysis rather than summary statistics, which is why we adopted a targeted approach to dataset selection.

To quantify climate, we used six climate layers included in the TerraClimate dataset (Abatzoglou et al., 2018) at monthly scales from 1958 to 2020: maximum

TABLE 1 Characteristics of studies included in the analysis.

Figure 1	Study	Species	Taxonomic group	Response	First year	Last year
1	McNulty, 2018	Ruffed grouse (*Bonasa umbellus*)	Birds	Date of first drumming	1984	2017
2	Ramakers et al., 2018	Great tit (*Parus major*)	Birds	Mean laying date	1973	2016
3	Ma et al., 2020	*Aesculus hippocastanum*	Plants	Leaf out date	1958	2013
4	Wadgymar et al., 2019	Two-lobe larkspur (*Delphinium nuttallianum*)	Plants	Date of first flowering	1973	2016
5	O'Keefe, 2021	Striped maple (*Acer pennsylvanicum*)	Plants	Budburst date	1990	2019
6	Visser et al., 2021	Caterpillar community	Insects	Total frass fall	1985	2020
7	Hinks et al., 2015	Great tit (*Parus major*)	Birds	Mean laying date	1965	2009
8	Pöysä, 2019	Common goldeneye (*Bucephala clangula*)	Birds	Mean settling date	1991	2018
9	Werner et al., 2016	Ips beetles (Scolytinae)	Insects	Abundance	1972	2012
10	Wiebe, 2020	Northern flicker (*Colaptes auratus*)	Birds	Abundance	1998	2013
11	Huenneke & Browning, 2022	Burrograss (*Scleropogon brevifolius*)	Plants	Percent fruiting in September	1992	2020
12	Donoso et al., 2016	Green-veined white (*Pieris napi*)	Insects	Abundance	1996	2012
13	McLean et al., 2020	Willow warbler (*Phylloscopus trochilus*)	Birds	Body condition	1994	2014
14	Cole et al., 2016	Great tit (*Parus major*)	Birds	Mean laying date	2001	2013
15	Valtonen et al., 2018	Moth community	Insects	Species richness	1962	2009
16	DeMay & Walters, 2020	Red-cockaded woodpecker (*Dryobates borealis*)	Birds	Abundance	1980	2015
17	Lightfoot, 2021	Grasshoppers	Insects	Spring abundance	1992	2019
18	Frigerio et al., 2021	Greylag goose (*Anser anser*)	Birds	First laying date	1990	2018
19	Weed et al., 2016	Southern pine beetle (*Dendroctonus frontalis*)	Insects	Abundance	1987	2009

Note: Studies are ordered from top to bottom in the same order as studies are grouped in Figure 1 from left to right based on similarity of response.

temperature, minimum temperature, precipitation, vapor pressure deficit, Palmer drought severity index, and soil moisture. For maximum and minimum temperature, the values represent the mean daily maximum and minimum temperature within a month. For each layer, we calculated the mean value across meteorological months for the previous fall, previous winter, current spring, current summer, and the annual mean in the current calendar year; for temperature, we also included extreme conditions (i.e., mean daily minimum temperature in the coldest month and mean daily maximum temperature in the warmest month for each season and annually). For each of these layer and season combinations, we extracted the value in the proximate year and also calculated the departure in that year from baseline conditions during the period from 1958 to 1987 using Mahalanobis distance (Abatzoglou et al., 2020). We only calculated univariate metrics for this analysis, as opposed to multivariate distance metrics, resulting in a total of 80 metrics. We extracted these metrics within a 4 km^2 cell, including the central coordinates of each study site for the years included in the study.

To avoid imposing excess shrinkage on parameters that truly have large effects, we used a Bayesian approach that included both global and local shrinkage parameters. For each ecological data set included in our case study, we fitted a Bayesian sparse regression model with a horseshoe estimator as the prior for coefficients associated with all climate variables. We assumed that each observation i of the population or community response y arose from a normal distribution with mean μ and a variance σ estimated from an uninformative gamma prior (Equations 1 and 10). We estimated the mean μ as a function of a random intercept, the effect of year t (using an uninformative normal prior), and the climate variables $\boldsymbol{X}$ (Equations 2, 11, and 12). For time series in which sampling effort can vary greatly from year to year and across study locations, it is critical to include sampling effort as a covariate that may explain differences in population responses due to observational rather than ecological processes (Botella et al., 2020; Rhodes & Jonzen, 2011). Sampling effort was largely not reported in the datasets selected as case studies, however, and thus we did not include it in the model below, but it could easily be incorporated in the same manner as the effect of year when it is not subject to variable selection if such data were available. For each climate variable in $\boldsymbol{X}$, we estimated if it should be included in the model with the parameter $\boldsymbol{\gamma}$ (drawn from a Bernoulli prior with probability 0.5) and subsequently modeled the effect $\boldsymbol{\beta}$ with a horseshoe estimator using half-Cauchy priors for the global shrinkage parameters and the local shrinkage parameter λ fixed to 0.1 (Equations 3–10, 13), following Roberts and Zhao (2022). This model structure allowed us to estimate both the probability of inclusion for every climate variable based on $\boldsymbol{\gamma}$ and generate parameter estimates for selected variables.

We fitted the models using the Markov chain Monte Carlo (MCMC) approach implemented in JAGS (Plummer, 2003) using saveJAGS v0.0.4.9002 (Meredith, 2021). After a burn-in of 100,000, we ran the models for one million iterations and sampled the posterior every 20 iterations, resulting in a posterior sample of 50,000 from three chains. We assessed model convergence with the Gelman–Rubin statistic ($\widehat{R}$) and assumed chains had converged when $\widehat{R}$ was less than or equal to 1.1. When models failed to converge, we increased the number of iterations in intervals of 200,000 until chains converged. Because models for different case studies ran for different numbers of iterations, we resampled the posterior to draw 5000 samples from each chain (for a total of 15,000 samples from the posterior) for primary results reporting. To identify patterns in which variables were selected across studies, we used the Euclidean distance between the mean $\boldsymbol{\gamma}$ values from the model output for hierarchical cluster analysis with the R package *stats* v4.1.2 (R Core Team, 2021):

$$y_i \sim N(\mu_i, \sigma), \tag{1}$$

$$\mu_i = \beta_0 + \beta_1 \times t_i + \boldsymbol{\gamma} \times \boldsymbol{\beta} \times \mathbf{X}, \tag{2}$$

$$\boldsymbol{\beta} \sim N(0, \lambda \times \boldsymbol{\nu} \times \omega), \tag{3}$$

$$\boldsymbol{\nu} = {}^{|\chi_1|}\!/\!_{\sqrt{\tau_1}}, \tag{4}$$

$$\omega = {}^{|\chi_2|}\!/\!_{\sqrt{t_2}}, \tag{5}$$

$$\tau_1 \sim \text{Gamma}(0.5, 0.5), \tag{6}$$

$$\tau_2 \sim \text{Gamma}(0.5, 0.5), \tag{7}$$

$$\chi_1 \sim N(0, 1), \tag{8}$$

$$\chi_2 \sim N(0, 1), \tag{9}$$

$$\sigma \sim \text{Gamma}(0.01, 0.01), \tag{10}$$

$$\beta_0 \sim N(0, 0.001), \tag{11}$$

$$\beta_1 \sim N(0, 0.001), \tag{12}$$

$$\gamma \sim \text{Bern}(0.5). \tag{13}$$

Climate variables tend to be highly correlated with one another, which can result in issues with collinearity and multicollinearity that are problematic for many approaches to variable selection (Lu & Lou, 2022). Previous simulations have found that the performance of horseshoe priors is relatively insensitive to increases in correlations among predictor variables from 0 to 0.9 (Lu & Lou, 2022). However, several of our climate variables were correlated with one another at greater than 0.90. To assess if this extreme multicollinearity affected variable selection, we performed a sensitivity analysis and re-ran the same models for all case studies but randomly dropped one variable from each pair of highly correlated predictors ($\rho > 0.95$). We then compared the mean posterior inclusion probability for the models fit with all variables, to those fit with a subset of predictors that all had correlations with one another <0.95.

RESULTS

Models for most case study data sets converged after one million iterations with $\widehat{R} < 1.1$. Seven datasets required additional iterations (Donoso et al., 2016; Huenneke & Browning, 2022; Ma et al., 2020; McLean et al., 2020; O'Keefe, 2021; Valtonen et al., 2018; Wadgymar et al., 2019), ranging from 1.2 million iterations to a maximum of 1.8 million iterations. Mean, standardized parameter estimates across case studies ranged from −0.490 to 0.34, for which larger absolute values indicate greater effects, with a mean overall estimate of 0.00 that would be expected given the horseshoe prior (Figure 1a). Here, we do not focus on the parameter estimates, but rather the probability of inclusion for each variable across the case studies. In some cases (Appendix S1: Figure S1), variables had high probabilities of inclusion but had negative effects on the population or community response (e.g., Figure 1a; proximate mean minimum temperature in the previous fall), whereas other variables had positive effects (e.g., Figure 1a; proximate mean maximum temperature in the previous winter). Out of all 80 variables, proximate mean spring maximum temperature had the highest probability of inclusion (mean = 0.62), across species and studies (Figure 1b). Univariate departures from baseline conditions had a low probability of inclusion for most variables across all case studies (Appendix S1: Figure S1). Based on similarity in the probability of inclusion for each variable, case studies clustered by type of response rather than by taxa or region (Figure 1b), with an emergent cluster of phenology studies. Models fit with all variables or a subset of variables as a sensitivity analysis for collinearity yielded similar mean posterior inclusion probabilities (Appendix S1: Figure S2).

Not all of the case study datasets had corresponding publications that assessed the impact of climate change. However, for nine studies, we were able to compare the variables that the original authors analyzed or found to be important to variables with a high probability of inclusion (arbitrarily set at 0.65) in the sparse regression model reported here. We qualitatively discuss those differences (see Appendix S1: Table S1), with the caveat that none of the original papers were describing variable selection on the scale considered here, and in some cases the questions addressed were tangential to the analyses discussed here. In half the cases, the variables with the probability of inclusion greater than 0.65 in the sparse regression model were very similar to those considered by the original authors. For example, Wadgymar et al. (2019) analyzed spring temperature and the timing of snowmelt in relation to the first flower date, and the sparse regression model indicated a high probability of inclusion for spring maximum temperature and summer soil moisture (presumably indirectly influenced by snowmelt). Spring maximum temperature was the only variable with a high probability of inclusion for two studies that only considered spring temperature effects (Hinks et al., 2015; Visser et al., 2021). In two cases, the original authors did not select any climate variables for analysis; Weed et al. (2016) represented weather in the error term, and Valtonen et al. (2018) discussed temperature and precipitation but found no significant change in climate over the course of their study. In both cases, the sparse regression model did not indicate a high probability of inclusion for any of the 80 variables we considered. Several papers analyzed metrics with no direct comparison to the 80 variables we considered (e.g., ice-out date, photoperiod, cloud cover, see Appendix S1: Table S1), or the model indicated a high probability of inclusion for variables not considered in the original paper.

DISCUSSION

The current biodiversity crisis is a product of the rapid environmental changes impacting nearly every ecosystem and ecological process (Parmesan & Yohe, 2003). Understanding how the biosphere will persist through these changes is the single greatest challenge facing the field of ecology. Among the threats facing biodiversity, climate change will continue to grow as a driver of population change across the globe, with a suite of complex and interrelated aspects of climate change affecting populations in different ways. While the approach we advocate for here is exploratory and has limited or no predictive ability, it can help to serve as the basis for future research by selecting variables that are associated with biodiversity responses. Because of its complexity, there are numerous ways to characterize climate change

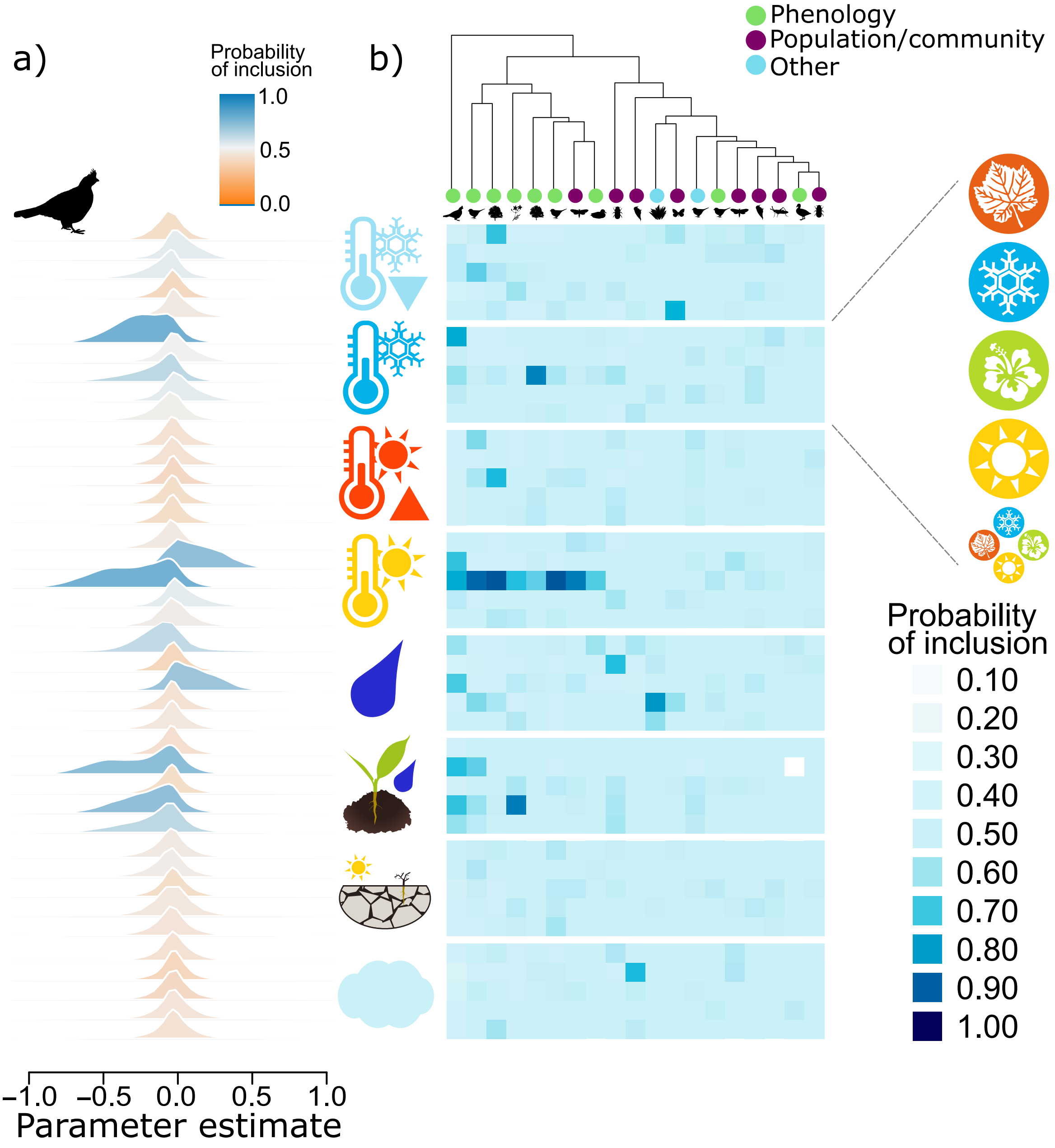

FIGURE 1 Probability of inclusion for different climate variables measured as proximate values within a study year for (a) McNulty (2018) as an example, and (b) across case studies. In (a) posterior distributions of parameter estimates are shown for all 40 proximate climate variables; color corresponds to mean probability of inclusion (γ) with red colors indicating probabilities less than 0.5 and blue colors indicating probabilities greater than 0.5. Climate variables are indicated at left from top to bottom as extreme minimum temperature, mean minimum temperature, extreme maximum temperature, mean maximum temperature, mean precipitation, soil moisture, Palmer drought severity index, and vapor pressure deficit. The top row in each block shows previous fall, followed by winter, spring, summer, and annual conditions. In (b) intensity of blue within cells indicates the mean probability of inclusion from the models. Studies are ordered from left to right based on similarity of response to climate variables, shown with a dendrogram at top, corresponding to the grouping order found in Table 1. For each study, the taxon is indicated with a silhouette and color indicating if the response is timing or phenology (green), a population or community response such as abundance or species richness (purple), or other metrics that are percent fruit and body condition (blue). Icons were created by Eliza Grames or are licensed in the public domain.

(Garcia et al., 2014). Sparse modeling can be used to select among the many measures of climate rather than oversimplifying it with one or a few broad measures, enabling researchers to build informed models with a small and more readily interpreted set of variables as the basis for future analyses (Hastie et al., 2009).

When applying sparse modeling to climate variable selection, we found emergent patterns in which variables had a high probability of inclusion for models in which the responses were phenological shifts but not for population and community responses. These results suggest that assumptions of which variables matter are warranted for some research questions related to phenology where many population responses are associated with similar predictor variables, but that these relationships are not universal, and in many cases, it is necessary to consider many potential climate variables. The modeling approach we advocate for here can help to identify associations between measures of climate and biodiversity responses in cases in which there are no *a priori* expectations for which variables matter.

Even within a relatively small set of case studies, we found similar results in which variables had a high probability of inclusion for phenology case studies (Figure 1b). The case studies we included were from temperate regions, where phenological responses have been studied much more than in tropical regions, especially in spring (Cohen et al., 2018). In the tri-trophic context of plants, insects, and birds considered here, many responses depend on spring bud burst or leaf out. For example, in temperate deciduous forests, bud burst often coincides with spring temperature, which also triggers insects to emerge from diapause, and the phenology of many insectivorous bird species is synchronized to the period of peak food availability in the form of insect resources (Visser & Both, 2005). Thus, finding a consistent relationship between temperature conditions and population responses is not wholly unexpected for phenology studies, and is in keeping with what has previously been considered in the literature (Appendix S1: Table S1). This consistency suggests that researchers studying the effects of climate on tri-trophic phenology in temperate regions of the Northern Hemisphere may be able to narrow down the set of candidate climate variables. The seasonality of temperature, however, does seem to matter for phenology studies and researchers should not default to using annual mean values for analyses when spring metrics are likely to be the most important when studying spring and summer phenology and indeed even more narrow windows are important and likely to be worth consideration for many studies (van de Pol et al., 2016).

Despite widely documented climate effects on population abundance, biomass, and community composition (McCarty, 2001), we found considerable heterogeneity of associations between responses among our case studies and climate variables. That is not to say that abundance and community composition are not shifting in response to climate change, but rather that climate change is multifaceted and different aspects may be responsible for observed responses across different taxa and regions. For example, we found that the grass fruit set (Huenneke & Browning, 2022) was primarily associated with summer precipitation, whereas northern flicker abundance (Wiebe, 2020) was associated with the previous winter vapor pressure deficit (presumably indirectly). Populations often exhibit delayed responses to extrinsic factors, with weather in previous years having lagged effects on abundance or species richness in future years (Evers et al., 2021; Salcido et al., 2020; Thompson & Ollason, 2001; Wu et al., 2015). We did not include lagged effects in our models because of the heterogeneity of datasets and our goal of emphasizing the method through case studies rather than aiming to make biological inferences in any one system; however, including lagged effects based on the biology of the system and more sophisticated time series analysis may help researchers working on tri-trophic responses to climate change to uncover these types of responses. Similarly, researchers could select more biologically relevant sets of candidate variables than the ones presented here. For example, snow cover is likely to be an important variable in several of the case studies (e.g., McNulty, 2018; Wadgymar et al., 2019), however, others were conducted in areas with no snow cover data (e.g., DeMay & Walters, 2020; Lightfoot, 2021) and for consistency we did not include snow in the models.

The method presented here is one of many sparse modeling approaches that could be used for climate variable selection (Hastie et al., 2009). The advantage of using a Bayesian method is that it produces posterior distributions for variables that can themselves be informative. For example, it can be useful to separate the probability that a variable should be included from its estimated effect. Because climate variables are frequently interrelated and collinearity can be an issue, skewed distributions or convergence issues can pinpoint variables that may be causing issues with model fit or convergence. The issue of collinearity is especially true of climate variables, and researchers should be aware that parameter values will frequently be estimated incorrectly in the presence of strong multicollinearity (Piironen & Vehtari, 2017). Horseshoe priors will tend to estimate a high probability of inclusion for only one of a set of highly correlated predictors. However, unlike frequentist approaches to variable selection, the overall model performance is relatively insensitive to high collinearity among predictors (Lu & Lou, 2022). Researchers should be cautious and not take the parameter estimates from Bayesian sparse modeling at face value, in part because the approach is exploratory, but also because the prior distribution is designed to draw estimates closer to zero (Roberts & Zhao, 2022). As with all statistical approaches to variable selection, the decision about which variables to include in the final models should be made by the researcher based on the biology of the system. With increased computational power and data storage capacity, ecology has entered a "big data" era in which the amount of available data layers to use as predictors often exceeds the number of observations that are still constrained by resources available for field data collection. Sparse modeling addresses this challenge by allowing researchers to select among the thousands of possible predictor variables to develop models understanding how biodiversity responds to global change.

With more and more studies documenting the effects of climate change on biodiversity, it has become apparent that population responses are often counterintuitive and that there are many unknown or unexpected effects of climate change. The sparse modeling approach for climate variable selection that we advocate for here is well suited to exploratory analyses and identifying associations between biodiversity responses and climate change. For example, previous winter vapor pressure deficit emerged as a variable with a high probability of inclusion in models of northern flicker abundance (Wiebe, 2020), but there is no plausible reason to expect that it has direct effects on the population. Rather, vapor pressure deficit probably has indirect effects on the plant and insect communities upon which the birds depend (Grossiord et al., 2020), and as such, this approach can highlight avenues for future research to identify more proximate factors that are influenced by climate change. Indeed, this approach can highlight suites of variables for subsequent analyses, including multivariate metrics or composite variables (Abatzoglou et al., 2020), using different historical periods for baseline conditions (Baker et al., 2016), or assessing responses at more fine-scale temporal windows (van de Pol et al., 2016). Simply including annual mean temperature as a measure of climate change can mask these more nuanced aspects of climate. Sparse modeling to select among many climate variables can highlight directions for future research to build toward more mechanistic models predicting the effects of climate change on biodiversity that are necessary for conservation planning.

ACKNOWLEDGMENTS

We are immensely grateful to the researchers who collected the population data used for these analyses and who made those data publicly available. We thank the editor and two anonymous reviewers for helpful comments on an earlier version of the manuscript, and the Modelscape Consortium for helpful discussion. We acknowledge support from the National Science Foundation (OIA-2019528, DEB-2114793, and DEB-2225092).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data and code (Grames & Forister, 2023) are available in Figshare at https://doi.org/10.6084/m9.figshare.24168159.

ORCID

Eliza M. Grames https://orcid.org/0000-0003-1743-6815

REFERENCES

Abatzoglou, J. T., S. Z. Dobrowski, and S. A. Parks. 2020. "Multivariate Climate Departures Have Outpaced Univariate Changes across Global Lands." *Scientific Reports* 10: 1–9.

Abatzoglou, J. T., S. Z. Dobrowski, S. A. Parks, and K. C. Hegewisch. 2018. "TerraClimate, a High-Resolution Global Dataset of Monthly Climate and Climatic Water Balance from 1958 to 2015." *Scientific Data* 5: 1–12.

Anderson, D. R., and K. P. Burnham. 2002. "Avoiding Pitfalls When Using Information-Theoretic Methods." *The Journal of Wildlife Management* 66(3): 912. https://doi.org/10.2307/3803155.

Baker, D. J., A. J. Hartley, S. H. Butchart, and S. G. Willis. 2016. "Choice of Baseline Climate Data Impacts Projected Species' Responses to Climate Change." *Global Change Biology* 22: 2392–2404.

Botella, C., A. Joly, P. Monestiez, P. Bonnet, and F. Munoz. 2020. "Bias in Presence-Only Niche Models Related to Sampling Effort and Species Niches: Lessons for Background Point Selection." *PLoS One* 15: e0232078.

Brito-Morales, I., J. G. Molinos, D. S. Schoeman, M. T. Burrows, E. S. Poloczanska, C. J. Brown, S. Ferrier, et al. 2018. "Climate Velocity Can Inform Conservation in a Warming World." *Trends in Ecology & Evolution* 33: 441–457.

Cahill, A. E., M. E. Aiello-Lammens, M. C. Fisher-Reid, X. Hua, C. J. Karanewsky, H. Yeong Ryu, G. C. Sbeglia, et al. 2013. "How Does Climate Change Cause Extinction?" *Proceedings of the Royal Society B* 280: 20121890.

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

Cohen, J. M., M. J. Lajeunesse, and J. R. Rohr. 2018. "A Global Synthesis of Animal Phenological Responses to Climate Change." *Nature Climate Change* 8: 224–28.

Cole, E. F., P. R. Long, P. Zelazowski, M. Szulkin, and B. C. Sheldon. 2016. "Data from: Predicting Bird Phenology from Space: Satellite-Derived Vegetation Green-up Signal Uncovers Spatial Variation in Phenological Synchrony between Birds and their Environment." Dryad. https://doi.org/10.5061/dryad.7v1qg.

DeMay, S. M., and J. R. Walters. 2020. "Data from: Variable Effects of a Changing Climate on Lay Dates and Productivity across the Range of the Red-Cockaded Woodpecker." Dryad. https://doi.org/10.5061/dryad.0qv86b5.

Dobrowski, S. Z., and S. A. Parks. 2016. "Climate Change Velocity Underestimates Climate Change Exposure in Mountainous Regions." *Nature Communications* 7: 1–8.

Donoso, I., C. Stefanescu, A. Martinez-Abrain, and A. Traveset. 2016. "Data from: Phenological Asynchrony in Plant–Butterfly Interactions Associated with Climate: A Community-Wide Perspective." Dryad. https://doi.org/10.5061/dryad.72551.

Evers, S. M., T. M. Knight, D. W. Inouye, T. E. Miller, R. Salguero-Gómez, A. M. Iler, and A. Compagnoni. 2021. "Lagged and Dormant Season Climate Better Predict Plant Vital Rates than Climate during the Growing Season." *Global Change Biology* 27: 1927–41.

Frigerio, D., P. Sumasgutner, K. Kotrschal, S. Kleindorfer, and J. Hemetsberger. 2021. "From Individual to Population Level: Temperature and Snow Cover Modulate Fledging Success through Breeding Phenology in Greylag Geese (*Anser Anser*)." Dryad. https://doi.org/10.5061/dryad.np5hqbztd.

Garcia, R. A., M. Cabeza, C. Rahbek, and M. B. Araújo. 2014. "Multiple Dimensions of Climate Change and their Implications for Biodiversity." *Science* 344: 1247579.

Grames, E. M., and M. Forister. 2023. "Data from: Sparse Modeling for Climate Variable Selection across Trophic Levels." Figshare. https://doi.org/10.6084/m9.figshare.24168159.

Grossiord, C., T. N. Buckley, L. A. Cernusak, K. A. Novick, B. Poulter, R. T. Siegwolf, J. S. Sperry, and N. G. McDowell. 2020. "Plant Responses to Rising Vapor Pressure Deficit." *New Phytologist* 226: 1550–66.

Halupka, L., and K. Halupka. 2017. "The Effect of Climate Change on the Duration of Avian Breeding Seasons: A Meta-Analysis." *Proceedings of the Royal Society B* 284: 20171710.

Hastie, T., R. Tibshirani, and J. H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Hinks, A. E., E. F. Cole, K. J. Daniels, T. A. Wilkin, S. Nakagawa, and B. C. Sheldon. 2015. "Data from: Scale-Dependent Phenological Synchrony between Songbirds and their Caterpillar Food Source." Dryad. https://doi.org/10.5061/dryad.n6g3q.

Huenneke, L., and D. Browning. 2022. "Plant Phenology Observations at 15 Sites at Jornada Basin LTER, 1992-Ongoing ver 84." EDI Data Portal. https://doi.org/10.6073/pasta/bf98e185c512a8dff9677a611dc05455.

Jactel, H., J. Koricheva, and B. Castagneyrol. 2019. "Responses of Forest Insect Pests to Climate Change: Not So Simple." *Current Opinion in Insect Science* 35: 103–8.

Kling, M. M., S. L. Auer, P. J. Comer, D. D. Ackerly, and H. Hamilton. 2020. "Multiple Axes of Ecological Vulnerability to Climate Change." *Global Change Biology* 26: 2798–2813.

Lightfoot, D. 2021. "Long-Term Core Site Grasshopper Dynamics for the Sevilleta National Wildlife Refuge, New Mexico ver 214969." EDI Data Portal. https://doi.org/10.6073/pasta/8da81e91177dc6029d147990d1181be3.

Lu, Z., and W. Lou. 2022. "Bayesian Approaches to Variable Selection: A Comparative Study from Practical Perspectives." *The International Journal of Biostatistics* 18: 83–108.

Ma, Q., J.-G. Huang, H. Hanninen, X. Li, and F. Berninger. 2020. "Data from: Climate Warming Prolongs the Time Interval between Leaf-out and Flowering in Temperate Trees: Effects of Chilling, Forcing and Photoperiod." Dryad. https://doi.org/10.5061/dryad.jdfn2z391.

Massad, T. J., and L. A. Dyer. 2010. "A Meta-Analysis of the Effects of Global Environmental Change on Plant-Herbivore Interactions." *Arthropod-Plant Interactions* 4: 181–88.

McCarty, J. P. 2001. "Ecological Consequences of Recent Climate Change." *Conservation Biology* 15: 320–331.

McLean, N., H. van der Jeugd, C. van Turnhout, J. Lefcheck, and M. van de Pol. 2020. "Reduced Avian Body Condition Due to Global Warming Has Little Reproductive or Population Consequences." Dryad. https://doi.org/10.5061/dryad.zs7h44j56.

McNulty, S. 2018. "Ruffed grouse (*Bonasa umbellus*) drumming surveys, 1987-2017, Adirondack Long-Term Ecological Monitoring Program Project No. 9." EDI Data Portal. https://doi.org/10.6073/pasta/39f495a15fa0667af635991d749741c4.

Meredith, M. 2021. "saveJAGS: Run JAGS and Regularly Save Output to Files." *R package version 0.0.4.9002*. https://github.com/mikemeredith/saveJAGS.

Nadeau, C. P., and A. K. Fuller. 2015. "Accounting for Multiple Climate Components when Estimating Climate Change Exposure and Velocity." *Methods in Ecology and Evolution* 6: 697–705.

Nicotra, A. B., O. K. Atkin, S. P. Bonser, A. M. Davidson, E. J. Finnegan, U. Mathesius, P. Poot, et al. 2010. "Plant Phenotypic Plasticity in a Changing Climate." *Trends in Plant Science* 15: 684–692.

O'Hara, R. B., and M. J. Sillanpää. 2009. "A Review of Bayesian Variable Selection Methods: What, how and which." *Bayesian Analysis* 4: 85–117.

O'Keefe, J. 2021. "Phenology of Woody Species at Harvard Forest Since 1990 ver 33." EDI Data Portal. https://doi.org/10.6073/pasta/91e3b7c2548a0f2e251729eeacbce312.

Parmesan, C., and G. Yohe. 2003. "A Globally Coherent Fingerprint of Climate Change Impacts across Natural Systems." *Nature* 421: 37–42.

Pelini, S. L., K. M. Prior, D. J. Parker, J. D. K. Dzurisin, R. L. Lindroth, and J. J. Hellmann. 2009. "Climate Change and Temporal and Spatial Mismatches in Insect Communities." In *Climate Change*, edited by T. M. Letcher, 215–231. Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/b978-0-444-53301-2.00011-7.

Piironen, J., and A. Vehtari. 2017. "Sparsity Information and Regularization in the Horseshoe and Other Shrinkage Priors." *Electronic Journal of Statistics* 11: 5018–51.

Plummer, M. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling." In *Proceedings of the Third International Workshop on Distributed Statistical Computing (DSC 2003)*, 20–22 March 2003, edited by K. Hornik, F. Leisch, and A. Zeileis, 20–22. Vienna, Austria: Technische Universitat. http://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf.

Pöysä, H. 2019. "Data from: Tracking Ice Phenology by Migratory Waterbirds: Settling Phenology and Breeding Success of Species with Divergent Population Trends." Dryad. https://doi.org/10.5061/dryad.b5mkkwh89.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ramakers, J. J. C., P. Gienapp, and M. E. Visser. 2018. "Data from: Phenological Mismatch Drives Selection on Elevation, but Not on Slope, of Breeding Time Plasticity in a Wild Songbird." Dryad. https://doi.org/10.5061/dryad.35k1n3m.

Renner, S. S., and C. M. Zohner. 2018. "Climate Change and Phenological Mismatch in Trophic Interactions among Plants, Insects, and Vertebrates." *Annual Review of Ecology, Evolution, and Systematics* 49: 165–182.

Rhodes, J. R., and N. Jonzen. 2011. "Monitoring Temporal Trends in Spatially Structured Populations: How Should Sampling Effort be Allocated between Space and Time?" *Ecography* 34: 1040–48.

Roberts, E., and L. Zhao. 2022. "A Bayesian Mixture Model for Changepoint Estimation Using Ordinal Predictors." *The International Journal of Biostatistics* 18: 57–72.

Salcido, D. M., M. L. Forister, H. Garcia Lopez, and L. A. Dyer. 2020. "Loss of Dominant Caterpillar Genera in a Protected Tropical Forest." *Scientific Reports* 10: 1–10.

Sen, B., C. Che-Castaldo, K. M. Krumhardt, L. Landrum, M. M. Holland, M. A. LaRue, M. C. Long, S. Jenouvrier, and H. J. Lynch. 2023. "Spatio-Temporal Transferability of Environmentally-Dependent Population Models: Insights from the Intrinsic Predictabilities of Adelie Penguin Abundance Time Series." *Ecological Indicators* 150: 110239.

Thompson, P. M., and J. C. Ollason. 2001. "Lagged Effects of Ocean Climate Change on Fulmar Population Dynamics." *Nature* 413: 417–420.

Thuiller, W. 2004. "Patterns and Uncertainties of species' Range Shifts under Climate Change." *Global Change Biology* 10: 2020–27.

Urban, M. C. 2015. "Accelerating Extinction Risk from Climate Change." *Science* 348: 571–73.

Valtonen, A., A. Hirka, L. Szocs, M. Ayres, H. Roininen, and G. Csoka. 2018. "Data from: Long-Term Species Loss and Homogenization of Moth Communities in Central Europe." Dryad. https://doi.org/10.5061/dryad.9m6vp.

van de Pol, M., L. D. Bailey, N. McLean, L. Rijsdijk, C. R. Lawson, and L. Brouwer. 2016. "Identifying the Best Climatic Predictors in Ecology and Evolution." *Methods in Ecology and Evolution* 7: 1246–57.

Visser, M. E., and C. Both. 2005. "Shifts in Phenology Due to Global Climate Change: The Need for a Yardstick." *Proceedings of the Royal Society B* 272: 2561–69.

Visser, M. E., M. Lindner, P. Gienapp, M. Long, and S. Jenouvrier. 2021. "Recent Natural Variability in Global Warming Weakened Phenological Mismatch and Selection on Seasonal Timing in Great Tits (*Parus major*)." Dryad. https://doi.org/10.5061/dryad.f1vhhmgx6.

Wadgymar, S. M., J. E. Ogilvie, D. W. Inouye, A. E. Weis, and J. T. Anderson. 2019. "Data from: Phenological Responses to Multiple Environmental Drivers under Climate Change: Insights from a Long-Term Observational Study and a Manipulative Field Experiment." Dryad. https://doi.org/10.5061/dryad.qr5vd.

Weed, A. S., M. P. Ayres, A. M. Liebhold, and R. F. Billings. 2016. "Data from: Spatio-Temporal Dynamics of a Tree-Killing Beetle and its Predator." Dryad. https://doi.org/10.5061/dryad.ms739.

Weiss-Lehman, C. P., C. M. Werner, C. H. Bowler, L. M. Hallett, M. M. Mayfield, O. Godoy, L. Aoyama, et al. 2022. "Disentangling Key Species Interactions in Diverse and Heterogeneous Communities: A Bayesian Sparse Modelling Approach." *Ecology Letters* 25: 1263–76.

Werner, R. A., K. Van Cleve, F. S. Chapin, R. W. Ruess, J. J. Kruse, and D. Wagner. 2016. "Bonanza Creek Experimental Forest Beetles per Trap 1975–2012—Werner ver 19." EDI Data Portal. https://doi.org/10.6073/pasta/c174404b0bb5d9a65bc8eccb40db825c.

Wiebe, K. 2020. "Local Recruitment in Northern Flickers Is Related to Environmental Factors at Multiple Scales and Provides Reproductive Benefits to Yearling Breeders Settling Close to Home." Dryad. https://doi.org/10.5061/dryad.bk3j9kd71.

Wu, D., X. Zhao, S. Liang, T. Zhou, K. Huang, B. Tang, and W. Zhao. 2015. "Time-Lag Effects of Global Vegetation Responses to Climate Change." *Global Change Biology* 21: 3520–31.

SUPPORTING INFORMATION

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

How to cite this article: Grames, Eliza M., and Matthew L. Forister. 2024. "Sparse Modeling for Climate Variable Selection across Trophic Levels." *Ecology* 105(3): e4231. https://doi.org/10.1002/ecy.4231