









## METHOD

# Disentangling key species interactions in diverse and heterogeneous communities: A Bayesian sparse modelling approach

Christopher P. Weiss-Lehman<sup>1</sup>  | Chhaya M. Werner<sup>1</sup> | Catherine H. Bowler<sup>2</sup>  |  
Lauren M. Hallett<sup>3,4</sup>  | Margaret M. Mayfield<sup>2</sup>  | Oscar Godoy<sup>5</sup> | Lina Aoyama<sup>3,4</sup>  |  
György Barabás<sup>6</sup> | Chengjin Chu<sup>7</sup>  | Emma Ladouceur<sup>8,9</sup>  | Lorelee Larios<sup>10</sup> |  
Lauren G. Shoemaker<sup>1</sup> 

<sup>1</sup>Botany Department, University of Wyoming, Laramie, Wyoming, USA

<sup>2</sup>School of Biological Sciences, University of Queensland, Brisbane, Queensland, Australia

<sup>3</sup>Biology Department, University of Oregon, Eugene, Oregon, USA

<sup>4</sup>Environmental Studies Program, University of Oregon, Eugene, Oregon, USA

<sup>5</sup>Departamento de Biología, Instituto Universitario de Investigación Marina (INMAR), Universidad de Cádiz, Puerto Real, Spain

<sup>6</sup>Division of Theoretical Biology, Department of IFM, Linköping University, Linköping, Sweden

<sup>7</sup>Department of Ecology, State Key Laboratory of Biocontrol and School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong, China

<sup>8</sup>German Centre for Integrative Biodiversity Research (iDiv) Leipzig-Halle-Jena, Leipzig, Germany

<sup>9</sup>Department of Physiological Diversity, Helmholtz Centre for Environmental Research -UFZ, Leipzig, Germany

<sup>10</sup>Department of Botany and Plant Sciences, University of California Riverside, Riverside, California, USA

## Correspondence

Christopher P. Weiss-Lehman, Botany Department, University of Wyoming, Laramie, Wyoming 82071, USA.  
Email: cweissle@uwyo.edu

## Funding information

NSF EPSCoR, Grant/Award Number: EPS-2019528; Australian Research Council, Grant/Award Number: DPI40100574; Vetenskapsrådet, Grant/Award Number: 2017-05245; European Social Fund, Grant/Award Number: RYC-2017-23666; Deutsches Zentrum für integrative Biodiversitätsforschung Halle-Jena-Leipzig, Grant/Award Number: FZT 118 and 02548816; Ministerio de Economía y Competitividad

Editor: Frederick Adler

## Abstract

Modelling species interactions in diverse communities traditionally requires a prohibitively large number of species-interaction coefficients, especially when considering environmental dependence of parameters. We implemented Bayesian variable selection via sparsity-inducing priors on non-linear species abundance models to determine which species interactions should be retained and which can be represented as an average heterospecific interaction term, reducing the number of model parameters. We evaluated model performance using simulated communities, computing out-of-sample predictive accuracy and parameter recovery across different input sample sizes. We applied our method to a diverse empirical community, allowing us to disentangle the direct role of environmental gradients on species' intrinsic growth rates from indirect effects via competitive interactions. We also identified a few neighbouring species from the diverse community that had non-generic interactions with our focal species. This sparse modelling approach facilitates exploration of species interactions in diverse communities while maintaining a manageable number of parameters.

Christopher P. Weiss-Lehman and Chhaya M. Werner contributed equally.

Chhaya M. Werner co-first author.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Ecology Letters* published by John Wiley & Sons Ltd.

**KEY WORDS**

coexistence, environmental gradients, pairwise interactions, parameter shrinkage, plant fecundity, species diversity

## INTRODUCTION

Understanding what maintains the diversity of life—where and how species abundances change through time—has long fascinated and challenged ecologists. It is widely accepted that community composition in any given time and place is driven by the interplay of species interactions, responses to environmental conditions, and feedbacks between local and regional dynamics (Chesson, 2000; HilleRisLambers et al., 2012; Vellend, 2020). However, incorporating natural levels of diversity into community models is difficult, given that many of our classic community models build directly from population models, meaning that myriad pairwise-interactions must be estimated in diverse systems (Bulleri et al., 2016; Germain et al., 2018; Letten et al., 2018). This quickly leads to experimental feasibility, model overfitting, and parameter non-identifiability concerns, especially when also accounting for environmental covariates and their impacts on species' demography and biotic interactions. Arguably, the magnitude of this methodological limitation has even shaped our historical theoretical frameworks and empirical tests. For example, classic species trait trade-offs, such as the competition–colonisation trade-off, and classic metacommunity paradigms often apply for species pairs (Leibold et al., 2004; Levins & Culver, 1971; Tilman, 1982). Yet species interactions in diverse systems and the effects of environmental variability on species' demography may yield complex dynamics that dramatically alter population growth and coexistence dynamics (Allesina & Levine, 2011; Li et al., 2021; May & Leonard, 1975; Mayfield & Stouffer, 2017). The further development and empirical testing of these theories thus requires a statistical approach that is applicable in diverse communities and capable of identifying and incorporating key species interactions and environmental covariates.

To date, empirical studies of population dynamics and species coexistence frequently take one of two approaches for dealing with parameterisation limitations arising in diverse communities and varied environments. In the first approach, experimental studies focus on a few focal species. For example, Wainwright et al. (2019) examined coexistence based on pairwise interaction coefficients between four annual forbs in two locations and across two water availability treatments—a lofty number of species interaction coefficients to estimate—but still a relatively small subset of the community's full diversity (10–14 species in 0.09 m<sup>2</sup>; Dwyer et al., 2015). Finer-scale environmental variation can further limit the number of species that can be feasibly incorporated

given experimental constraints: in a study of grass and forb coexistence under variable rainfall regimes, Hallett et al. (2019) considered four rainfall conditions, requiring estimates of eight distinct species interaction coefficients even with only two species. Isolating pairwise species interactions across environmental conditions is a high barrier in species-rich communities, whether in the laboratory or field (Hallett et al., 2019; Letten et al., 2018).

In the second approach, often used with observational data, species are grouped into broad categories. At the most extreme, a single interaction coefficient is calculated between the focal species and all heterospecific individuals—regardless of their identity (Clark, Ann Turnbull, et al., 2020; Uriarte et al., 2004). Heterospecifics may also be grouped more finely, for example according to their taxonomic relationship (Uriarte et al., 2004) or origin status and life form (e.g. native versus exotic and grasses vs. forbs) (Martyn et al., 2020). Alternatively, functional groups can be created according to species' traits (e.g. specific leaf area, canopy height, seed number) (Kühner & Kleyer, 2008; Uriarte et al., 2004). However, this methodological approach often necessitates a priori knowledge of the system and makes an underlying assumption that species grouped together will interact similarly with each other and with the focal species. These assumptions are often not met (Mayfield & Levine, 2010), suggesting a need for a more parsimonious and robust methodology that allows the data to inform species groupings.

Various alternative statistical approaches have been proposed to assess species interactions in diverse communities using observational data. For example, joint species distribution modelling has become a common approach to infer species interactions from co-occurrence patterns (Legendre & Gauthier, 2014; Ovaskainen et al., 2017b, 2019). However, patterning in co-occurrence also results from environmental sorting and dispersal (Schamp et al., 2015), meaning with this approach species interaction strengths cannot be accurately inferred from observational patterns (Barner et al., 2018). Furthermore, these co-occurrence patterns are scale dependent and regional analyses are not suited to assessing local-scale species interactions (König et al., 2021). Recognising a need to directly estimate species interaction coefficients, recent work has expanded multivariate autoregressive models for use in more diverse communities (Picoche & Barraquand, 2020), including examining which linear combinations of species abundances best predict future growth rates (Ovaskainen et al., 2017a). This approach is effective for binning species based on their competitive effects, but does not account



for environmental variation. Clark, Scher, et al. (2020) developed a state-space hierarchical Bayesian model to assess the effect of environmental gradients on non-linear species abundance patterns, incorporating environmental responses in species' density-independent growth rates but not in species interactions (Clark, Scher, et al., 2020). Independently, these multiple methodological developments each address one of the largest hurdles for modelling species abundances in diverse communities: (1) identifying important species interactions and (2) accounting for the mediating effect of the environment (here referred to as species–environment interactions). Addressing these two aspects simultaneously would solidify a path forward for characterising species interactions in diverse communities and across environmental gradients.

Here, we present an approach for modelling dynamics in diverse communities and across environmental gradients. The approach balances realism and complexity without extensive experimental manipulation or *a priori* assumptions regarding species groupings. Our method is based on two innovations to standard population and community ecology models. First, we define heterospecific species interaction coefficients as linear combinations of the average interaction strength and species-specific deviations from this average. In parallel, we allow environmental covariates to modify species intrinsic growth rates and the strength of biotic interactions—both the average and species deviation terms. We implement this approach using a Beverton–Holt model of community dynamics (Beverton & Holt, 1957) within a single growing season, although the method can easily be adapted to other models of population abundance (e.g. Mayfield & Stouffer, 2017; Ricker, 1954) or incorporate additional dynamics such as seed banks or dispersal (Levine & HilleRisLambers, 2009; Thompson et al., 2020). Second, we extend Bayesian statistical methods for variable selection via sparsity-inducing priors in linear models (such as Lasso and Ridge regression; Hastie et al., 2015; Piironen and Vehtari 2017) to our non-linear abundance model, thereby reducing the number of terms included in the final model fit, yielding a 'sparse model'. By coupling these two modelling approaches, we can identify heterospecific species that deviate in their interaction strength, and how environmental gradients alter species' density-independent growth rates and biotic interactions. We explore model effectiveness using simulated data and apply the model to empirical data from a highly diverse (45 species) annual plant community.

## METHODS

### Deconstructing species interaction coefficients and fecundity

Models of community dynamics incorporate pairwise species-specific interaction coefficients, commonly denoted as  $\alpha_{i,j}$ , the effect of species  $j$  on species  $i$ , resulting in many parameters required to model diverse

communities or environmental relationships. To reduce the number of parameters required in such models, we start with a partitioning approach. We first define these interaction terms as follows:

$$\ln(\alpha_{e,i,j}) = \bar{a}_{0,i} + \hat{a}_{0,i,j} + (\bar{a}_{e,i} + \hat{a}_{e,i,j}) X_e, \quad (1)$$

where  $\alpha_{e,i,j}$  is the effect of species  $j$  on species  $i$  in environment  $e$  with  $i \neq j$ . In Equation (1),  $\bar{a}_{0,i}$  is the effect of an average heterospecific individual on individuals of species  $i$ ,  $\hat{a}_{0,i,j}$  is the deviation from this average effect associated with species  $j$ ,  $\bar{a}_{e,i}$  is the average slope of species  $i$ 's interaction coefficients with environmental covariate  $X_e$  and  $\hat{a}_{e,i,j}$  is the deviation from this slope associated with species  $j$ . At first, Equation (1) may seem counterproductive as it increases the number of parameters compared to traditional interaction coefficients. However, in the next section, we describe how coupling this approach with sparsity inducing priors in a Bayesian context can dramatically reduce the number of required parameters by identifying only the necessary species-specific terms ( $\hat{a}_{0,i,j}$  and  $\hat{a}_{e,i,j}$ ) for accurately modelling population dynamics of species  $i$ .

While intraspecific competition ( $\alpha_{e,i,i}$ ) could in principle be modelled according to Equation (1), we define it separately as follows:

$$\ln(\alpha_{e,i,i}) = a_{0,i,i} + a_{e,i,i} X_e, \quad (2)$$

where  $a_{0,i,i}$  and  $a_{e,i,i}$  are the intercept and slope for the effect of intraspecific individuals. As both theoretical expectations (Chesson, 2000) and empirical results (Adler et al., 2018) point to the importance of intraspecific competition, we use Equation (2) to explicitly exclude the intraspecific terms from the sparsity inducing process defined in the next section. These terms, therefore, are always included in the final model.

Interaction coefficients (Equations 1 and 2) can be incorporated in many different models of community dynamics. We use the Beverton–Holt model due to its legacy in studies of annual plant communities and coexistence theory (e.g. Godoy & Levine, 2014; Kraft et al., 2015). We emphasise, however, that our general statistical approach can be adapted to other population models. In the Beverton–Holt model, the fecundity  $F_{e,i}$  of focal species  $i$  in environment  $e$  is modelled as follows:

$$F_{e,i} = \frac{\lambda_{e,i}}{1 + \alpha_{e,i,i} N_{e,i} + \sum_{j \neq i}^S \alpha_{e,i,j} N_{e,j}}. \quad (3)$$

Fecundity depends on a species' intrinsic growth rate (i.e. density-independent seed production;  $\lambda_{e,i}$ ) and the competitive effects of all  $S$  species in the community ( $\alpha_{e,i,j}$  terms as defined by Equations 1 and 2) scaled by each

species' abundance ( $N_{e,j}$ ) (Levine & HilleRisLambers, 2009; Pérez-Ramos et al., 2019; Shoemaker & Melbourne, 2016). To incorporate environmental variation in intrinsic growth rates, we model  $\lambda_{e,i}$  as follows:

$$\ln(\lambda_{e,i}) = b_{0,i} + b_{e,i}X_e, \quad (4)$$

where  $b_{0,i}$  is the intercept of the intrinsic growth rate and  $b_{e,i}$  its slope with environmental covariate  $X_e$ . We use Equation (3) to model observed fecundity within a single growing season.

## Incorporating sparsity-inducing priors

By deconstructing interaction coefficients into a combination of species-specific and generic terms, we can determine which, if any, species-specific terms are necessary for the final model. Allowing only a subset of parameters to take non-zero values is referred to as 'sparse modelling' and various techniques exist to induce sparsity in linear models (Hastie et al., 2015; O'Hara and Sillanpää, 2009).

To extend a sparse modelling approach to our non-linear model of fecundity (Equation 3), we employ sparsity-inducing priors which act to shrink all but a subset of parameters to 0, thus producing a sparsely parameterised model. Specifically, we model  $\hat{a}_{0,i,j}$  and  $\hat{a}_{e,i,j}$ , the species-specific intercepts and slopes of the interspecific interaction coefficients (Equation 1), with regularised horseshoe priors which more accurately estimate large parameter values compared to other sparsity-inducing priors (Bhadra et al., 2019; Carvalho et al., 2009; Piironen and Vehtari, 2017; Van Erp et al., 2019). Parameters  $\hat{a}_{0,i,j}$  and  $\hat{a}_{e,i,j}$  are given priors  $\text{Normal}(0, \tau\tilde{\beta}_{0,i})$  and  $\text{Normal}(0, \tau\tilde{\beta}_{e,i})$ , respectively. (Note that since we fit the model for a single focal species, we drop the  $i$  subscript from the priors for simplicity). In these priors,  $\tau$  defines the global tendency towards sparsity through its effect on the priors' standard deviations. Essentially, smaller values of  $\tau$  will more tightly centre the priors for all  $\hat{a}_{0,i,j}$  and  $\hat{a}_{e,i,j}$  parameters on 0. Conversely, the  $\tilde{\beta}$  terms allow specific parameters to escape this global trend towards sparsity. As an individual  $\tilde{\beta}$  term becomes large, its associated prior becomes wider, and that species-specific term is more likely to be included in the final model. In the regularised horseshoe prior, these  $\tilde{\beta}$  terms are defined as follows:

$$\begin{aligned} \tilde{\beta}_j &= \frac{c\beta_j}{\sqrt{c^2 + \tau\beta_j^2}} \\ \beta_j &\sim \text{half-Cauchy}(0, 1) \\ c^2 &\sim \text{inverse-gamma}\left(\frac{\nu}{2}, \frac{\nu s^2}{2}\right). \end{aligned} \quad (5)$$

Defining  $\tilde{\beta}_j$  as the combination of a half-Cauchy and inverse-gamma distribution causes large coefficients to be shrunk towards 0 by a Student's  $t$  distribution with  $\nu$  degrees of freedom and a scale of  $s^2$  (Piironen and Vehtari, 2017; Van Erp et al., 2019). Following the recommendations of Piironen and Vehtari (2017), we set  $\nu$  to 4 and  $s^2$  to 2. Rather than setting the global shrinkage parameter  $\tau$  to a fixed value, we give it a half-Cauchy prior ( $\tau \sim \text{half-Cauchy}(0, 1)$ ) and allow the data to inform the posterior distribution of  $\tau$  (Piironen and Vehtari, 2017; Van Erp et al., 2019).

We employ a hybrid approach in which we first fit the full model with regularised horseshoe priors to induce sparsity in the species-specific terms; we subsequently fit a final model using traditional, non-sparse methods. From the preliminary model fit, we calculate credible intervals (CIs) for each species-specific term and identify terms with sufficient evidence for inclusion in the final model fit as those whose CIs do not overlap 0. Using this approach, we can directly adjust how conservative we wish to be in including model parameters, balancing model prediction, the proportion of variance explained, and simplicity depending on modelling goals (Tredennick et al., 2021) (i.e. using a 50% CI leads to more parameter inclusion than if we use a 95% CI; illustrated in Figure S1d–f). Thus, our model incorporates both the possibility of fully generic species interactions and of fully species-specific interactions, identifying the appropriate compromise between these two extremes that can best explain the data.

For the final model fit, included species-specific terms ( $\hat{a}_{0,i,j}$  and  $\hat{a}_{e,i,j}$ ) are given standard normal priors (i.e.  $\text{Normal}(0, 1)$ ). In both preliminary and final model fits, terms defining  $\lambda_{e,i}$  ( $b_{0,i}$  and  $b_{e,i}$ ; Equation 4) are also given standard normal priors. The intercept and slope terms defining intraspecific competition ( $a_{0,i,i}$  and  $a_{e,i,i}$ ; Equation 2) and the generic intercept and slope defining interspecific competition ( $\bar{a}_{0,i}$  and  $\bar{a}_{e,i}$ ; Equation 1) are given weakly informative priors in each model fit, matching the expected scale of these interaction coefficients:  $a_{0,i,i}$ :  $\text{Normal}(-6, 3)$ ,  $\bar{a}_{0,i}$ :  $\text{Normal}(-6, 3)$ ,  $a_{e,i,i}$ :  $\text{Normal}(0, 0.5)$  and  $\bar{a}_{e,i}$ :  $\text{Normal}(0, 0.5)$ . All models were fit using the stan language with the rstan package (version 2.18.2; Stan Development Team 2018 in R (version 3.5.3; R Core Team, 2019). All data and code for the analyses and simulations presented here can be found at <https://doi.org/10.5281/zenodo.5828361>.

## Simulation tests of model performance

To test our ability to predict changes in population size and recover true parameter values, we first paired our Bayesian sparse modelling approach with simulated Beverton–Holt data using Equations (3) and (4).

We generated communities of 15 species in different plots, where each plot was a unique run of the simulation with a given environmental condition  $X_e$ . These environmental conditions were drawn from a Gaussian distribution with mean 0 and standard deviation of 1 to represent any standardised environmental variable of interest. Species were assigned parameters defining intrinsic growth rates  $\lambda_{e,i}$  and intraspecific competition  $\alpha_{e,i}$  following Equations (4) and (2), respectively (Table S1). Similar to intraspecific competition, pairwise interspecific competitive interactions in the simulations were defined as  $\ln(\alpha_{e,i,j}) = a_{0,i,j} + a_{e,i,j}X_e$  where  $a_{0,i,j}$  terms were drawn from a uniform distribution and  $a_{e,i,j}$  terms were drawn from a normal distribution (Table S1). Each plot simulation was run deterministically for 20 time steps with each time step  $N_{t+1} = F_t N_t$  using  $F_t$  from Equation (3). This resulted in some subset of the 15 species remaining in each plot. Then each population was perturbed by drawing from a normal distribution with mean and standard deviation equal to the previous population size (i.e. scaling with population size), truncated at 0 to prevent negative population sizes. This perturbed state and the following time step generated our simulated ‘full-community’ data. In addition to 500 full-community plots, we simulated 500 ‘no-competition’ treatments with a single phytometer individual of the focal species per plot, running the Beverton–Holt function for one time step. This simulated treatment matches methods commonly used in experimental studies to parse intrinsic growth rates from competition parameters (Hallett et al., 2019; Wainwright et al., 2019). Simulation details are included in Supplement 1.

We used these simulations to measure our sparse modelling approach's ability to predict population growth in diverse communities and recover underlying parameters. For 1000 simulations, we randomly selected one focal species to use in fitting our statistical model (described above) using varying numbers of full-community and no-competition plots (10, 50, or 200). For each model fit, we calculated (1) the average deviation of each parameter as the difference between the posterior mean and the true value, (2) the number of species identified by the model as requiring species-specific intercepts and/or slopes, and (3) the model's predictive accuracy as measured by the root-mean-square error (RMSE) of model predictions compared to 200 simulated full-community plots not used to fit the model. This allowed us to assess model performance, both in terms of parameter accuracy and predictive power, when used with varying amounts of input data.

Additionally, we explored the effect of drawing interspecific competitive terms from distributions of varying widths to determine the performance of the model when fit to communities with different underlying competitive structures. We simulated data from a low, medium and high range of potential pair-wise

interaction values (Table S1) and fit the model to 150 independent simulations of 50 full-community and 50 no-competition plots for each range. For each model fit, we recorded the same three quantities described above.

## Empirical application

We additionally applied our model to species interactions and their environmental dependencies in the annual plant understory of the York gum (*Eucalyptus loxophleba* Benth)–jam (*Acacia acuminata*) woodlands of south-western Western Australia. This community is highly diverse and heterogeneous, with local composition of annual forbs and grasses influenced by gradients in soil nutrients and shade from York gum and jam trees (Dwyer et al., 2015; Lai et al., 2015). We focused on two York gum–jam woodland remnants: West Perenjori Nature Reserve (29°47'S, 116°20'E) and Bendering Nature Reserve (32°23'S, 118°22'E). Both sites experience a Mediterranean climate with mild winters and long, dry summers (Suppiah et al., 2007) and have high overlap in annual species composition. Data used here were originally collected as part of a larger experiment described in Wainwright et al. (2019). We focus on two species used in the original study and common to both reserves: *Waitzia acuminata*, an abundant native annual forb, and *Arctotheca calendula*, a prevalent exotic annual forb.

We used data from 11 experimental blocks in Bendering Nature Reserve and 18 blocks in West Perenjori Nature Reserve. Each block was  $\approx 15 \times 15$  m, a size selected to account for previously identified soil-nutrient turn-over rates (Dwyer et al., 2015). Each block was split into  $50 \times 50$  cm plots and each plot was further subdivided into four  $25 \times 25$  cm quadrats. One individual near the centre of each quadrat was assigned as the focal individual for that quadrat. Which focal species were in a given quadrat depended on the natural distribution of individuals. This experiment employed five thinning treatments at the plot level to manipulate local community compositions (individual focal individuals with no competitors, native dominated competitors, exotic dominated competitors, monocultures with only conspecific competitors and unmanipulated plots) (Wainwright et al., 2019). This ensured a range of densities in the background communities to inform model estimates of competition coefficients and intrinsic growth rates. Across both reserves we used data from 129 focal individuals in 69 plots interacting with 45 neighbouring species for *W. acuminata* and 95 focal individuals in 54 plots interacting with 40 species for *A. calendula*.

We applied our sparse modelling framework to quantify the competitive environment's effect on fecundity in *W. acuminata* and *A. calendula* under different environmental conditions. Fecundity  $F_{e,i}$  was measured as the number of flowers produced by each focal individual.

The competitive environment was characterised as the number of individuals of each interacting species in the quadrat after the experimental treatment had been applied ( $N_{e,j}$ ). We considered two aspects of the physical environment  $X_e$ : percent overhead tree canopy cover, measured at the plot scale, and soil Colwell P (mg/kg), measured at the block scale. Both environmental covariates were standardised for inclusion in the model. We ran separate models for each focal species and environmental covariate (four total model fits). To account for regional differences between the Bending and Perenjori reserves, we incorporated a fixed reserve effect into our sparse modelling approach by allowing  $b_{0,i}$ ,  $b_{e,i}$ ,  $\hat{a}_{0,i,j}$ , and  $\hat{a}_{e,i,j}$  to differ between reserves. Using this approach, we quantified  $\lambda_{e,i}$  and  $\alpha_{e,i,j}$  for both species across both environmental gradients in the York gum-jam woodland communities.

## RESULTS

### Simulations

Our model accurately predicted growth rates in simulated communities even with relatively low sample sizes (Figures 1 and 2) and across different model formalisations (Box 1). For a single simulation, growth rate predictions improved in both accuracy and precision with increasing amounts of input data (Figure 1b,d and f). When comparing across multiple simulations, the average RMSE of model predictions when using 10 full-community and 10 no-competition plots was only 0.43. While increasing sample size further increased model accuracy (average RMSE of 0.32 for 50 plots and 0.29 for 200 plots), these results indicate the model can accurately predict species' realised growth rates using limited data (Figure 2c). Furthermore, species' growth rates can be accurately predicted using observed competitive communities paired with no-competition plots, rather than necessitating common manipulative experimental designs where each possible species combination is paired across a gradient of densities (Hallett et al., 2019; Kraft et al., 2015).

Our model also accurately predicted individual parameter estimates for simulated communities, with relatively low deviations between the estimated and true values (Figures 1 and 2). Even with just 10 plots of each type, parameters relating to intrinsic growth rate ( $\lambda_{e,i}$ ), intraspecific competition ( $\alpha_{e,i,i}$ ) and the relationship of generic and species-specific competition with the environment ( $\bar{a}_{e,i}$  and  $\hat{a}_{e,i,j}$ , respectively) are estimated reasonably accurately on average (Figure 2). In contrast, the parameters relating to the intercepts of generic ( $\bar{a}_{0,i}$ ) and species-specific ( $\hat{a}_{0,i,j}$ ) competition are estimated with less accuracy when only using 10 plots. However, while all deviations in parameter estimates shrunk with increasing numbers of input plots, the intercept of generic

competition benefited the least from increasing amounts of data. This is likely because the model correctly identified a larger number of species-specific terms on average with more data (Figure 2b), decreasing the total number of species contributing to the estimation of the generic competition terms.

Parameter accuracy, number of species-specific terms estimated and predictive power of the model were similar regardless of the underlying distribution of pair-wise interaction values (Figure S2).

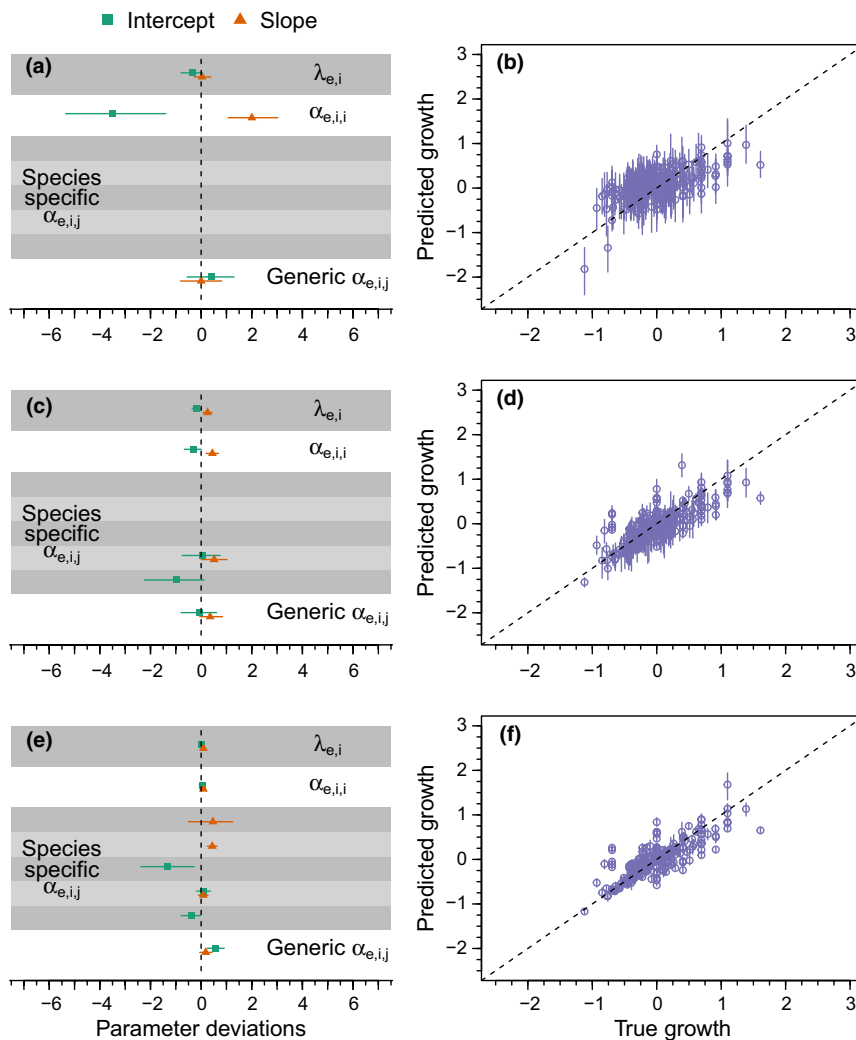
### Empirical application

Our method identified environmental dependencies in intrinsic growth rates (Figures 4a,b and 5a,b), relative strengths of intraspecific competition and average interspecific competition, along with competition–environment interactions (Figures 4c,d and 5c,d), all of which differed between our two focal species. Additionally, our model highlighted three species with deviations from the average interspecific effects on native *W. acuminata*, but no such species when fit to data on exotic *A. calendula*.

*W. acuminata* and *A. calendula*'s intrinsic growth rates differed in their relationship with the environmental gradients and reserves. The intrinsic growth rate of *W. acuminata* across both environmental gradients varied between the Bending and Perenjori reserves (Figure 4a,b). In contrast,  $\lambda_{e,i}$  for *A. calendula* was quite similar between the two reserves as it varied with both phosphorous and canopy cover (Figure 5a,b). This could reflect local adaptation in regional populations of the native *W. acuminata* but not the newly introduced *A. calendula*. Importantly, the intrinsic growth rate of *W. acuminata* declined with high phosphorous (marginally in Bending, but substantially in Perenjori) while *A. calendula*'s intrinsic growth rate increased with phosphorous, potentially explaining the high prevalence of invasive species in areas with increased phosphorous (Dwyer et al., 2015).

Relative competitive effects between conspecifics versus heterospecifics also differed between the two focal species. For *W. acuminata*, the relationship between intraspecific competition and average interspecific competition varied with the environmental gradients. At low levels of phosphorous and high levels of canopy cover, intraspecific competition in *W. acuminata* was greater than average interspecific competition (Figure 4c,d). However, at high levels of phosphorous and low levels of canopy cover, intra- and interspecific competition converged to similar values. In contrast, intraspecific competition for *A. calendula* was similar to or lower than generic interspecific competition across both environmental gradients (Figure 5c,d). This likely contributes to the invasive status of *A. calendula* in this ecosystem, whereas *W. acuminata* populations self-regulate under





**FIGURE 1** Fitted model parameter estimates and predicted growth rates. We fit the model to simulated data with 10 (a and b), 50 (c and d) and 200 (e and f) full-community and no-competition plots. The left column (a, c and e) shows the deviation of parameter values from the true value used in the simulations (points are posterior means and lines are 95% credible intervals). The right column (b, d and f) shows model accuracy of the focal species' growth rate for 200 simulated full-community plots not included in the model fitting. Growth rates were calculated as  $m\left(\frac{N_{t+1}}{N_t}\right)$ . The dashed line is the 1–1 line indicating a perfect match. Points show mean estimates and lines are 95% CIs

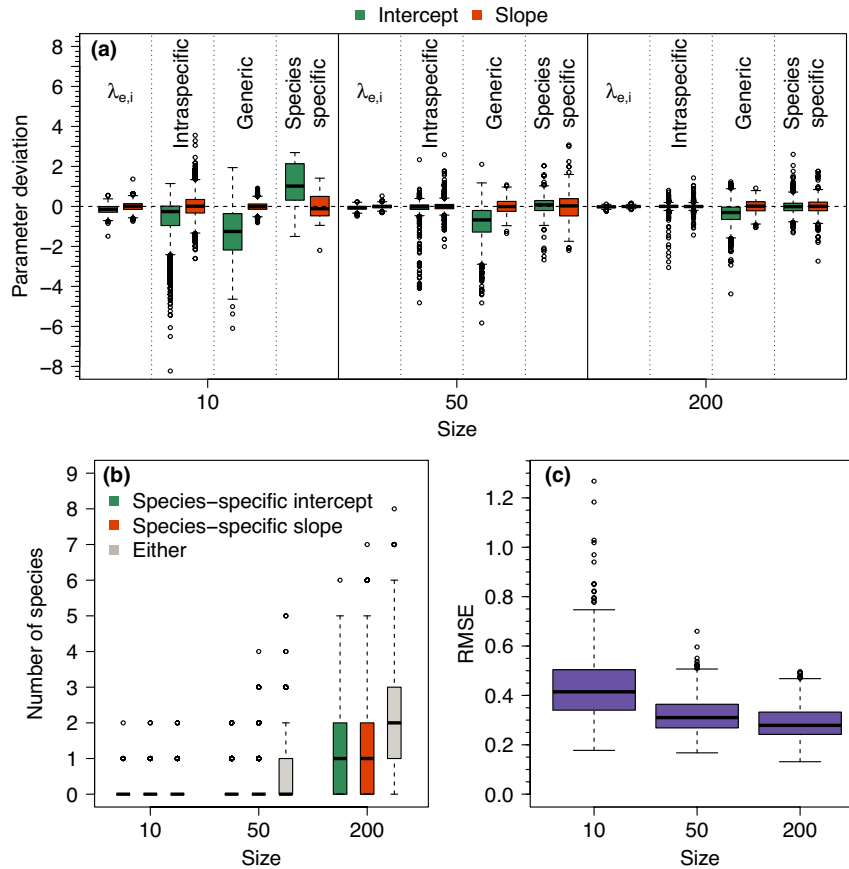
certain environmental conditions—a necessary component of stable coexistence.

Our model highlighted multiple species with non-generic competitive effects on *W. acuminata*. Across the gradient in phosphorous, *Hyalosperma glutinosum* had a higher than average effect on *W. acuminata* in Perenjori while *Schoenus nanus* had a lower than average effect in Bendering (Figure 4c). Across the gradient in canopy cover, *Hypochaeris glabra* had a much higher than average effect on *W. acuminata* in Bendering (Figure 4d). In contrast, all heterospecific interactive effects on *A. calendula* remained grouped in the generic competition term. The lack of species-specific effects on *A. calendula* (Figure 5c,d) could be due to its exotic status (Lai et al., 2015). With no shared evolutionary history in the community, *A. calendula* could be experiencing a form of competitive release, wherein the identity of competitor species matters less than simply the presence of additional individuals.

## DISCUSSION

Given the inherent complexity of ecological communities, ecologists are often forced to rely on simplifying assumptions to keep analyses tractable, for example limiting the number of species considered or ignoring environmental variation. The sparse modelling approach presented here provides an alternative method to analyse communities without requiring extensive additional data or sacrificing complexity. This approach enabled us to accurately predict population growth rates with limited data and identify how species' demographic rates and competitive interactions vary across environmental gradients. Our results identify environment by species interactions that deviate from the species-averaged community effects without making *a priori* assumptions about species interaction strengths or species groupings (Figure 4c,d). Output from the sparse modelling





**FIGURE 2** Model performance across 1000 independent simulations. For each simulation, we fit the model to sets of 10, 50 and 200 full-community and no-competition plots. (a) For each model fit, we quantified the parameter deviation as the difference between the posterior mean and the true value, such that positive values indicate an overestimate by the model. (b) We also recorded the number of species identified by each model fit as having a species-specific intercept, species-specific slope and the total number of species with either species-specific term identified by the model. For context, the maximum number of species that could be selected is 14 (out of 15 total species, leaving a single species to be defined by the generic parameters). (c) Finally, we calculated the root-mean-square error (RMSE) of the model predictions for 200 simulated full-community plots excluded from the model fitting. In all panels, standard Tukey box plots are used to show the distribution of results from all 1000 simulations

approach generates concrete, testable hypotheses about species interactions and environmental conditions. For example, does the number of non-generic parameters systematically differ for native and exotic species or with global change drivers? We see broad potential for this method's implementation in community ecology, from theory development to management applications.

The sparse modelling approach's flexibility allows easy adjustments for the best match between underlying model structure and the given study system and research questions. As shown in Box 1 and Figure 3, these models can successfully be applied to different forms of species–environment interactions and complexity can be modified based on motivating questions and data availability. For example, the functional form of the relationship between intrinsic growth rate and the environment likely depends on spatial scale. For localised studies, a simple monotonic relationship (Figure 3a) might be appropriate to capture species' expected responses across a small range of environmental variation. However, studies over larger spatial scales might require a functional form with

optimal intrinsic growth at an intermediate environmental value (Figure 3b), mimicking expected patterns of adaptation across species' ranges (Angert et al., 2020). Additionally, while we used a Beverton–Holt framework, sparse approaches could be used with different functional forms of competition (García-Callejas, 2020) or with models incorporating facilitative interactions (Stachowicz, 2001). Future work could expand beyond our consideration of plant fecundity, to instead model changes in biomass or plant cover depending on experimental design. Given the importance of temporal stochasticity to community dynamics (Shoemaker et al., 2020) and the need to predict community responses to changing anthropogenic pressures (Ma et al., 2017), sparse modelling with time series could provide invaluable insight into the importance of species-specific interactions through time and across space. We see future methodological expansions to additional biological models and response variables as exciting avenues to increase the generality of sparse approaches in community ecology.

### BOX 1 Adapting the sparse modelling method to different ecological questions

This sparse modelling method is generalisable to a variety of underlying ecological models. The method's flexibility allows researchers to use sparse approaches with different underlying ecological models depending on their study system and questions of interest.

For example, the relationship between species' growth rates and the environment can be modelled in multiple ways. A monotonic relationship would be appropriate for a study concentrated within a small spatial scale, while a humped-shape relationship would match expectations for a study over a broad environmental gradient. To demonstrate how our method can be modified for different underlying ecological models, we simulated environmental responses in growth rate two ways: with a monotonic relationship between species and the environmental conditions  $\lambda_{e,i}$  and with a curved environmental optimum with a defined niche breadth for each species  $\lambda_{e,i}^*$ . In the first model formulation:

$$\lambda_{e,i} = e^{b_{0,i} + b_{e,i} X_e}, \quad (6)$$

for which we estimate the mean intrinsic growth rate ( $b_{0,i}$ ) and the slope ( $b_{e,i}$ ) across an environmental gradient. In the second:

$$\lambda_{e,i}^* = b_{max,i} e^{-\left(\frac{z_i - X_e}{2\sigma_i}\right)^2}, \quad (7)$$

for which we estimate the maximum intrinsic growth rate ( $b_{max,i}$ ), the environmental optimum ( $z_i$ ) and the environmental niche breadth ( $\sigma_i$ ) (following the parameterisation in Thompson et al. (2020)). We tested these models using samples of 200 full community plots and 200 no-competition plots.

Both the monotonic and curved environmental optimum conditions are correctly captured by the sparse modelling approach (Figure 3a,b), with all growth rate parameters in both models falling within the 95% credible intervals. In the monotonic  $\lambda_{e,i}$  model, both the intercept  $b_{0,i}$  and the slope  $b_{e,i}$  deviated from the true values by less than 5%. In the optimum  $\lambda_{e,i}^*$  model, the maximum  $b_{max,i}$  and the niche breadth  $\sigma_i$  deviated from the true values by less than 1%, and the location of the environmental optimum  $z_i$  deviated by 47%, which is an absolute difference of 0.11.

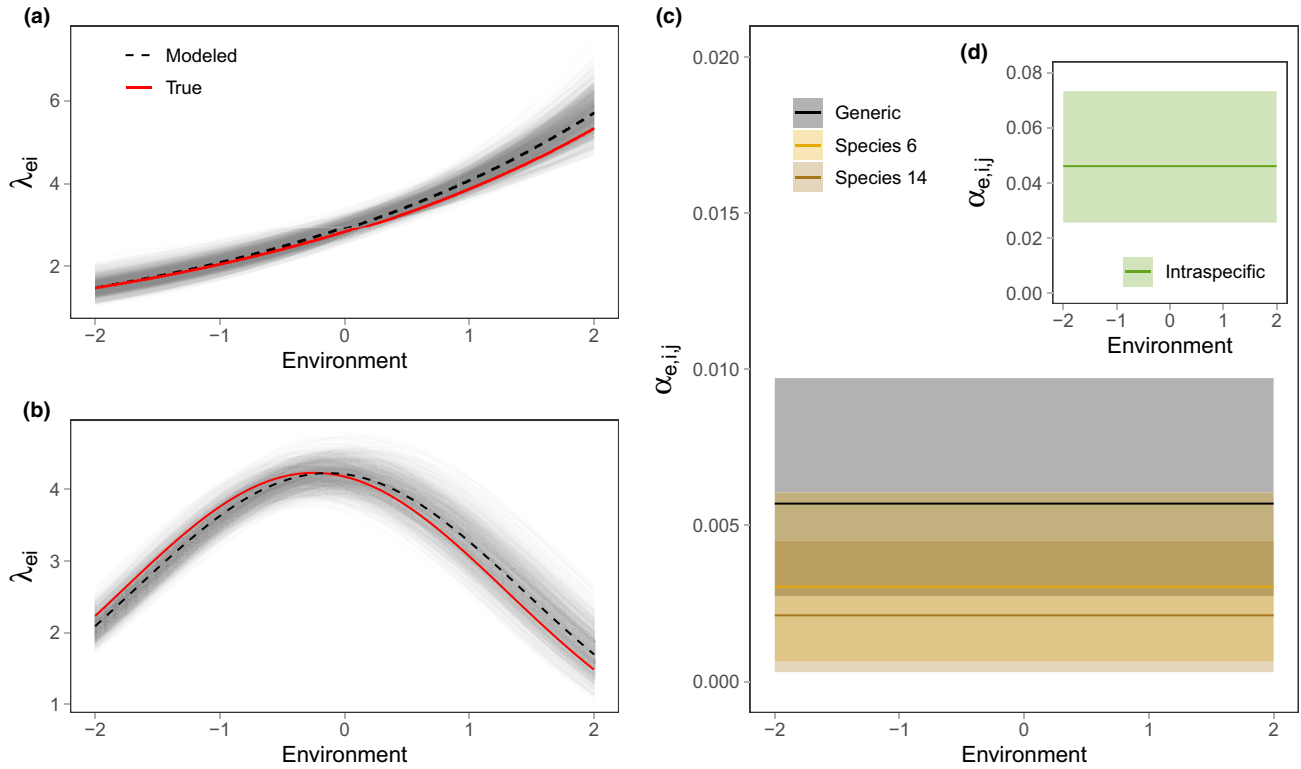
As a further example, the species interaction components of the model can be adjusted depending on the main research questions of interest. For questions focused on species interaction strengths, one may wish to model interaction coefficients independent of environmental conditions. We demonstrate this approach with a simple simulation with competitive interactions independent of environmental conditions  $\alpha_{i,j}$ , modelled accordingly (Figure 3c,d).

We present these options as launching-off points for researchers to adapt the sparse modelling approach to their study systems and questions. Even more extensive modifications are possible, such as replacing the Beverton–Holt community framework with a different underlying ecological model.

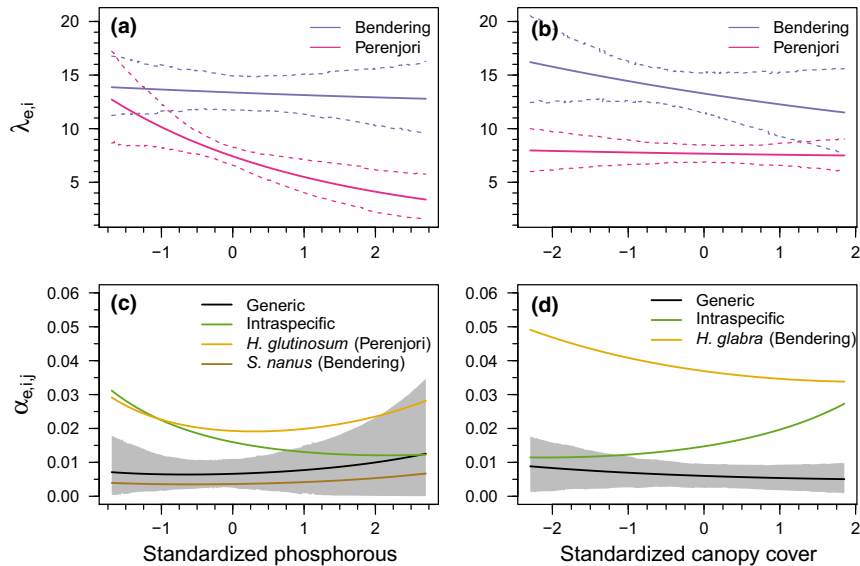
Sparse modelling can be a powerful tool to accelerate the development of community ecology theory and practice. It can provide important insights into the covariation of environmental conditions, species' demographic rates and competitive effects—critical aspects of modern coexistence theory (Chesson, 2000). This includes quantifying the relative strengths of intra- versus inter-specific competition, a key metric for stable coexistence (Adler et al., 2018; Chesson, 2000). Furthermore, the approach elucidates effects of environmental conditions on species' density-independent growth rates versus competitive interactions, potentially allowing for quantification of variation-dependent coexistence mechanisms, such as the storage effect, in diverse communities (Chesson, 2000; Spaak

et al., 2021). Additionally, sparse model fits across environmental gradients can be used to quantify the relative importance of environmental (abiotic) filtering, biotic interactions, and their joint effect on species occurrence (Cadotte & Tucker, 2017). Applying such an approach is especially exciting for linking community theory to global change predictions, depending on the underlying environmental gradient.

In addition to expanding theory, we see exciting potential for sparse modelling to address applied questions and inform management strategies. This includes quantifying how environmental manipulations could, in conjunction with community manipulations, control invasive species or promote native species. For example, our results from the York gum-jam woodlands of Western



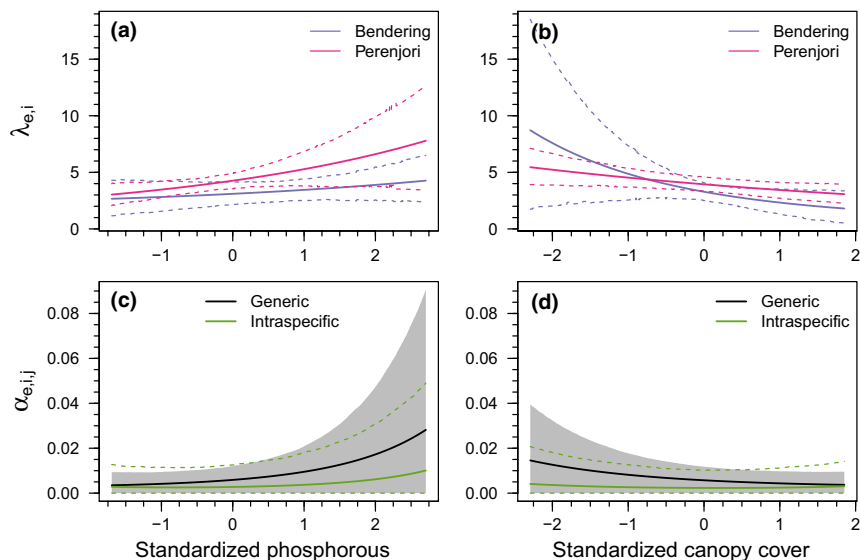
**FIGURE 3** Estimates for  $\lambda_{e,i}$  for simulated data with (a) a monotonic response to the environment and (b) an environmental optimum, with the true values as a solid red line, model means as a dashed black line and individual model posterior draws as thin grey lines. Both models were run with 200 full community plots and 200 no-competition plots. All growth rate parameters fell within the 95% credible intervals for parameters in both models. (c) Model fit of a simulation with only  $\alpha_{i,j}$  species pair intercepts (no  $\alpha_{e,i,j}$  species pair by environment slopes) with inset (d) showing the intraspecific  $\alpha_{i,i}$  intercept



**FIGURE 4** Model estimates for *W. acuminata*. Means (solid lines) and 95% CIs (dashed lines) are shown for  $\lambda_{e,i}$  across a gradient of phosphorous (a) and canopy cover (b). Colours indicate the Bendering and Perenjori reserves. The mean (black line) and 95% CI for generic interspecific competition are shown across a phosphorous (c) and canopy cover (d) gradient. In both c and d, the mean intraspecific competition coefficient is shown in green and different species identified by the model as non-generic in each reserve are shown with other colours as indicated in the legends. All non-generic results are shown with 95% CIs in Figure S3

Australia suggest the native *W. acuminata* experiences declining fitness with increasing levels of phosphorous and in response to the exotic *Hyalosperma glutinosum* in

the Perenjori reserve (Figure 4). Taken together, these results suggest that reserve managers could promote growth of *W. acuminata* populations in Perenjori by



**FIGURE 5** Model estimates for *A. calendula*. Means (solid lines) and 95% CIs (dashed lines) are shown for  $\lambda_{e,i}$  across a gradient of phosphorous (a) and canopy cover (b). Colours indicate the Bendering and Perenjori reserves. The mean (black line) and 95% CI for generic interspecific competition are shown across a phosphorous (c) and canopy cover (d) gradient. In both c and d, the mean intraspecific competition coefficient is shown in green with dashed lines indicating the CI. The model did not identify any interspecific competitors in either reserve that impacted *A. calendula* differently from a generic competition coefficient

mitigating phosphorous run-off while simultaneously removing *Hyalosperma glutinosum*. In contrast, our results for the invasive *A. calendula* suggest that neighbour species identity is unimportant (Figure 5c,d) and management strategies focusing solely on environmental factors would be most impactful.

Extending the implementation presented here, the underlying model structure can be further adjusted to align with focal management questions. For example, if a management goal only requires knowledge of species interactions within a community, the model could be simplified to remove environmental covariates (Box 1, Figure 3c,d). Alternatively, the global shrinkage parameter  $\tau$  could be fixed to induce more or less sparsity in the final model results. Such a change could allow users to manually explore the trade-off between parameter inclusion and precision, finding the balance that best suits their particular goals. For example, fixing  $\tau$  to a higher value yields more estimates of species-specific parameters, which could help to inform future research priorities, but those estimates tend to be less precise, limiting their utility in predicting community dynamics (Figure S1a–c). Small adjustments such as these empower ecologists and managers to match the tool to their questions and aims.

Currently, the sparse modelling framework is most beneficial when applied to high-diversity communities with limited data. As demonstrated in Figure S1, the number of non-generic terms is limited by  $\tau$  and a traditional, non-sparse model in which every interaction term is estimated may be preferable in situations with abundant data, lower-diversity communities, or when answering questions requiring individual estimates of

all potential species interactions. In contrast, the sparse approach is particularly helpful with limited data and in cases where traditional models struggle to converge or provide overly broad parameter estimates. In fact, the sparse approach could serve as a temporary solution in some cases, providing accurate predictions while data availability is limited, then transitioning to a non-sparse approach once sufficient data has been collected to estimate every species-specific parameter. Applying this sparse modelling approach does not require novel experimental designs nor data collection. It does, however, allow for a ‘breadth over depth’ approach, with sampling spread across an environmental gradient, and minimal replication per location (Figures 1 and 2), matching the suggestions of Cottingham et al. (2005). We advocate for including a ‘thinning’ treatment across the environmental gradient (e.g. Wainwright et al., 2019) (or, equivalently a single ‘phytometer’ treatment as is commonly done in empirical tests of modern coexistence theory Godoy & Levine, 2014; Hallett et al., 2019); otherwise the focal species’ density-independent growth rate and intraspecific competitive coefficient can strongly covary. However, this statistical identifiability issue arises from the biological models of fecundity, and is not a property of the sparse-modelling approach (Guillaume et al., 2019).

Sparse modelling approaches have proved immensely valuable in fields as diverse as genomics (Gianola & Fernando, 2020) to economics (Fan et al., 2011). By dramatically reducing the parameter load required to model diverse communities across environmental gradients, we show these sparse modelling approaches may serve as a valuable tool for linking community theory with empirical tests in species rich ecosystems. We demonstrate

the flexibility of this approach across different ecological models and underlying biological assumptions, and are excited to see it expanded and applied to a variety of ecological questions and applications. Although the implementation of the sparse method requires an initial conceptual investment, the output results are easily interpretable—a particularly important quality for linking models to practice. The sparse modelling approach eliminates the need for *a priori* assumptions regarding species' groupings or the exclusion of all but a handful of focal species, providing a critical method and step forward in expanding ecological theory and linking models to observational and experimental datasets of diverse communities.

## ACKNOWLEDGEMENTS

This paper is a joint effort of the working group sToration kindly supported by sDiv, the Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118, 02548816). CWL, CMW and LGS were supported by Modelscapes, NSF award #EPS-2019528. OG was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) and by the European Social Fund through the Ramón y Cajal Program (RYC-2017-23666); GB was supported by the Swedish Research Council (Vetenskapsrådet), grant 2017-05245. MMM was supported by the Australian Research Council (DP140100574).

## AUTHORSHIP

CWL and CMW contributed equally to this study. CWL developed and implemented the statistical approach with input from CMW, CHB, OG and LGS. CMW developed and implemented the simulations. Data were contributed by MM. CWL, CMW, LMH and LGS wrote the first manuscript draft. The manuscript was developed through conversations with the entire author list and all authors contributed to manuscript edits.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/ele.13977>.

## DATA ACCESSIBILITY STATEMENT

All data and model code necessary to recreate the analyses presented in this manuscript are archived at <https://doi.org/10.5281/zenodo.5828361>.

## ORCID

Christopher P. Weiss-Lehman  <https://orcid.org/0000-0003-1793-235X>

Catherine H. Bowler  <https://orcid.org/0000-0002-0662-1540>

Lauren M. Hallett  <https://orcid.org/0000-0002-0718-0257>

Margaret M. Mayfield  <https://orcid.org/0000-0002-5101-6542>

Lina Aoyama  <https://orcid.org/0000-0001-9677-7268>

Chengjin Chu  <https://orcid.org/0000-0002-0606-449X>

Emma Ladouceur  <https://orcid.org/0000-0002-4943-4358>

Lauren G. Shoemaker  <https://orcid.org/0000-0002-4465-8432>

## REFERENCES

- Adler, P.B., Smull, D., Beard, K.H., Choi, R.T., Furniss, T., Kulmatiski, A. et al. (2018) Competition and coexistence in plant communities: intraspecific competition is stronger than interspecific competition. *Ecology Letters*, 21, 1319–1329.
- Allesina, S. & Levine, J.M. (2011) A competitive network theory of species diversity. *Proceedings of the National Academy of Sciences*, 108, 5638–5642.
- Angert, A.L., Bontrager, M.G. & Ågren, J. (2020) What do we really know about adaptation at range edges? *Annual Review of Ecology, Evolution, and Systematics*, 51, 341–361.
- Barner, A.K., Coblenz, K.E., Hacker, S.D. & Menge, B.A. (2018) Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. *Ecology*, 99, 557–566.
- Beverton, R.J. & Holt, S.J. (1957) *On the dynamics of exploited fish populations*. 11. London: Ministry of Agriculture, Fisheries and Food.
- Bhadra, A., Datta, J., Polson, N.G., & Willard, B. (2019) Lasso meets horseshoe: a survey. *Statistical Science*, 34, 405–427.
- Bulleri, F., Bruno, J.F., Silliman, B.R. & Stachowicz, J.J. (2016) Facilitation and the niche: implications for coexistence, range shifts and ecosystem functioning. *Functional Ecology*, 30, 70–78.
- Cadotte, M.W. & Tucker, C.M. (2017) Should environmental filtering be abandoned? *Trends in Ecology & Evolution*, 32, 429–437.
- Carvalho, C.M., Polson, N.G. & Scott, J.G. (2009) Handling sparsity via the horseshoe. In: *Artificial Intelligence and Statistics*
- Chesson, P. (2000) Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, 31, 343–366.
- Clark, A.T., Ann Turnbull, L., Tredennick, A., Allan, E., Harpole, W.S., Mayfield, M.M. et al. (2020) Predicting species abundances in a grassland biodiversity experiment: Trade-offs between model complexity and generality. *Journal of Ecology*, 108, 774–787.
- Clark, J.S., Scher, C.L. & Swift, M. (2020) The emergent interactions that govern biodiversity change. *Proceedings of the National Academy of Sciences*, 117, 17074–17083.
- Cottingham, K.L., Lennon, J.T. & Brown, B.L. (2005) Knowing when to draw the line: designing more informative ecological experiments. *Frontiers in Ecology and the Environment*, 3, 145–152.
- Dwyer, J.M., Hobbs, R.J., Wainwright, C.E. & Mayfield, M.M. (2015) Climate moderates release from nutrient limitation in natural annual plant communities. *Global Ecology and Biogeography*, 24, 549–561.
- Fan, J., Lv, J. & Qi, L. (2011) Sparse high-dimensional models in economics. *Annual Review of Economics*, 3, 291–317.
- García-Callejas, D., Godoy, O. & Bartomeus, I. (2020) cxr: a toolbox for modelling species coexistence in r. *Methods in Ecology and Evolution*, 11, 1221–1226.
- Germain, R.M., Mayfield, M.M. & Gilbert, B. (2018) The ‘filtering’ metaphor revisited: competition and environment jointly structure invasibility and coexistence. *Biology Letters*, 14, 20180460.
- Gianola, D. & Fernando, R.L. (2020) A multiple-trait bayesian lasso for genome-enabled analysis and prediction of complex traits. *Genetics*, 214, 305–331.
- Godoy, O. & Levine, J.M. (2014) Phenology effects on invasion success: insights from coupling field experiments to coexistence theory. *Ecology*, 95, 726–736.
- Guillaume, J.H.A., Jakeman, J.D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B. et al. (2019) Introductory overview of



- identifiability analysis: a guide to evaluating whether you have the right type of data for your modeling purpose. *Environmental Modelling & Software*, 119, 418–432.
- Hallett, L.M., Shoemaker, L.G., White, C.T. & Suding, K.N. (2019) Rainfall variability maintains grass-forb species coexistence. *Ecology Letters*, 22, 1658–1667.
- Hastie, T., Tibshirani, R. & Wainwright, M. (2015) *Statistical learning with sparsity: the lasso and generalizations*. Boca Raton: CRC Press.
- HilleRisLambers, J., Adler, P.B., Harpole, W., Levine, J.M. & Mayfield, M.M. (2012) Rethinking community assembly through the lens of coexistence theory. *Annual Review of Ecology, Evolution, and Systematics*, 43, 227–248.
- König, C., Wüest, R.O., Graham, C.H., Karger, D.N., Sattler, T., Zimmermann, N.E. et al. (2021) Scale dependency of joint species distribution models challenges interpretation of biotic interactions. *Journal of Biogeography*, 48(7), 1541–1551.
- Kraft, N.J., Godoy, O. & Levine, J.M. (2015) Plant functional traits and the multidimensional nature of species coexistence. *Proceedings of the National Academy of Sciences*, 112, 797–802.
- Kühner, A. & Kleyer, M. (2008) A parsimonious combination of functional traits predicting plant response to disturbance and soil fertility. *Journal of Vegetation Science*, 19, 681–692.
- Lai, H.R., Mayfield, M.M., Gay-des-combes, J.M., Spiegelberger, T. & Dwyer, J.M. (2015) Distinct invasion strategies operating within a natural annual plant system. *Ecology Letters*, 18, 336–346.
- Legendre, P. & Gauthier, O. (2014) Statistical methods for temporal and space–time analysis of community composition data. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20132728.
- Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F. et al. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, 7, 601–613.
- Letten, A.D., Dhimi, M.K., Ke, P.-J. & Fukami, T. (2018) Species coexistence through simultaneous fluctuation-dependent mechanisms. *Proceedings of the National Academy of Sciences*, 115, 6745–6750.
- Levine, J.M. & HilleRisLambers, J. (2009) The importance of niches for the maintenance of species diversity. *Nature*, 461, 254.
- Levins, R. & Culver, D. (1971) Regional coexistence of species and competition between rare species. *Proceedings of the National Academy of Sciences*, 68, 1246–1248.
- Li, Y., Mayfield, M.M., Wang, B., Xiao, J., Kral, K., Janik, D. et al. (2021) Beyond direct neighbourhood effects: higher-order interactions improve modelling and predicting tree survival and growth. *National Science Review*, 8, nwa244.
- Ma, Z., Liu, H., Mi, Z., Zhang, Z., Wang, Y., Xu, W. et al. (2017) Climate warming reduces the temporal stability of plant community biomass production. *Nature Communications*, 8, 1–7.
- Martyn, T.E., Stouffer, D.B., Godoy, O., Bartomeus, I., Pastore, A. & Mayfield, M.M. (2020) Identifying ‘useful’ fitness models: balancing the benefits of added complexity with realistic data requirements in models of individual plant fitness. *The American Naturalist*, 197(4):415–433.
- May, R.M. & Leonard, W.J. (1975) Nonlinear aspects of competition between three species. *SIAM Journal on Applied Mathematics*, 29, 243–253.
- Mayfield, M.M. & Levine, J.M. (2010) Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*, 13, 1085–1093.
- Mayfield, M.M. & Stouffer, D.B. (2017) Higher-order interactions capture unexplained complexity in diverse communities. *Nature Ecology & Evolution*, 1, 1–7.
- O’Hara, R.B., Sillanpää, M.J. (2009) A review of bayesian variable selection methods: what, how and which. *Bayesian Analysis*, 4, 85–117.
- Ovaskainen, O., Rybicki, J. & Abrego, N. (2019) What can observational data reveal about metacommunity processes? *Ecography*, 42, 1877–1886.
- Ovaskainen, O., Tikhonov, G., Dunson, D., Grøtan, V., Engen, S., Sæther, B.-E. et al. (2017a) How are species interactions structured in species-rich communities? a new method for analysing time-series data. *Proceedings of the Royal Society B: Biological Sciences*, 284, 20170768.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D. et al. (2017b) How to make more out of community data? a conceptual framework and its implementation as models and software. *Ecology Letters*, 20, 561–576.
- Pérez-Ramos, I.M., Matías, L., Gómez-Aparicio, L. & Godoy, Ó. (2019) Functional traits and phenotypic plasticity modulate species coexistence across contrasting climatic conditions. *Nature Communications*, 10, 1–11.
- Picoche, C. & Barraquand, F. (2020) Strong self-regulation and widespread facilitative interactions in phytoplankton communities. *Journal of Ecology*, 108, 2232–2242.
- Piironen, J., Vehtari, A. et al. (2017) Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11, 5018–5051.
- R Core Team (2019) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. R foundation for statistical computing: Vienna, Austria. Available from: <https://www.R-project.org/>
- Ricker, W.E. (1954) Stock and recruitment. *Journal of the Fisheries Board of Canada*, 11, 559–623.
- Schamp, B.S., Arnott, S.E. & Joslin, K.L. (2015) Dispersal strength influences zooplankton co-occurrence patterns in experimental mesocosms. *Ecology*, 96, 1074–1083.
- Shoemaker, L.G. & Melbourne, B.A. (2016) Linking metacommunity paradigms to spatial coexistence mechanisms. *Ecology*, 97, 2436–2446.
- Shoemaker, L.G., Sullivan, L.L., Donohue, I., Cabral, J.S., Williams, R.J., Mayfield, M.M. et al. (2020) Integrating the underlying structure of stochasticity into community ecology. *Ecology*, 101, e02922.
- Spaak, J.W., Godoy, O. & De Laender, F. (2021) Mapping species niche and fitness differences for communities with multiple interaction types. *Oikos*, 130(12), 2065–2077.
- Stachowicz, J.J. (2001) Mutualism, facilitation, and the structure of ecological communities: positive interactions play a critical, but underappreciated, role in ecological communities by reducing physical or biotic stresses in existing habitats and by creating new habitats on which many species depend. *BioScience*, 51, 235–246.
- Stan Development Team. (2018). RStan: the R interface to Stan. R package version 2182. Available from: <http://mc-stan.org/>
- Suppiah, R., Hennessy, K., Whetton, P., McInnes, K., Macadam, I., Bathols, J. et al. (2007) Australian climate change projections derived from simulations performed for the IPCC 4th assessment report. *Australian Meteorological Magazine*, 56, 131–152.
- Thompson, P.L., Guzman, L.M., De Meester, L., Horváth, Z., Ptacnik, R., Vanschoenwinkel, B. et al. (2020) A process-based metacommunity framework linking local and regional scale community ecology. *Ecology Letters*, 23, 1314–1329.
- Tilman, D. (1982) *Resource competition and community structure*. Princeton: Princeton University Press.
- Tredennick, A.T., Hooker, G., Ellner, S.P. & Adler, P.B. (2021) A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, 102(6), e03336.
- Uriarte, M., Condit, R., Canham, C.D. & Hubbell, S.P. (2004) A spatially explicit model of sapling growth in a tropical forest: does the identity of neighbours matter? *Journal of Ecology*, 92, 348–360.
- Van Erp, S., Oberski, D.L. & Mulder, J. (2019) Shrinkage priors for bayesian penalized regression. *Journal of Mathematical Psychology*, 89, 31–50.

- Vellend, M. (2020) *The theory of ecological communities (MPB-57)*. 57, Princeton: Princeton University Press.
- Wainwright, C.E., HilleRisLambers, J., Lai, H.R., Loy, X. & Mayfield, M.M. (2019) Distinct responses of niche and fitness differences to water availability underlie variable coexistence outcomes in semi-arid annual plant communities. *Journal of Ecology*, 107, 293–306.

### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Weiss-Lehman, C.P., Werner, C.M., Bowler, C.H., Hallett, L.M., Mayfield, M.M., Godoy, O., et al. (2022) Disentangling key species interactions in diverse and heterogeneous communities: A Bayesian sparse modelling approach. *Ecology Letters*, 00, 1–14. Available from: <https://doi.org/10.1111/ele.13977>