Manuscript title: **Predictive ability of climate on vital rates and population dynamics depends on vital rate and model complexity**

Author names: Aldo Compagnoni[1,2,3.], Dylan Childs[4], Tiffany M. Knight[1,2,3], Roberto Salguero-Gómez[6,7,8]

Affiliations: [1] Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108, Halle (Saale), Germany.

Deutscher Platz 5e, Leipzig, 04103, Germany; [2] Department of Community Ecology, Helmholtz Centre for Environmental Research–UFZ, 06120 Halle (Saale), Germany; [3] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstrasse 4, Leipzig 04103, Germany; [4] Department of Animal and Plant Sciences, University of Sheffield. Western Bank, Sheffield S10 2TN, UK; [6] Department of Zoology, University of Oxford. 11a Mansfield Road, Oxford, OX1 3SZ, United Kingdom; [7] Max Planck Institute for Demographic Research, Konrad Zuse Strasse 1, Rostock, DE-18057, Germany; [8] Centre of Excellence in Environmental Decisions. The University of Queensland, St Lucia QLD 4071, Australia

Corresponding Author: Aldo Compagnoni; aldo.compagnoni@gmail.com

Open Research statement: Data and code to reproduce the analyses will be made available on zenodo and github upon publication. The code use to produce our results is available at https://github.com/AldoCompagnoni/climate_drivers_methods.

Keywords: Bayesian moving window, COMPADRE Plant Matrix Database, cross-validation, exploratory data analysis, forecasting, population growth rate, regularization, vital rate.

## Abstract

Understanding mechanisms and predicting trends in vital rates and population growth responses to climate is a key goal of ecology. It is unclear whether predictive ability changes based on the response variable (e.g. survival, reproduction, population growth rate). Studies explicitly linking climate to vital rates and population dynamics remain limited, and use different statistical tools that vary in their complexity. Antecedent effect models are the most complex statistical tools, but are hypothesized to have better predictive ability because they capitalize on the evidence provided by climate and population data to select time windows correlated with a population level response. We compare the predictive performance of antecedent effect models against simpler models. We fit three antecedent effect models: (1) weighted mean models (WMM), which weigh the importance of monthly anomalies based on a Gaussian curve, (2) stochastic antecedent models (SAM), which weigh the importance of monthly anomalies using a Dirichlet process, and (3) regularized regressions using the Finnish Horseshoe prior (FHM), which estimate a separate effect size for each monthly anomaly. We compare these approaches to a linear model using a yearly climatic predictor and a null model with no predictors. We use demographic data from 77 natural populations of 3 plant species ranging between 7 and 36 years of length. We fit models to the overall asymptotic population growth rate ($\lambda$) and its underlying vital rates: survival, development (transition between stage classes), and reproduction. Contrary to our expectations, we found that simple models using yearly climate as a predictor showed, on average, slightly better predictive ability than antecedent effect models. These models had a higher ability to predict development than survival and reproduction and a higher ability to predict vital rates than population growth rate. We find that development is more predictable than other vital rates and especially population growth rates. Improving the predictive ability of

survival and reproduction might be inherently more difficult, and might require larger datasets or conceptual advances. Conceptual advances could be facilitated by antecedent effect models which, despite not improving predictive ability, are well suited as exploratory tools.

## Introduction

Understanding and predicting population dynamics is a key objective of ecology. In population ecology, the last decades have witnessed the development and application of many statistical methods aimed precisely at this objective (Caswell 2001, Ellner et al. 2016). Increasingly, emphasis is being devoted to understanding how climatic drivers such as temperature and precipitation affect demographic rates (henceforth "vital rates") such as survival, development (*i.e.*, changes in stage along the life cycle), and reproduction (Ehrlén et al. 2016; Teller et al. 2016). The ultimate goal of this ambitious research agenda is to achieve a mechanistic understanding of what makes species vulnerable to climate change worldwide. Progress in this understanding requires synthetic studies that link vital rates to climatic drivers using standardized methods (Compagnoni et al. 2021).

In order to understand and predict the effects of climate drivers on populations, ecologists need to first understand how predictable different response variables are. While ecologists are ultimately interested in linking population growth rates to climatic drivers, this link has often poor predictive ability (Knape and de Valpine 2011). Poor predictive performance might reflect that variation in population growth rates is buffered when compared to variation in the underlying vital rates (Hilde et al. 2020). Thus, we might have a better understanding and ability to predict the effects of climate on vital rates (e.g. Clark et al., 2016, 2021; Schulze et al., 2019).

Currently, we do not know our relative ability to predict vital rates and population growth rate, and how this predictability varies across species. Establishing this predictive ability in synthesis study requires addressing the technical hurdles associated with linking climatic drivers to population dynamics.

Studies that explicitly link population responses to climate drivers typically identify climatic predictors *a priori*. Most population studies are short-term (Salguero-Gómez et al. 2015, 2016), and these low sample sizes have historically pushed investigators to consider annual or seasonal climatic effects (reviewed in Evers et al., 2021). This choice has practical reasons: first, monthly or annual climate data are widely available (Daly et al. 1994, Karger and Zimmermann 2018). Second, annual anomalies are a logical predictor of annual demographic variation, and seasonal anomalies can be justified by physiological or behavioural considerations (White et al. 1999, Catchpole et al. 2000). However, there is no consensus on which season affects vital rates the most. *A priori* expectations based on physiology, such as the importance of the growing season in plants, or the importance of exceeding minimum thermal limits in insects (Angilletta, 2009), are often met with opposing evidence (Kreyling 2010, Czachura and Miller 2020), or evidence of lagged effects from previous years (Hacket□Pain et al. 2018, Tenhumberg et al. 2018, Evers et al. 2021).

Ecologists need to simultaneously identify the critical time window and magnitude of climatic effects that drive their system dynamics. To address this need, several models have been proposed to automatically select the timing most relevant to population processes (van de Pol and Cockburn 2011, Ogle et al. 2015, Simmonds et al. 2019). Henceforth, we refer to these models as “antecedent effect models”. These models control for multiple comparisons, and weigh the time windows when drivers are most influential for population dynamics, dividing the

climate (or any other relevant factor) of a time window (*e.g.*, a year) into sub-periods of equal lengths – usually months. Such weighing can occur by estimating the importance of each monthly anomaly using weights, which can be constrained to follow a specific function (van de Pol and Cockburn 2011) or not (Ogle et al. 2015). Alternatively, the weighing can be performed by estimating a different effect size for each monthly climate anomaly, using regularization or informative priors to minimize overfitting (Hoerl & Kennard, 1970). These models have potential as both exploratory and predictive tools. The automatic estimation of the timing and magnitude of climatic effects can help generate hypotheses. Moreover, these models could increase predictive ability by reflecting the temporal processes that link climatic drivers to population dynamics.

Here, we systematically evaluate our performance to predict the out-of-sample variation of the vital rates of survival, development, and reproduction and the overall population growth rate ($\lambda$) of 77 natural populations of 34 plant species. We compare and contrast the predictive performance of three types of antecedent effect models to that of null models and models based on annual climatic anomalies. We use monthly temperature and precipitation as climatic drivers, and demographic information ranging from 7-36 years from the COMPADRE Plant Matrix Database (Salguero-Gómez *et al.* 2015). We evaluate the predictive performance of these five models to address three objectives: (1) evaluate the predictive performance of each different statistical model; (2) determine whether the predictive ability of antecedent effect models increase with temporal replication; and (3) compare the predictive ability among vital rates and population growth rate.

## Methods

*Data and response variables*

To compare the predictive performance of antecedent effect models to that of simpler models, we used long-term time series of vital rate data obtained from matrix population models (MPMs, hereafter) contained in the COMPADRE Plant Matrix Database (Salguero-Gómez et al. 2015). COMPADRE contains MPMs of natural populations where individuals have been categorized into discrete stages and their vital rates (*e.g.*, survival, development, reproduction) tracked in discrete time steps (Caswell 2001). We selected datasets from COMPADRE with at least seven contiguous years (*i.e.*, six MPMs) for a single natural population. We chose seven years as the shortest study length that allowed comparing a sufficient number of species: only half of plant demographic studies exceed four years in duration (Salguero-Gómez et al. 2015, Römer et al. 2021). This criterion resulted in 1,057 MPMs originating from 77 populations across 34 species, with an average length of 15 years (median = 10; range = 7-36), and an average number of two populations (median = 1, range=1-6). We note that in our analysis one species, *Cirsium pitcheri*, is repeated three times because each original study examined a different population, or a different set of populations (Supplementary Online Materials Table S1).

We used these MPMs to calculate the expectation of survival, development, reproduction, and the population growth rate ($\lambda$) associated with each annual MPM. By "development", we refer to changes in life cycle stage. These four year-specific metrics refer to a population at its stable stage distribution: the distribution of discrete stages that would result if each yearly MPM's vital rates were kept constant. We calculated survival, development, and reproduction using standard methods (Franco & Silvertown, 2004), and $\lambda$ as the dominant eigenvalue of each MPM (Caswell 2001) with the R package popbio (Stubben and Milligan 2007). Here, $\lambda$ > (or <) 1 corresponds to a population that is expected to grow (decline) at the stable stage distribution

(Caswell, 2001). We calculated net annual vital rates by taking a weighted mean of the stage-specific vital rates, where the weights were derived from the stable stage distribution.

We used gridded and weather station climatic data to generate the climatic predictors for our statistical models. The climatic predictors in our models include mean monthly air temperature and precipitation at the location of each of our 77 populations. We obtained these data from the CHELSA (Karger et al. 2017) and PRISM (Daly et al. 1994) datasets, and a single Global Historical Climatology Network (GHCN) daily weather station (Menne et al. 2012). Both CHELSA and PRISM provide gridded climatic data at a 30 arcsec (~1 $km^2$) resolution which approximate weather station climate averages well (Behnke et al. 2016), while providing a standardized source of climatic data. We used CHELSA data for North American populations sampled between 1979 and 2013 (Supplementary Online Materials Table S1). We used PRISM data for the two populations of *Astragalus scaphoides* which extend to 2014 (Tenhumberg *et al.* 2018). We used data from weather station number USC00143527 of the GHCN for the 10 populations collected between 1938 and 1972 in Hays, Kansas, USA (Supplementary Online Materials Table S1).

Our climatic predictors consisted of the monthly climatic anomalies observed in the 36 months preceding each vital rate observation (*i.e.*, preceding the end month of a given demographic projection interval). Specifically, vital rate observations describe a transition from year $t$ to year $t + 1$: our models use the 36 monthly anomalies preceding year $t + 1$ as predictors. These anomalies were $z$-scores (mean = 0, SD = 1) with respect to the long-term monthly mean. We calculated long-term means exceeding the recommended minimum of 30 years of data (World Meteorological Organization 2017). For CHELSA and PRISM datasets, we calculated the long-term monthly mean using a 35-year period between 1979 and 2013, and 1980 and 2014,

respectively. We calculated the long-term monthly mean for the 10 datasets collected before 1979 using 50 years of climate data. The difference between anomalies computed using the 50- versus 35-year mean was negligible, as their correlations were 0.996 and 0.995 for temperature and precipitation, respectively.

*Modeling overview*

Using vital rates and log($\lambda$) as response variables, we compared and contrasted the predictive performance of five models (Supplementary Online Materials: Fig. S1): the (1) Null Model (NM), with no climatic covariates, the (2) Climate Summary Model (CSM, Dalgleish et al., 2011), using yearly climate as a predictor, and three types of antecedent effect models: the (3) Weighted Means Models (WMM, van de Pol & Cockburn, 2011b), the (4) Stochastic Antecedent Model (SAM, Ogle et al., 2015), and the (5) Finnish Horseshoe Model (FHM, Hastie et al., 2009). Models 1-5 are ranked by increasing flexibility, by progressively relaxing key assumptions (Supplementary Online Materials: Fig. S1). NMs are a “model of the mean” that do not account for the effect of climatic drivers. CSMs assume that the monthly anomalies observed within a specific period have all equal weight and that these anomalies affect population dynamics in the same direction. Therefore, this model uses as a predictor the mean of 12 monthly anomalies. WMM's weigh the contribution of the 12 monthly anomalies based on a Gaussian function. For example, monthly anomalies could decrease in weight depending on their distance to the demographic census (Supplementary Online Materials: Fig. S1). Like in WMMs, SAMs assume that monthly anomalies influence population dynamics in the same direction. SAMs construct relative weights of monthly anomalies using coefficients drawn from a Dirichlet distribution (Supplementary Online Materials: Fig. S1; Ogle et al., 2015). Finally, FHMs fit a

separate coefficient for each of the 12 monthly anomalies, shrinking these toward zero through their prior distribution. In regularized horseshoe regressions, monthly anomalies can have positive or negative effects. FHMs were not developed to model climatic effects (Piironen and Vehtari 2017), though regularization approaches have been previously employed to estimate climatic effects on vital rate information (e.g. Tredennick et al., 2016).

We fit data from three separate years to account for the potential of lagged climatic effects. There is convincing evidence of lagged climatic effects in several case studies, including for three populations analyzed in this study: one population of *Cryptantha flava* (Evers et al. 2021) and two of *Astragalus scaphoides* (Tenhumberg et al., 2018). We re-fit each model using climate data from three separate years: each year corresponding to a subset of the 36 monthly anomalies preceding each demographic observation. We subdivided these 36 anomalies as belonging to the year (year *t*), one year (*t*-1), and two years (*t*-2) preceding each demographic observation. Hence, we fit 13 models for each vital rate or estimation of population growth rate and climate variable combination.

Below, we describe the equations of the models we tested assuming a normal response variable, log($\lambda$). To facilitate model comparison and model convergence, we used weakly informative priors for the intercept and variance terms (Lemoine 2019). We document the priors for all models and the prior predictive checks we ran in the section of the Supplementary Online Materials on prior predictive checks. Our null is a model of the mean, such that

$$y_{ti} \sim Normal(\hat{y}, \sigma) \qquad (1a)$$

$$\hat{y} = \alpha, \qquad (1b)$$

$$\alpha \sim Normal(0,0.5), \qquad (1c)$$

$$\sigma \sim Gamma(0.01,0.01), \qquad (1d)$$

where $y_{ti}$ refers to a log($\lambda$) value, $t$ refers to the year of observation, $i$ to a replicate within years ($i$ exceeds one in datasets that are spatially replicated), $\hat{y}$ is the mean model prediction, and $\alpha$ is a model intercept, the mean. In the CSM, we model the response variable as a function of $\bar{x}_t$, the mean of the 12 monthly anomalies (z-scores) over year $t$, as

$$\hat{y}_t = \alpha + \beta\bar{x}_t, \qquad (2a)$$

$$\beta \sim Normal(0,1), \qquad (2b)$$

where $\beta$ is the coefficient for the effect of $\bar{x}_t$, and $\hat{y}$ is the mean model prediction referred to year $t$.

The first antecedent effect model is the WMM, which models the response as a function of $x_{tk}$, the monthly temperature deviations for year $t$ and month $k$:

$$\hat{y}_t = \alpha + \beta \sum_k w_k x_{kt} \qquad (3a)$$

$$w_k = g_k / \sum_k g_k \qquad (3b)$$

$$g_k = \frac{1}{\sigma_w\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_w}\right)^2} \qquad (3c)$$

$$\mu \sim Normal(6.5,12)[0,12], \qquad (3d)$$

$$\sigma_w \sim Normal(1,12)[1,], \qquad (3e)$$

where the 12 $x_k$ monthly anomalies prior to the end of the relevant demographic year, are each weighted by $w_k$. In turn, $w_k$ is a normalized vector of Gaussian weights, $g_k$, with mean $\mu$ and scale $\sigma_w$. The square brackets in Eq. 3d and 3e show the limits beyond which the potential values of these parameters are truncated.

The SAM (Ogle et al. 2015) adds flexibility to the WMM above,

$$\hat{y}_t = \alpha + \beta \sum_k w_k x_{kt}, \qquad (4a)$$

$$\boldsymbol{w} = \text{Dirichlet}\left(\alpha, \overbrace{\ldots, \alpha}^{n=12}\right), \qquad (4b)$$

where vector $\boldsymbol{w}$ contains 12 weights that sum to 1, and $w_k$ is a single weight that refer to monthly anomaly $x_{kt}$ observed in month $k$ and year $t$. We set the concentration parameter ($\alpha$) to 1, corresponding to a uniform distribution on the 12-dimensional simplex of weights. The added flexibility originates because the values in vector $\boldsymbol{w}$ do not follow a functional form (albeit they are correlated). Henceforth, we refer to the term $\sum_k w_k x_{kt}$, in both Eq. 3a and 4a, as the "antecedent" of year $t$.

Finally, the FHM is a multiple regression whose coefficients are shrunk via the "Finnish" horseshoe prior proposed by Piironen & Vehtari (2017). These models estimate a separate slope value, $\beta_k$, for each $k$ month. The Finnish horseshoe prior prevents overfitting by shrinking estimates towards zero. This model is fit as follows:

$$\hat{y}_t = \alpha + \sum_k \beta_k x_{kt}, \qquad (5a)$$

$$\beta_k \sim Normal\left(0, \tau\tilde{\theta}_k\right), \qquad (5b)$$

$$\tilde{\theta}_k = \frac{c\theta_k}{\sqrt{c^2+\tau^2\theta_k^2}}, \qquad (5c)$$

$$\theta_k \sim HalfCauchy(0,0.1), \qquad (5d)$$

$$c^2 \sim InvGamma\left(\frac{v}{2}, \frac{v}{2}s^2\right), \qquad (5e)$$

$$\tau \sim HalfCauchy(0, \tau_0), \qquad (5f)$$

$$\tau_0 = \frac{m_0}{M-m_0}\frac{\sigma}{\sqrt{N}}. \qquad (5g)$$

Here, 12 $\beta_k$ values depend on a normal distribution whose standard deviation is controlled by $\tau$, the "global scale", and 12 $\tilde{\theta}_k$ "local scales". The global scale $\tau$ tends to shrink every monthly estimate $\beta_k$ toward zero. On the other hand, the $\tilde{\theta}_k$ values allow the $\beta_k$ monthly estimates to become relatively large. In the FHM, the combination of $\tau$ and $\tilde{\theta}_k$ shrinks all $\beta_k$ coefficients toward zero, while controlling the magnitude of $\beta_k$ values for which there is evidence for a

substantial departure from zero. We modified these models to fit survival, development, and reproduction data, which are non-normally distributed. We describe these modifications in the Supplementary Online Material.

*Model fitting*

We fit models using a Hamiltonian Monte Carlo sampler via the *R* package *rstan* (Stan Development Team 2018). We fit models running four parallel chains for 4,000 iterations, with a burn-in of 1,000 iterations, and thinning every other iteration. Because we had 13 alternative models, four demographic responses (log($\lambda$) and three vital rates), two climatic predictors, and 36 datasets, we fit a total of 3,744 models. Fitting these many models made it unfeasible to check models one at a time. Therefore, we opted to perform model checks across model types, response variables, and dataset length. In the Supplementary Online Material, we provide the methods and results to assess the convergence of the Hamiltonian Monte Carlo samplers and assess model fit (Bayesian model checks).

*Model comparison*

We compared model performance for each vital rate, climate predictor, and species combination. We performed comparisons based on the log pointwise predictive density (LPPD) (Bernardo and Smith 2009, Hooten and Hobbs 2015) calculated through a leave-one-year-out (LOYO) cross-validation (e.g. Tredennick et al., 2016). A LOYO cross-validation creates a training set by leaving out an entire year of data rather than a single data point. The higher the value of LPPD of the held-out data, the better the predictive performance of the model. Our model selection focused on LOYO cross-validation, because we aim to predict year-level

variation in vital rates. We performed this procedure because the climate is strongly correlated among different sites in our spatially replicated datasets. In particular, 16 of our 36 datasets are spatially replicated, and their mean correlation is 0.98 (range = 0.74-1).

We compared model performance by plotting the difference in ΔLPPD — the difference in LPPD with the best model. The $\Delta LPPD_i$ of model *i* is calculated as LPPD*i* – max(LPPD). For each species, we plotted heat maps of the $\Delta LPPD_i$ values for each model, flagging the best three models. We then ordered heat maps by data length along the y-axis, and by the year of climate data (*t*, *t*-1, and *t*-2) on the x-axis.

We addressed our three objectives by testing for differences in the relative support of our models using Pseudo-Bayesian Model Averaging. We calculated model weights using Pseudo-Bayesian Model Averaging for each combination of species, response variable, and predictor (temperature or precipitation). We calculated model weights following equations 8 and 9 in Yao et al. (2018). LPPD relative weights using the same formula used for AIC weights, so that

$$w_k = \frac{\exp\left(LPPD_k - \frac{1}{2}\mathrm{se}(\mathrm{LPPD_k})\right)}{\sum_{k=1}^{K}\exp\left(LPPD_k - \frac{1}{2}se(LPPD_k)\right)} \qquad (8)$$

where *k* refers to each model, *se* is the standard error, and *K* is 13, the total number of models tested for each combination of species, response variable, and climatic predictor. We arcsine transformed these LPPD weights to make them normally distributed. We then used these acrsine transformed LPPD weights as the response variable to address our three objectives. We addressed these objectives using graphs and average statistics. We plotted boxplots of model weights across all datasets, grouping them based on model types. This evaluated (1) whether antecedent effect models outperformed NMs and CSMs, and which antecedent effect model had the best predictive performance. Then, we plotted model weights against the temporal replication

of each dataset to (2) verify whether such replication increased the predictive ability of antecedent effect models. We fit quantile regressions on these plots using the R package quantreg (Koenker 2022), from the $10^{th}$ to the $90^{th}$ quantile in increments of 10. Finally, we plotted model weights of NMs grouping them by response variable to verify (3) which response variable climate is a better predictor of. Note that in this case, the lower the model weight of NMs, the higher the predictive ability of climate.

*Case study and exploratory analysis*

We present a case study to showcase the potential of antecedent effect models as exploratory and predictive tools. First, we show how antecedent effect models identify monthly climatic drivers of demographic responses, and we discuss how these inferences can inform biological hypotheses. Second, we show how antecedent effect models improve model fit. Here, we present data from five natural populations of *Astragalus tyghensis*, an endemic legume plant species with an extremely restricted range encompassing arid lands of Western North America (Thorpe & Kaye, 2008). We present models fit on the reproduction of *Astragalus tyghensis* using precipitation anomalies one year prior to the last demographic observation (year *t*-1). This climatic predictor had high predictive power in our model comparisons.

We use this case study to show for each model its fit to data and the effect sizes of monthly climatic anomalies. To show model fit to data, we plotted each raw reproduction data point against 200 randomly selected posterior samples predictions. To display the monthly effect sizes, we produced estimates which were directly comparable among CSMs, WMMs, SAMs, and FHMs. For the CSM, we divided the regression slope, $\beta$ (Eq. 2a), by 12, the number of months.

For the WMMs and SAMs, we multiplied the regression slope, $\beta$ (Eq. 3-4a), by the 12 monthly weights, $w_k$ (Eq. 3-4b). Finally, the FHMs directly estimate 12 $\beta_k$ separate regression slopes. We represented these monthly effect sizes by plotting the median and the 95% credible intervals of the posterior on the y-axis, against the month number on the x-axis.

## Results

*Model comparison*

We found no evidence that antecedent effect models have higher predictive power than climate summary models. When comparing LPPD weights, the NMs were on average the best performing, followed by CSMs, and then antecedent effect models (Fig. 1). The variance in model weights was smaller in NMs models, while the models including climatic predictors could reach very high or very low model weights (Fig. 1). This variance in the LPPD weights suggests that estimating climate effects is advantageous for a subset of populations which are presumably particularly sensitive to climate. However, estimating climate effects can substantially decrease predictive ability in the other populations. The best performing antecedent effect models were the FHMs, followed by WMMs, and SAMs. However, these average differences were small, potentially reflecting, as shown in our case study, that these three models tend to provide similar information (Supplementary Online Materials: Fig. S9). The superior performance of FHMs is in part driven by the relatively good performance of these models in log(λ) models (Supplementary Online Materials: Fig. S10-12).

The temporal replication of the studies did not improve average model performance, but it decreased the likelihood of fitting models with very low LPPD relative weights. The only

quantile regression that shows a significant trend is linked to the 10$^{th}$ quantile, for which temporal replication increases LPPD relative weights (Supplementary Online Materials: Fig. S13). This result suggests that sample size only limits the negative effects of fitting models with large numbers of parameters on predictive ability.

Models including climate tended to predict development better than the other response variables (Fig. 2). This higher predictive ability was due to the higher predictive performance of CSM models (Supplementary Online Materials: Fig S10-12). Moreover, models including climate predicted log($\lambda$) worse than the other response variables (Fig. 2).

*Case study and exploratory analysis*

The results of antecedent effect models were biologically interpretable, but they did not substantially improve the ability to predict the reproduction of *Astragaluys tygensis*. All four models show that precipitation is negatively associated with reproduction. According to the WMM, the monthly anomalies preceding a demographic census tend to be the most important in driving reproduction. The SAM and FHM provide more precise insights, indicating that the most important climate windows are located two (May) and six (January) months prior to the demographic census. Thus, antecedent effect models indicate that the demography of this species is correlated to the precipitation of the growing and dormant seasons. However, the model fits of antecedent effect models and the CSM are similar (Supplementary Online Materials: Fig S9). Hence, despite the predictive performance of antecedent effect models is slightly better in this case, this is a quantitatively small difference.

## Discussion

Understanding how climate affects the vital rates (survival, development, reproduction) of natural populations and their overall performance (*e.g.*, population growth rate, $\lambda$) is a core ecological question. Contrary to our expectations, we found that simple climate summary models (CSMs) showed, on average, slightly better predictive ability than the substantially more complex antecedent effect models. Moreover, we found dataset length had little role in modulating these results. Therefore, we advocate using antecedent effects models as exploratory tools to facilitate hypothesizing the windows of time most strongly associated with demographic variation. Finally, we found predicting climatic effects is easier for development and harder to for population growth rate. This last result suggests that development is more responsive to relatively short-term climatic drivers, and that species responses to climate might reflect "demographic buffering".

Our initial expectation was that antecedent effect models would improve predictions because they accommodate a variety of potential links between demography and climate (Kreyling 2010, Evers et al. 2021). Our lack of support for this expectation could be due to the small temporal length of most of our datasets. Selecting the most accurate model among a candidate set trades off the bias and variance of model predictions (Hastie et al. 2009). In small datasets, complex models with many parameters are likely estimated with high uncertainty (variance). Bayesian regularization aims to address this problem by limiting the prior variance of the model parameters. Nonetheless, a simple model with few parameters may achieve similar levels of predictive accuracy when its structural bias is sufficiently low. Our case study suggests another reason why antecedent effect models may not outperform simpler models: the monthly precipitation anomaly with the strongest effect, May, was positively correlated with the annual

precipitation anomaly. Such correlations will reduce the predictive performance gained through the additional flexibility of the antecedent effect models.

The case study exemplified the exploratory power of antecedent effect models. In *Astragalus tyghensis*, both the SAM and FHM show a strong negative association between reproduction and the precipitation anomalies that occurred in the May and January of the preceding year. These results raise three hypotheses. First, such negative association with precipitation anomalies occurring a year prior to a demographic observation could result from indirect effects mediated through competition (Suttle et al. 2007, Adler et al. 2012) or physiological tradeoffs (*e.g.*, costs of reproduction, Crone et al., 2009; Miller et al., 2012). Second, the correlation with the precipitation in January, a month when the mean temperature is just above or below 0° C, could reflect the effect of snowfall. This effect is often distinct from rain (Dalgleish et al. 2011, Compagnoni and Adler 2014) and could arise in different ways: for example, because snow affects pathogens (Smull et al. 2019), it protects plants from frost (Inouye, 2000), or because snowmelt provides a large pulse of resources (Noy-Meir 1973). Third, the large effect of May precipitation could reflect a key developmental phase occurring either in May or June. All three hypotheses call for further tests to be performed via analysis of existing data or field experiments.

We hypothesize that the higher predictability of development with respect to other vital rates is a general pattern in plants. First, predicting log($\lambda$) should be more challenging, because evolution should favor buffering its variation through time (Hilde et al. 2020). Evidence of such buffering is typically tested, across species, using data on the temporal variability of vital rates (McDonald et al. 2017, Le Coeur et al. 2022). Here we provide the first evidence, to our knowledge, that across many species, demographic buffering is reflected in the link between

climatic drivers and vital rates. Second, understanding how climate affects plant survival is still a rapidly developing field (Trugman et al. 2021). For example, the survival of plants to drought is specific to the species, plant functional type, and a large number of contextual factors (Volaire 2018). Third, understanding the response of reproduction to climate is not straightforward: seed production reflects indirect effects of development (Clark et al. 2021), and the transition from seed to new recruits is a source of uncertainty (Paniw et al. 2017). On the other hand, the strong dependence of development to climate is posited by physiology and demonstrated by ecosystem models (Schulze et al. 2019).

Our experience building and comparing antecedent effect models informs our suggestions on hyperparameter priors and emphasizes the role of prior predictive checks. The choice of hyperparameter priors has strong effects on model fitting and performance. In weighted mean models (WMMs), truncated hyperparameter priors (Eq. 3d-e) prevented the sampler from exploring unrealistic parameter values, greatly facilitating model convergence. In SAMs, we chose a vector of ones as the Dirichlet concentration parameter (Eq. 4b) to not favor any particular lag, and to ensure that prior weights had a uniform distribution. A much lower number than 1 would favor models with a single non-zero weight. In FHMs, we suggest using a standard deviation of 0.1 for the prior on monthly effect sizes (Eq. 5d). Using a a standard deviation of 1 (*e.g.*, Piironen & Vehtari, 2017) generated unrealistically high variation in our prior predictive checks (Supplementary Online Materials: prior predictive checks). More generally, prior predictive checks were essential because our initial prior choices produced very different variance in predictions among model types and response variables. Moreover, priors generating unrealistically high variance had model convergence issues and lower predictive performance.

Our results justify the continued data and conceptual efforts to understand and predict the

effects of climatic drivers on survival and reproduction. Compared with development, the link between climate and these vital rates is likely more complex. This complexity could in part be understood using antecedent effect models as hypothesis generation tools. However, improvements in predictive ability will rely on datasets with larger replication, especially temporal replication. Moreover, conceptual breakthroughs such as those regarding the importance of indirect effects (Clark et al., 2021; Hacket□Pain et al., 2018), might also prove useful.

## Acknowledgements

AC and TMK were supported by the Alexander von Humboldt Foundation under the framework of the Humboldt Professorship, and by the Helmholtz Recruitment Initiative of the Helmholtz Association (both to TMK); RSG was supported by NERC (NE/M018458/1). This work emanated from research discussions during the sAPROPOS (Analysis of PROjections of POpulationS) working group led by RSG and TMK at the German Centre for Integrative Biodiveristy Research (iDiv), which is funded by the German Research Foundation (FZT-118-202548816). We thank the Max Planck Institute for Demographic Research for the support of the COMPADRE Plant Matrix Database and the hundreds of plant population ecologists who have contributed with demographic data to COMPADRE. We thank B. Teller and P. Barks for discussions in the early phases of this project.

**Author Contributions**

RS-G and AC conceived the study; AC, RS-G, and DC designed the research; AC, performed the research, analyzed the data, and wrote the article; RS-G, DC, and TMK provided critical revisions of the article, and approved the final version of the article.

**Conflict of Interest statement**

We declare no conflict of interest

# References


Adler, P. B., H. J. Dalgleish, and S. P. Ellner. 2012. Forecasting plant community impacts of climate variability and change: when do competitive interactions matter? Journal of Ecology 100:478–487.

Angilletta, M. J. J. 2009. Thermal Adaptation: A Theoretical and Empirical Synthesis. OUP Oxford.

Behnke, R., S. Vavrus, A. Allstadt, T. Albright, W. E. Thogmartin, and V. C. Radeloff. 2016. Evaluation of downscaled, gridded climate data for the conterminous United States. Ecological Applications 26:1338–1351.

Bell, T. J., K. I. Powell, and M. L. Bowles. 2013. Viability model choice affects projection accuracy and reintroduction decisions. The Journal of Wildlife Management 77:1104–1113.

Bernardo, J. M., and A. F. M. Smith. 2009. Bayesian Theory. John Wiley & Sons.

Caswell, H. 2001. Matrix population models. Construction, analysis, and interpretation. Sinauer Associates, Sunderland, MA.

Catchpole, E. A., B. J. T. Morgan, T. N. Coulson, S. N. Freeman, and S. D. Albon. 2000. Factors influencing Soay sheep survival. Journal of the Royal Statistical Society: Series C (Applied Statistics) 49:453–472.

Clark, J. S., R. Andrus, M. Aubry-Kientz, Y. Bergeron, M. Bogdziewicz, D. C. Bragg, D. Brockway, N. L. Cleavitt, S. Cohen, B. Courbaud, R. Daley, A. J. Das, M. Dietze, T. J. Fahey, I. Fer, J. F. Franklin, C. A. Gehring, G. S. Gilbert, C. H. Greenberg, Q. Guo, J. HilleRisLambers, I. Ibanez, J. Johnstone, C. L. Kilner, J. Knops, W. D. Koenig, G. Kunstler, J. M. LaMontagne, K. L. Legg, J. Luongo, J. A. Lutz, D. Macias, E. J. B. McIntire, Y. Messaoud, C. M. Moore, E. Moran, J. A. Myers, O. B. Myers, C. Nunez, R. Parmenter, S. Pearse, S. Pearson, R. Poulton-Kamakura, E. Ready, M. D. Redmond, C. D. Reid, K. C. Rodman, C. L. Scher, W. H. Schlesinger, A. M. Schwantes, E. Shanahan, S. Sharma, M. A. Steele, N. L. Stephenson, S. Sutton, J. J. Swenson, M. Swift, T. T. Veblen, A. V. Whipple, T.

481 G. Whitham, A. P. Wion, K. Zhu, and R. Zlotin. 2021. Continent-wide tree fecundity driven by
482 indirect climate effects. Nature Communications 12:1242.
483 Clark, J. S., L. Iverson, C. W. Woodall, C. D. Allen, D. M. Bell, D. C. Bragg, A. W. D'Amato, F. W. Davis, M.
484 H. Hersh, I. Ibanez, S. T. Jackson, S. Matthews, N. Pederson, M. Peters, M. W. Schwartz, K. M.
485 Waring, and N. E. Zimmermann. 2016. The impacts of increasing drought on forest dynamics,
486 structure, and biodiversity in the United States. Global Change Biology 22:2329–2352.
487 Compagnoni, A., and P. B. Adler. 2014. Warming, soil moisture, and loss of snow increase Bromus
488 tectorum's population growth rate. Elementa: Science of the Anthropocene 2:000020.
489 Compagnoni, A., S. Levin, D. Z. Childs, S. Harpole, M. Paniw, G. Römer, J. H. Burns, J. Che-Castaldo, N.
490 Rüger, G. Kunstler, J. M. Bennett, C. R. Archer, O. R. Jones, R. Salguero-Gómez, and T. M. Knight.
491 2021. Herbaceous perennial plants with short generation time have stronger responses to
492 climate anomalies than those with longer generation time. Nature Communications 12:1824.
493 Crone, E. E., E. Miller, and A. Sala. 2009. How do plants know when other plants are flowering? Resource
494 depletion, pollen limitation and mast-seeding in a perennial wildflower. Ecology Letters
495 12:1119–1126.
496 Czachura, K., and T. E. X. Miller. 2020. Demographic back-casting reveals that subtle dimensions of
497 climate change have strong effects on population viability. Journal of Ecology 108:2557–2570.
498 Dalgleish, H. J., D. N. Koons, M. B. Hooten, C. A. Moffet, and P. B. Adler. 2011. Climate influences the
499 demography of three dominant sagebrush steppe plants. Ecology 92:75–85.
500 Daly, C., R. P. Neilson, and D. L. Phillips. 1994. A Statistical-Topographic Model for Mapping
501 Climatological Precipitation over Mountainous Terrain. Journal of Applied Meteorology 33:140–
502 158.
503 Ehrlén, J., W. F. Morris, T. von Euler, and J. P. Dahlgren. 2016. Advancing environmentally explicit
504 structured population models of plants. Journal of Ecology 104:292–305.

Ellis, M. M., J. L. Williams, P. Lesica, T. J. Bell, P. Bierzychudek, M. Bowles, E. E. Crone, D. F. Doak, J. Ehrlén, A. Ellis-Adam, K. McEachern, R. Ganesan, P. Latham, S. Luijten, T. N. Kaye, T. M. Knight, E. S. Menges, W. F. Morris, H. den Nijs, G. Oostermeijer, P. F. Quintana-Ascencio, J. S. Shelly, A. Stanley, A. Thorpe, T. Ticktin, T. Valverde, and C. W. Weekley. 2012. Matrix population models from 20 studies of perennial plant populations. Ecology 93:951–951.

Ellner, S. P., D. Z. Childs, and M. Rees. 2016. Data-driven Modelling of Structured Populations: A Practical Guide to the Integral Projection Model. Springer International Publishing.

Elton, C. S. 1924. Periodic Fluctuations in the Numbers of Animals: Their Causes and Effects. Journal of Experimental Biology 2:119–163.

Evers, S. M., T. M. Knight, D. W. Inouye, T. E. X. 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 27:1927–1941.

Franco, M., and J. Silvertown. 2004. A Comparative Demography of Plants Based upon Elasticities of Vital Rates. Ecology 85:531–538.

Hacket-Pain, A. J., D. Ascoli, G. Vacchiano, F. Biondi, L. Cavin, M. Conedera, I. Drobyshev, I. D. Liñán, A. D. Friend, M. Grabner, C. Hartl, J. Kreyling, F. Lebourgeois, T. Levanič, A. Menzel, E. van der Maaten, M. van der Maaten-Theunissen, L. Muffler, R. Motta, C.-C. Roibu, I. Popa, T. Scharnweber, R. Weigel, M. Wilmking, and C. S. Zang. 2018. Climatically controlled reproduction drives interannual growth variability in a temperate tree species. Ecology Letters 21:1833–1844.

Harper, J. L. 1967. A Darwinian Approach to Plant Ecology. Journal of Ecology 55:247–270.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer Science & Business Media.

Hilde, C. H., M. Gamelon, B.-E. Sæther, J.-M. Gaillard, N. G. Yoccoz, and C. Pélabon. 2020. The Demographic Buffering Hypothesis: Evidence and Challenges. Trends in Ecology & Evolution 35:523–538.

Hoerl, A. E., and R. W. Kennard. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12:55–67.

Hooten, M. B., and N. T. Hobbs. 2015. A guide to Bayesian model selection for ecologists. Ecological Monographs 85:3–28.

Inouye, D. w. 2000. The ecological and evolutionary significance of frost in the context of climate change. Ecology Letters 3:457–463.

Jolls, C. L., J. E. Marik, S. I. Hamzé, and K. Havens. 2015. Population viability analysis and the effects of light availability and litter on populations of Cirsium pitcheri, a rare, monocarpic perennial of Great Lakes shorelines. Biological Conservation 187:82–90.

Karger, D. N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, and M. Kessler. 2017. Climatologies at high resolution for the earth's land surface areas. Scientific Data 4:170122.

Karger, D. N., and N. E. Zimmermann. 2018. CHELSAcruts – High resolution temperature and precipitation timeseries for the 20th century and beyond. http://dx.doi.org/10.16904/envidat.159.

Kaye, T. N., and D. A. Pyke. 2003. The Effect of Stochastic Technique on Estimates of Population Viability from Transition Matrix Models. Ecology 84:1464–1476.

Knape, J., and P. de Valpine. 2011. Effects of weather and climate on the dynamics of animal population time series. Proceedings of the Royal Society B: Biological Sciences 278:985–992.

Koenker, R. 2022, May 2. quantreg: Quantile Regression.

Kreyling, J. 2010. Winter climate change: a critical factor for temperate vegetation performance. Ecology 91:1939–1948.

Le Coeur, C., N. G. Yoccoz, R. Salguero-Gómez, and Y. Vindenes. 2022. Life history adaptations to fluctuating environments: Combined effects of demographic buffering and lability. Ecology Letters 25:2107–2119.

Lemoine, N. P. 2019. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. Oikos 128:912–928.

McDonald, J. L., M. Franco, S. Townley, T. H. G. Ezard, K. Jelbert, and D. J. Hodgson. 2017. Divergent demographic strategies of plants in variable environments. Nature Ecology & Evolution 1:0029.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston. 2012. An Overview of the Global Historical Climatology Network-Daily Database. Journal of Atmospheric and Oceanic Technology 29:897–910.

Miller, T. E. X., J. L. Williams, E. Jongejans, R. Brys, and H. Jacquemyn. 2012. Evolutionary demography of iteroparous plants: incorporating non-lethal costs of reproduction into integral projection models. Proceedings of the Royal Society B: Biological Sciences 279:2831–2840.

Noy-Meir, I. 1973. Desert Ecosystems: Environment and Producers. Annual Review of Ecology and Systematics:25–51.

Ogle, K., J. J. Barber, G. A. Barron-Gafford, L. P. Bentley, J. M. Young, T. E. Huxman, M. E. Loik, and D. T. Tissue. 2015. Quantifying ecological memory in plant and ecosystem processes. Ecology Letters 18:221–235.

Paniw, M., P. F. Quintana-Ascencio, F. Ojeda, and R. Salguero-Gómez. 2017. Accounting for uncertainty in dormant life stages in stochastic demographic models. Oikos 126:900–909.

Piironen, J., and A. Vehtari. 2017. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electronic Journal of Statistics 11:5018–5051.

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 7:1246–1257.

van de Pol, M., and A. Cockburn. 2011. Identifying the Critical Climatic Time Window That Affects Trait Expression. The American Naturalist 177:698–707.

Römer, G., J. P. Dahlgren, R. Salguero-Gómez, I. M. Stott, and O. R. Jones. 2021, April 26. Plant demographic knowledge is biased towards short-term studies of temperate-region herbaceous perennials. bioRxiv.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, C. Bein, H. de Buhr, C. Farack, F. Gottschalk, A. Hartmann, A. Henning, G. Hoppe, G. Römer, T. Ruoff, V. Sommer, J. Wille, J. Voigt, S. Zeh, D. Vieregg, Y. M. Buckley, J. Che-Castaldo, D. Hodgson, A. Scheuerlein, H. Caswell, and J. W. Vaupel. 2016. COMADRE: a global data base of animal demography. Journal of Animal Ecology 85:371–384.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, Y. M. Buckley, J. Che-Castaldo, H. Caswell, D. Hodgson, A. Scheuerlein, D. A. Conde, E. Brinks, H. de Buhr, C. Farack, F. Gottschalk, A. Hartmann, A. Henning, G. Hoppe, G. Römer, J. Runge, T. Ruoff, J. Wille, S. Zeh, R. Davison, D. Vieregg, A. Baudisch, R. Altwegg, F. Colchero, M. Dong, H. de Kroon, J.-D. Lebreton, C. J. E. Metcalf, M. M. Neel, I. M. Parker, T. Takada, T. Valverde, L. A. Vélez-Espino, G. M. Wardle, M. Franco, and J. W. Vaupel. 2015. The COMPADRE Plant Matrix Database: an open online repository for plant demography. Journal of Ecology 103:202–218.

Schulze, E.-D., E. Beck, N. Buchmann, S. Clemens, K. Müller-Hohenstein, and M. Scherer-Lorenzen. 2019. Dynamic Global Vegetation Models. Pages 843–863 *in* E.-D. Schulze, E. Beck, N. Buchmann, S. Clemens, K. Müller-Hohenstein, and M. Scherer-Lorenzen, editors. Plant Ecology. Springer, Berlin, Heidelberg.

Simmonds, E. G., E. F. Cole, and B. C. Sheldon. 2019. Cue identification in phenology: A case study of the predictive performance of current statistical tools. Journal of Animal Ecology 88:1428–1440.

Smull, D. M., N. Pendleton, A. R. Kleinhesselink, and P. B. Adler. 2019. Climate change, snow mold and the Bromus tectorum invasion: mixed evidence for release from cold weather pathogens. AoB PLANTS 11:plz043.

Stubben, C., and B. Milligan. 2007. Estimating and Analyzing Demographic Models Using the popbio Package in R. Journal of Statistical Software 22:1–23.

Suttle, K. B., M. A. Thomsen, and M. E. Power. 2007. Species Interactions Reverse Grassland Responses to Changing Climate. Science 315:640–642.

Teller, B. J., P. B. Adler, C. B. Edwards, G. Hooker, and S. P. Ellner. 2016. Linking demography with drivers: climate and competition. Methods in Ecology and Evolution 7:171–183.

Tenhumberg, B., E. E. Crone, S. Ramula, and A. J. Tyre. 2018. Time-lagged effects of weather on plant demography: drought and Astragalus scaphoides. Ecology 99:915–925.

Thorpe, A. S., and T. N. Kaye. 2008. Astragalus tyghensis: Actual vs. predicted population sizes. Page 18. Institute for Applied Ecology for Prineville District BLM.

Tredennick, A. T., M. B. Hooten, and P. B. Adler. 2016. Do we need demographic data to forecast plant population dynamics? Methods in Ecology and Evolution 8:541–551.

Trugman, A. T., L. D. L. Anderegg, W. R. L. Anderegg, A. J. Das, and N. L. Stephenson. 2021. Why is Tree Drought Mortality so Hard to Predict? Trends in Ecology & Evolution.

Volaire, F. 2018. A unified framework of plant adaptive strategies to drought: Crossing scales and disciplines. Global Change Biology 24:2929–2938.

White, M. A., S. W. Running, and P. E. Thornton. 1999. The impact of growing-season length variability on carbon assimilation and evapotranspiration over 88 years in the eastern US deciduous forest. International Journal of Biometeorology 42:139–145.

World Meteorological Organization. 2017. WMO Guidelines on the Calculation of Climate Normals. World Meteorological Organization. Geneva.

Yao, Y., A. Vehtari, D. Simpson, and A. Gelman. 2018. Using Stacking to Average Bayesian Predictive Distributions (with Discussion). Bayesian Analysis 13.

## Figure legends

**Figure 1**. Null models (NMs) perform better than models using climate as predictor. Box and whisker plots and point clouds showing the arcsine transformed model weights of the five model types for 34 species. Panels separate models fit using precipitation or air temperature as predictor variables. The middle line of the boxplots shows the median, the upper and lower hinges delimit the first and third quartile, and the whiskers extend 1.5 times beyond the first and third quartile. Each point refers to one of the 3,744 model fits, and colors refer to the response variable of each model.

**Figure 2.** Climate has higher predictive power for development than the other response variables. Box and whisker plots and point clouds showing the arcsine transformed model weights of the null models based on the type of response variable. In this plot, lower values reflect a higher predictive ability of the models including a climatic predictor. The middle line of the boxplots shows the median, the upper and lower hinges delimit the first and third quartile, and the whiskers extend 1.5 times beyond the first and third quartile. Points represent 288 model fits: one for each species, response variable, and predictor variable (Null models were only one of the 13 models we fit on data).

**Figure 1.**

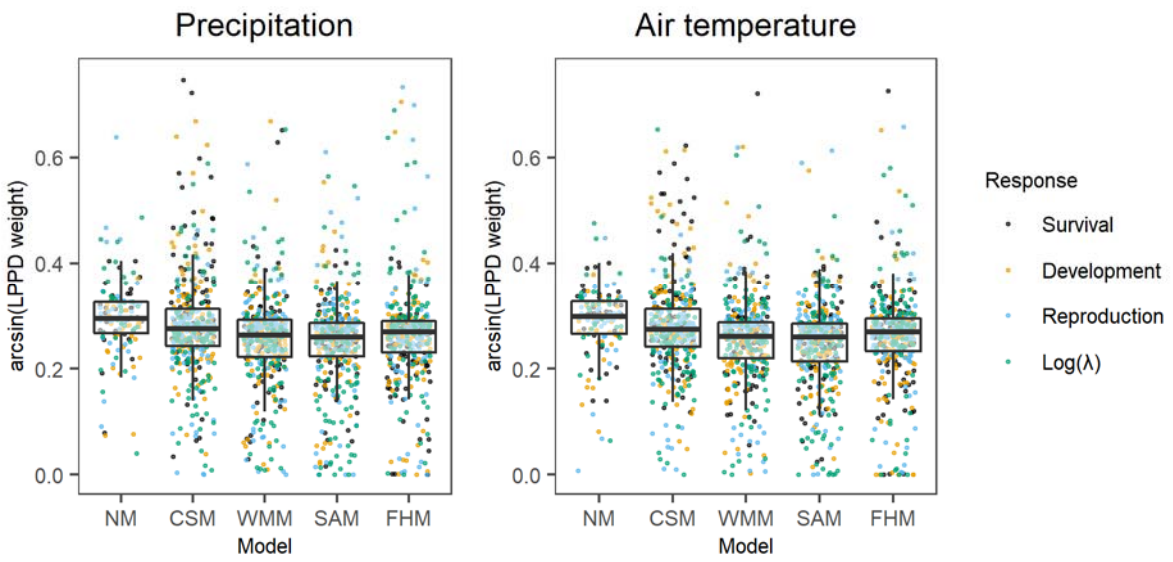
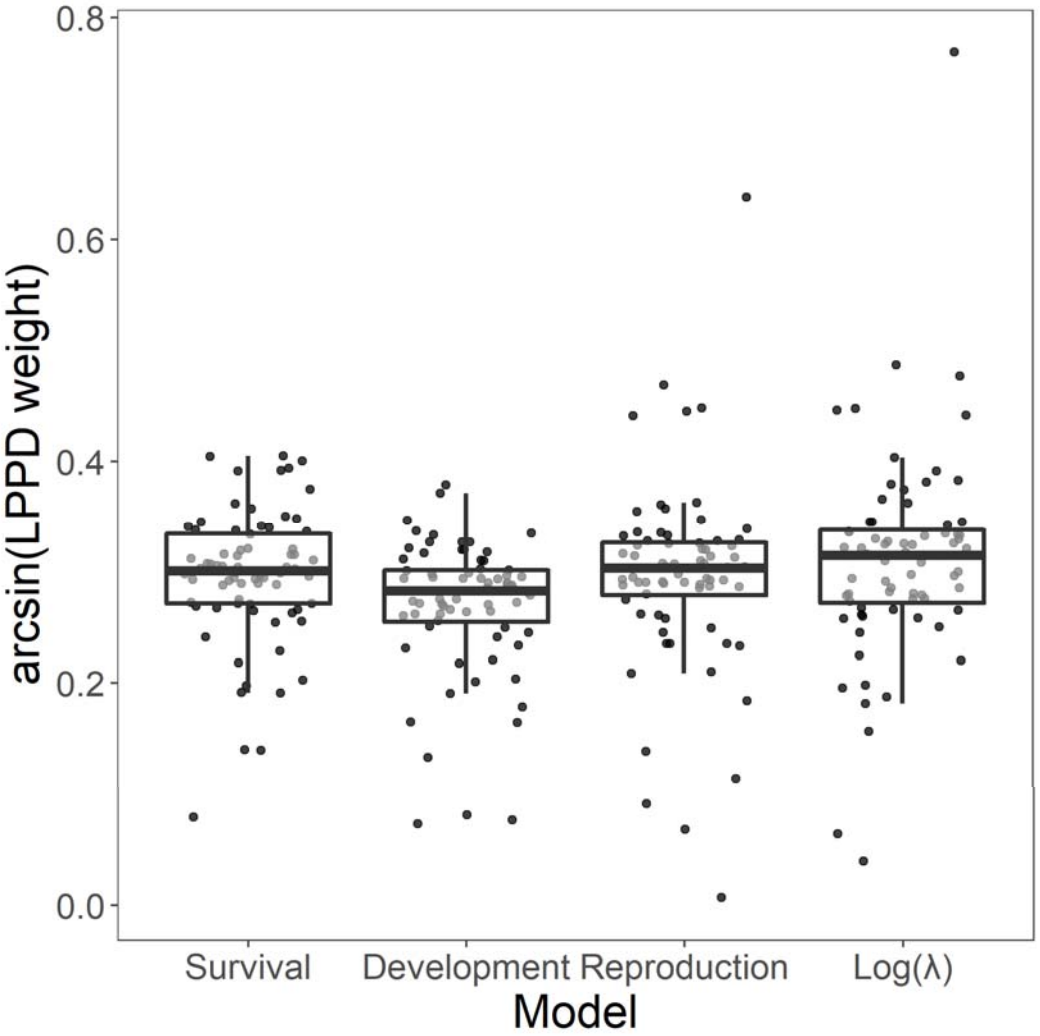

Precipitation
Air temperature
arcsin(LPPD weight)
0.0
0.2
0.4
0.6
NM
CSM
WMM
SAM
FHM
Model
Response
Survival
Development
Reproduction
Log(λ)

**Figure 2.**


0.8
0.6
0.4
0.2
0.0
arcsin(LPPD weight)
Survival
Development
Reproduction
Log(λ)
Model