Supplementary Material

How are species interactions structured in species rich communities? A new method for analysing time-series data

Otso Ovaskainen, Gleb Tikhonov, David Dunson, Vidar Grøtan, Steinar Engen, Bernt-Erik Sæther & Nerea Abrego

Generation of simulated data

We generated the simulated data sets with the MAR(1) model

$$y_{i,t} = c_i + \sum_{j=1}^{m} \alpha_{i,j} y_{j,t-1} + e_{i,t}.$$

For all scenarios, we sampled the intercepts c_i from the standard normal distribution $c_i \sim N(0,1)$, independently among the species i. We also sampled the noise terms $e_{i,t}$ from the standard normal distribution $e_{i,t} \sim N(0,1)$, independently among the species i and among the time steps t. The parameterization of the interaction matrix **A** depended on the model used for data generation as follows.

- In Model 1 (no inter-specific interactions), we sampled the diagonal matrix elements as $\alpha_{i,i} \sim N(0,1)$ independently among the species, and set $\alpha_{i,j} = 0$ for $i \neq j$.
- In Model 2 (full interactions), we sampled all matrix elements as $\alpha_{i,j} \sim N(0,1)$ independently among the species pairs.
- In Model 3 (sparse interactions), we proceeded first as in Model 2, and then set each off-diagonal element to zero with probability 0.9.
- In Model 4 (community-level drivers), the interaction terms are defined as $\alpha_{i,j} = \sum_{k=1}^{r} q_{i,k} w_{k,j} + \delta_{ij} \alpha_i$. We set $r = 2$, and sampled each of the $q_{i,k}$, $w_{k,j}$ and α_i from $N(0,1)$, independently among the species i and drivers k.

The MAR(1) model leads to stable dynamics if all eigenvalues of **A** have absolute value not greater than 1 (1). To ensure stable dynamics, we set the largest absolute eigenvalue of **A** to the value of 0.8 by multiplying the matrix by the appropriate constant.

We generated simulated data for 200 time steps, out of which the latter 100 time steps were stored as validation data. Out of the first 100 time steps, either all (long time-series) or the last 10 time steps (short time-series) were used as training data.

Details on model fitting

We extended the Matlab-HMSC code of Ovaskainen et al. (2) by implementing variable selection to fit Model 3, and row-column interaction models (3) to fit Model 4.

For variable selection, we modelled the interaction coefficients as $\alpha_{i,j} = \beta_{i,j} q_{i,j}$, where $q_{i,j} \in \{0,1\}$ is an indicator variable describing whether the interaction coefficient $\alpha_{i,j}$ is included in the model (i.e., non-zero) or not. We sampled the regression coefficients $\beta_{i,j}$ from their full conditional distribution separately for each focal species i using otherwise the same approach as in Ovaskainen et al. (2) but multiplying the elements of the design matrix (here, the matrix of species abundances in the previous time step) by zero if $q_{i,j} = 0$. We sampled the $q_{i,j}$ values separately for each i and j from their full conditional distribution by computing the likelihood of the data both for $q_{i,j} = 1$ and $q_{i,j} = 0$, weighting these by the prior likelihoods $p_{i,j}$ and $1 - p_{i,j}$, and sampling a new value based on these likelihoods.

For the estimation of the interaction matrices through community-level drivers (excluding the intraspecific terms, which were estimated separately), we implemented the row-column interaction model

$$\alpha_{i,j} = \sum_{k=1}^{n_d} w_{k,j} q_{i,k}$$

through a latent variable approach. Our implementation was similar to that of the co-occurrence model of Ovaskainen et al. (2), which model residual variation in occurrences and co-occurrences as

$$e_{i,t} = \sum_{k=1}^{n_f} \eta_{t,k} \lambda_{k,i}.$$

Here i is the species, t is the sampling unit (here time), n_f is the number of factors, the $\eta_{t,k}$ are the latent factors, and $\lambda_{k,i}$ are the factor loadings. We treated the parameters $w_{k,j}$ as the latent factors $\eta_{t,k}$, thus giving them a prior $w_{k,j} \sim N(0,1)$. As the prior is independent among the drivers k, they are considered a priori to be independent of each other. However, as the model does not force orthogonality among them, they may be correlated in the posterior distribution. We treated the $q_{i,k}$ as the factor loadings $\lambda_{k,i}$, thus assuming for them a multiplicative gamma prior (4) which implies increasing level of shrinkage as the function of the number of the driver k.

As prior distributions, we assumed the default priors of Ovaskainen et al. (2) for all other parameters than those extended here. For the sparse interaction model (Model 3), we assumed the priors $\beta_{i,j} \sim N(0,1)$ and $p_{i,j} \sim \text{Bernoulli}(0.1)$ for $j \neq i$. For the community-drivers model (Model 4), we assumed the parameters $\nu = 3$, $a_1 = 10$, $b_1 = 1$, $a_2 = 10$, $b_2 = 1$ for the multiplicative gamma prior of the parameters $q_{i,k}$. For the interpretation of these parameters, see Bhattacharya and Dunson (2011).

Additional measures of model fit and sensitivity of the results to prior distributions

As shown by Figure S1, the log(x+1) transformed OTU data were well in line with the assumption of residual normality.
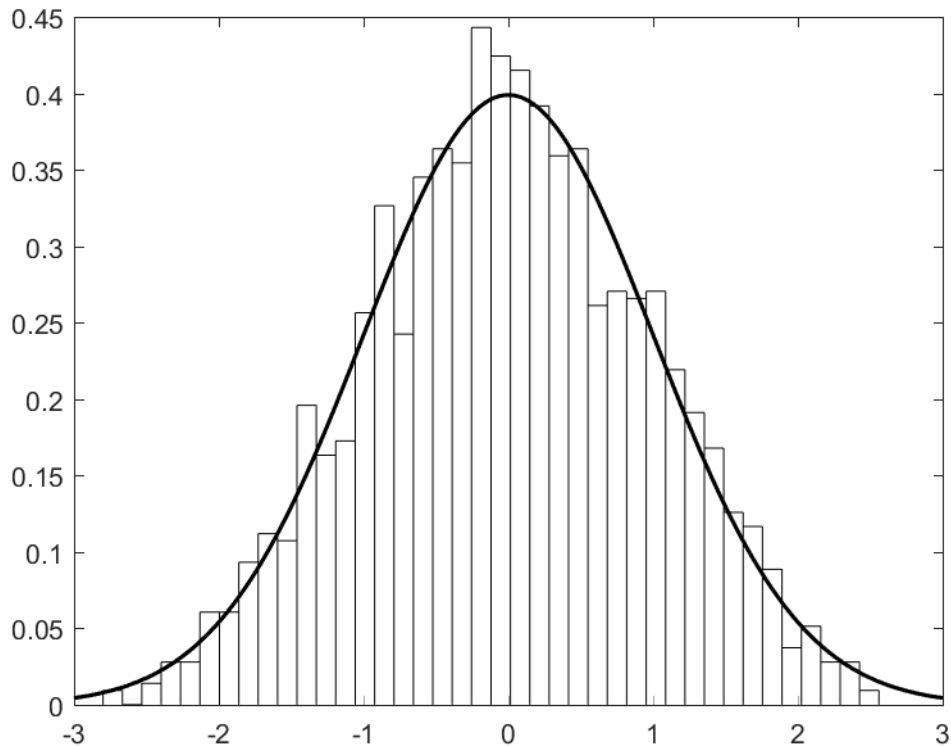

Figure S1. Comparison of model residuals (histogram) with the theoretical expectation given by the normal distribution (the line). The residuals have been computed by subtracting from the validation data the posterior mean model prediction (based on Model 4), and normalizing the residuals (separately for each species) to mean zero and unit variance.

As an alternative measure for model comparison, we used mean squared error between the validation data and the posterior mean model predictions. To enable comparison between the two measures of model fit, we repeated also here the values for correlation given in the main text. The correlation [mean squared error] between model prediction and validation data was 0.37 [4.8] (Model 1), 0.14 [23.7] (Model 2), 0.32 [12.2] (Model 3), and 0.46 [5.5] (Model 4). The correlation [mean squared error] between model prediction and training data was 0.60 [2.7] (Model 1), 0.98 [0.19] (Model 2), 0.73 [7.1] (Model 3), and 0.86 [0.98] (Model 4). Thus, these two measures ranked the models roughly similarly, high correlation corresponding to small mean squared error.

To test the sensitivity of the results on the prior distributions, we repeated the empirical case study with alternative priors for Models 3 and 4. For the sparse interaction model (Model 3), in the first alternative prior we assumed that $p_{i,j} \sim \text{Bernoulli}(0.01)$ instead $p_{i,j} \sim \text{Bernoulli}(0.1)$, and thus that the matrix is more spare than with the default prior. For the community-level drivers model (Model 4), in the first alternative prior we assumed that $a_1 = a_2 = 50$ instead of $a_1 = a_2 = 10$, and thus

assumed that the interaction matrix has a smaller effective dimensionality than with the default prior.

As described in the main text, for the default priors the correlation between model prediction and training data was 0.73 (Model 3), and 0.86 (Model 4). For the first alternative priors, the corresponding numbers were 0.63 (Model 3), and 0.76 (Model 4). Thus, as expected, imposing stronger constraints on the sparsity or dimensionality of the interaction matrix reduced the level of overfitting. As further described in the main text, for the default priors the correlation between model prediction and validation data was 0.32 (Model 3), and 0.46 (Model 4). For the first alternative priors, the corresponding numbers were 0.39 (Model 3), and 0.40 (Model 4). Thus, with the first alternative priors, the predictive performance of the sparse interaction model increased, and it slightly overperformed the predictive performance of the model without interspecific interactions (Model 1, correlation 0.37). In contrast, the predictive performance of the community-level drivers model decreased, but it still slightly overperformed the sparse interactions model.

We next shifted the prior distribution further to the direction that had thus far (default prior vs. first alternative prior) yielded best performance. As the second alternative prior, we set $p_{i,j} \sim \text{Bernoulli}(0.001)$ for Model 3 and $a_1 = a_2 = 5$ for Model 4. With these choices, the correlation between model prediction and validation data was 0.37 (Model 3), and 0.30 (Model 4). Thus, we conclude that Model 3 is not likely to overperform Model 4 with the default prior, whichever prior distribution is chosen for Model 3.
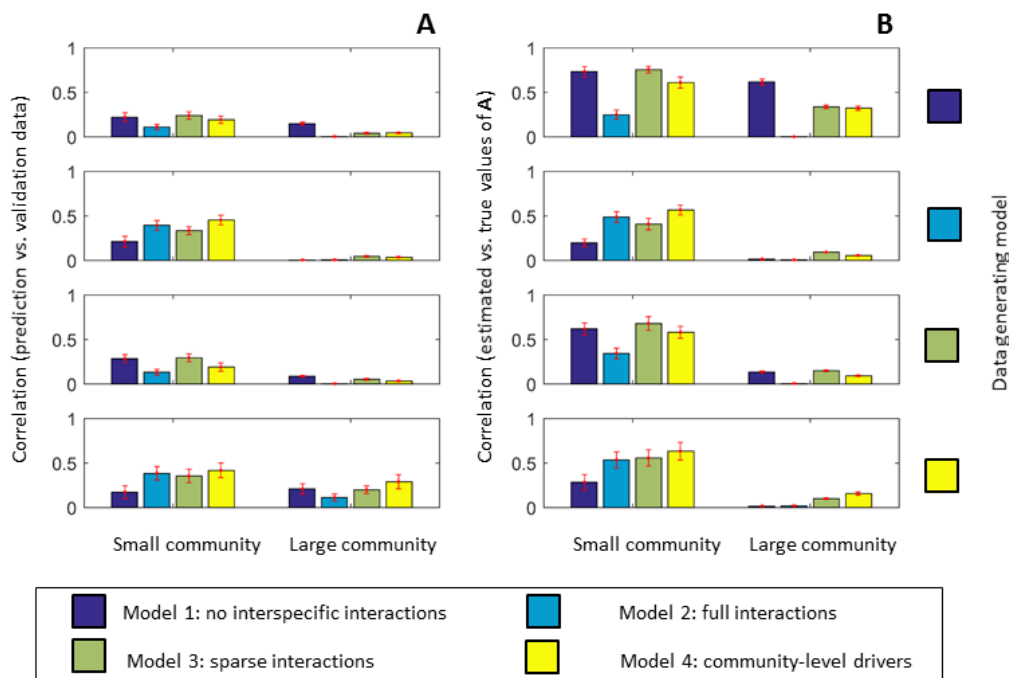
Additional results for simulated data

Figure S2. Comparison of the performance of the alternative statistical modelling frameworks based on simulated data. Panel A shows correlation between model prediction and validation data (averaged over species), and panel B the correlation between elements of true and estimated (posterior mean) interaction matrices **A**. In both panels, the rows correspond to the data generating models, the columns to small ($m = 5$) or large ($m = 100$) communities, and the colours to the models used for inference. The bars show the mean and the error bars $\pm$ two standard errors over the ten replicates. The figure shows the results for a short time series ($n = 10$), corresponding results for a long time series ($n = 100$) being shown in the main article.

References:

(1) Ives AR, Dennis B, Cottingham KL, Carpenter SR. 2003. Estimating community stability and ecological interactions from time-series data. Ecol Monogr, 73(2): 301-330. DOI: 10.1890/0012-9615(2003)073[0301:ECSAEI]2.0.CO;2.

(2) Ovaskainen O, Tikhonov G, Norberg A, Blanchet FG, Duan L, Dunson D, Roslin T, Abrego N. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecol Lett, DOI: 10.1111/ele.12757.

(3) Yee TW, Hadi AF. 2014. Row-column interaction models, with an R implementation. Comput Stat, 29(6): 1427-1445. DOI: 10.1007/s00180-014-0499-9.

(4) Bhattacharya A, Dunson DB. 2011. Sparse Bayesian infinite factor models. Biometrika, 98(2): 291-306. DOI: 10.1093/biomet/asr013.