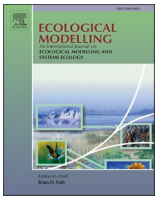
Ecological Modelling 446 (2021) 109504

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel



A method for predicting species trajectories tested with trees in barro colorado tropical forest

Hugo Fort [a,*], Tomás S. Grigera [b,c]

[a] Institute of Physics, Faculty of Science, Universidad de la República, Iguá 4225, Montevideo 11400 Uruguay
[b] Instituto de Física de Líquidos y Sistemas Biológicos, CONICET & Universidad Nacional de La Plata, La Plata, Argentina
[c] CCT CONICET La Plata, Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina, Departamento de Física, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina

ARTICLE INFO

Keywords:
Prediction of changes in species abundances
Tropical forest dynamics
Quantitative community ecology
Time series forecasting

ABSTRACT

The ability to predict changes in the abundances of the species in ecological communities is essential for sustainable management, biodiversity conservation, and community restoration.

We propose a framework to predict such changes. We test our method, which uses the linear Lotka-Volterra equations (LLVE) as well as other empirical predictors (linear least squares regression, quadratic extrapolation, simple exponential smoothing), against the measured abundances of trees from the long-term 50-ha plot on Barro Colorado Island (BCI), along eight censuses.

To obtain the parameters of the LLVE -the intrinsic growth rate r and the carrying capacity K of each species and the interspecific interaction matrix A- we first estimate A through the Maximum Entropy (MaxEnt) method. Next, using A as input, we fit r and K. Then, feeding the LLVE with these parameters, we obtain predicted species trajectories along censuses. Since for this particular community the interspecific interaction coefficients are much smaller than the intraspecific ones, keeping only intraspecific competition is enough to predict the evolution of the abundances of several tree species, i.e. the LLVE reduce to a set of uncoupled logistic equations. However, this simplification is not a requirement of the method. We define P-values to establish when the predicted trajectory for a species is statistically significant; this is crucial in determining the set of species over which a particular predictor can be meaningfully applied.

To illustrate a possible application of the method, we present our predictions for the abundances of tree species for the currently underway BCI 2020 census, which provide warnings regarding species that are likely to experience important population loss.

1. Introduction

Understanding the abundance and distribution of species at large spatial and temporal scales is the focus of macroecology (Brown and Maurer 1989; Blackburn and Gaston 1998). Indeed, abundance is one of the most important measures of species conservation status (Hui et al. 2001). Estimates of species abundances are useful to understand community structure (Gaston and Blackburn 2000), for monitoring changes in biodiversity (Wilson et al., 2004), and is the state variable of interest in most population–level ecological research and in most programs involving management and conservation of populations (Nichols and MacKenzie, 2004).

Therefore, it is hard to overstate the usefulness of any efficient technique of forecasting the variation of species abundances for biodiversity conservation, especially given global environmental change (Brook et al., 2008).

In a recent paper on the demographic trends and climate over 35 years in the long-term 50-ha plot on Barro Colorado Island (BCI), Panama, Condit et al. (2017) notice that the BCI forest has exhibited considerable dynamism in this period. This 50-ha plot was established in 1981 and fully censused in 1982 then every 5 years from 1985 through 2015. All stems $\geq$ 1 cm diameter-at breast-height (dbh) were mapped, measured, and identified to species in each census. The tree community has been in taxonomic nonequilibrium with many species' abundances

Manuscript statistics: Words in abstract: 262. Words in main text (including captions, acknowledgements and references): 7200.**Data accessibility statement:** The data used is fully available from https://doi.org/10.15146/5xcp-0d46.

* **Corresponding author:** Phone: +598-25258624.
E-mail address: hugo@fisica.edu.uy (H. Fort).

https://doi.org/10.1016/j.ecolmodel.2021.109504
Received 5 November 2020; Received in revised form 24 February 2021; Accepted 25 February 2021
Available online 3 March 2021

fluctuating over 35 years with no known cause (Condit et al., 2017).

The degree of stability in ecological communities, is related to a major debate regarding whether species coexistence emerges from equilibrium niche partitioning or from non-equilibrium stochastic dispersal-assembly (Clark and McLachlan 2003; Ishida et al. 2003). According to the first hypothesis, in a community at equilibrium each species occupies a different niche that results from and reduces direct competition (Whittaker 1975). Stabilizing mechanisms −like tradeoffs between species in terms of their capacities to disperse to sites where competition is weak, to exploit abundant resources effectively and to compete for scarce resources (Clark and McLachlan 2003)− play an important role. The size of each species' niche is determined by abiotic or biotic factors, such as resource constraints or predators. Therefore, if a species becomes too common its per-capita growth rate declines because of resource limitation or increased predation, conversely if it becomes too rare its per-capita growth rate increases. Hence niche stabilization is synonym of negative density dependence. Alternatively, the dispersal-assembly hypothesis assumes that communities are open nonequilibrium assemblages of species that coexist only transiently by chance, history, and random dispersal rather than by the stabilizing effects of niche differentiation, regarded as superfluous (Hubbell 2008, 2009). This 'neutral model' thus emphasizes 'equalizing' mechanisms (Chesson 2000), because competitive exclusion of similar species is slow. The relative importance of stabilizing and equalizing mechanisms in real communities is unknown.

Here we propose a general series forecasting method to predict future species abundances, based on the time-discrete Lotka-Volterra equations (Pastor 2008; Fort 2020). We use the BCI forest as a case study to test this method. For the tree community of the 50-ha BCI plot the interaction strengths between species were indirectly estimated by Volkov et al. (2009) using the principle of maximum entropy (MaxEnt) (Jaynes 1957, 2003) for the first census. They found that the effects of the pairwise interspecific interaction coefficients are generally *much smaller* than the intraspecific ones, by approximately an order of magnitude. Using the same MaxEnt recipe we found that the interspecific interaction coefficients are much smaller than the intraspecific ones for the eight BCI censuses (see SI). Although this is not a general result, it was empirically found through competition experiments and field studies, that pairwise interspecific interaction strengths are tend to be weak compared with the intraspecific ones for plants (Adler et al. 2018) and a quite common finding across taxonomic groups (Fort and Segura 2018). Therefore, neglecting the interspecific interaction coefficients is a sensible approximation for the BCI tree community, leading to a series of uncoupled logistic growth equations, one for each species. Firstly we show that this set of equations can explain the overall evolution of the abundances of species that make up the community of trees in the BCI plot. Secondly, we demonstrate that our method leads to statistically significant predictions for the *trajectories* of several of the species (their sequence of abundances over censuses). Thirdly, we show that using alternative purely phenomenological fits we can obtain additional statistically significant predictions for the trajectories of many species in which the logistic growth model fails.

We conclude with our predictions for the abundances of 77 tree species for the future census to be held in BCI in 2020, as well as warning signals regarding some species that are likely to experience important population loss.

2. Materials and methods

2.1. Prediction of future species trajectories from time series of empirically observed abundances

The procedure to predict future species trajectories is based on a novel general method, based on the generalized linear Lotka-Volterra equations (Pastor 2008; Fort 2020), supplemented with three different empirical fitting predictors commonly used in time series forecasting. These additional predictors serve both to assess the goodness of the method for each species and to use them for such species in which the method fails. Therefore, we start with the proposed method, predictor i) below, and then we select those species whose trajectories validate the method. For the other species me do the same with each of the empirical fittings, predictors ii) to iv). This allows to classify most of the species into different sets according which predictor best reproduces their trajectories. In fact, there are species which belong to more than one of these sets. Then we apply the prediction methods and compare predictions.

i) The linear Lotka-Volterra *finite difference* equations for describing S interacting tree species are:

$$n_i(c+1) - n_i(c) = r_i n_i(c)\left(1 + \frac{\sum_{j=1}^{S} \alpha_{ij}(c) n_j(c)}{K_i}\right), \quad (1)$$

where i denotes the tree species number; $n_i(c)$ is the density (dimensions of individuals.area^{-1}) for the census c; r_i is the intrinsic growth rate (with dimension time^{-1}); $K_i > 0$ is the carrying capacity (individuals. area^{-1}) *i.e.* the maximum density of individuals that the environment can sustain indefinitely for this species alone; and $\alpha_{ij}(c)$ is the interaction coefficient quantifying the effect of species j on species i for census c. A negative (positive) α_{ij} reflects a competitive (facilitative) interaction of species j on species i, namely, the abundance of the i^{th} species decreases (increases) with the increase of abundance of the j^{th} species. For each census, to estimate the interaction matrix $\boldsymbol{A}=[\alpha_{ij}(c)]$, we proceed as in Volkov et al. (2009) and used the MaxEnt principle (Jaynes 1957, 2003). The idea of this principle is setting up the least biased probability distribution, the one that maximizes the Shannon information entropy, on the basis of partial knowledge or constraints we have on a system. If one of such constraints is that we know the covariance matrix $\mathbf{C}$ this maximization implies that the interaction matrix between species is given by $\mathbf{A} \equiv -\boldsymbol{C}^{-1}$ [see Supporting Information (SI)]. To compute $\mathbf{C}$, we divided the 50-ha BCI plot into equally sized quadrats of side L. Thus, for each quadrat we have a vector of abundances $[n_1,n_2,\ldots,]$, and averaging over these quadrats with the probability distribution $P[n_1,n_2,\ldots,]$, we obtained $\boldsymbol{C} = [C_{ij}] \equiv \overline{n_i n_j} - \overline{n_i}\,\overline{n_j}$.

We used time series forecasting to estimate parameters r_i and K_i of equations (1) for all species as well as to validate our method. In time series forecasting, it is customary first to partition the data series of length T into an earlier training period ($t = 1, 2, \ldots t_{tr}$) and a later validation period ($t = t_{tr}+1, \ldots T$). The training period is used to estimate the parameters of the model, and then the model with estimated parameters is used to generate forecasts to be compared with data corresponding to the validation period to assess the predictive performance of a model on new data (Shmueli and Lichtendahl 2016).Varying the training period t_{tr} allows one to develop multiple model instances. The validation partition is used to assess the performance of each model so that we can compare models and pick the best one.

To estimate r_i and K_i, the general recipe is:

1 We first rewrite Eq. (1) as:

$$\frac{n_i(c+1) - n_i(c)}{n_i(c)} = \frac{r_i}{K_i}\sum_{j=1}^{S} \alpha_{ij}(c) n_j(c) + r_i. \quad (2)$$

2 For each species i we can use the empirical densities, which thereafter we will denote as n_i^e with the superscript e to

distinguish them from the theoretical ones, for the training period ($c = 1, 2,\ldots, c_{tr}$) to evaluate both sides of Eq. (2). Let us define:

$$y_i(c) \equiv \frac{n_i^e(c+1) - n_i^e(c)}{n_i^e(c)}, \tag{3-a}$$

$$x_i(c) \equiv \sum_{j=1}^{S} \alpha_{ij}(c) n_j^e(c), \tag{3-b}$$

in such a way that Eq. (2) becomes the equation of a straight line with slope r_i/K_i and intercept r_i.

1 From these c_{tr} values, since the numerator of y_i in Eq. (2) involves a difference between consecutive censuses, we get $c_{tr}-1$ points in the x_i-y_i plane (Fig. S6 in SI). To estimate r_i and r_i/K_i we apply a least squares fitting. Since several of the 328 species recorded in the BCI plot have very low abundances, their $x_i(c)$ and $y_i(c)$ exhibit drastic variations. Therefore, to make this fitting meaningful, we require a minimum density $n^e_{\min}$, *i.e.* only those S species such that $n_i^e(c) \geq n^e_{\min}$ for all c are taken into account (see Table 2).
2 For each of these species $i = 1,\ldots,S$, since the species carrying capacity, by definition, must be positive, we only take into account thus species with $K_i > 0$ (which are the great majority, as shown in Table S4).

Once we have obtained the model parameters from the training set of data, we can compare the abundances generated by the model against the empirical abundances of the validation set. Starting with $n_i(c_{tr}) = n_i^e(c_{tr})$ the subsequent theoretical abundances are generated via Eq. (1) as:

$$n_i(c+1) = n_i(c) + r_i n_i(c)\left(1 + \frac{\sum_{j=1}^{S} \alpha_{ij} n_j(c)}{K_i}\right), \; c = c_{tr}, \ldots, 8. \tag{4}$$

Eq. (4), combined with Eq. (3-b), constitutes the basis of the Lotka-Volterra predictor. In fact, for the eight BCI censuses, we found that the interspecific interaction coefficients α_{ij} ($i \neq j$) are much smaller than the intraspecific ones (see SI). We checked that including these off-diagonal elements of the interaction matrix do not change the accuracy of the predictions for BCI (see SI). When "turning off" the interspecific interactions (and taking $\alpha_{ii} = -1$), the Lotka-Volterra predictor reduces to a logistic predictor, i.e. equations (3-b) and (4) simplify, respectively, to

$$x_i(c) \equiv n_i^e(c), \tag{3′ − b}$$

$$n_i(c+1) = n_i(c) + r_i n_i(c)\left(1 - \frac{n_i(c)}{K_i}\right), \quad c = c_{tr}, \ldots, 8. \tag{4′}$$

As mentioned, we also consider three empirical extrapolations of abundance data that are commonly used in time series forecasting.

i) A linear least squares (LLS) regression.
ii) Simple exponential smoothing (SES) on the abundance increment. The abundance increment for all censuses beyond the training period is taken as constant and equal to $\Delta n_i = \alpha\, \Delta n_i^e(c_{tr}) + \alpha\,(1\text{-}\alpha)\,\Delta n_i^e\,(c_{tr}\text{-}1) + \alpha\,(1\text{-}\alpha)^2 + \ldots$ This is a sort of linear extrapolation but with weights such that the older values are gradually "forgotten" by the predictor. Rather than fitting α for each species, we found $\alpha = 0.2$ gives good results for many species.
iii) A second order polynomial fitting in the census number c, *i.e.* $n_i'(c) = \alpha_i c^2 + \beta_i c + \gamma_i$.

In the same way as we did for the logistic model, the parameters of predictors ii) and iii) are estimated using the training period.

2.2. Global accuracy metrics

A commonly used metric for characterizing a model's predictive capabilities is the coefficient of determination, R^2. To assess how well our method explains globally the abundances of *all* species of trees we considered the three more stringent error/accuracy metrics summarized in Table 1. For more details about these metrics and why we prefer them over R^2 as measures of global accuracy see SI.

2.3. Significance test for trajectories for individual species

In addition to measuring the accuracy of abundance predictions by means of the above three global metrics we performed a basic significance test for each predicted *species trajectory*. This significance test with the P-value can identify species for which the particular procedure yields better predictions than would be expected simply by chance. This test consists of reshuffling the time series data for each species, its $c_{tr}+k$ ($1 \leq k \leq 8 - c_{tr}$) measures of abundance, to obtain a statistical ensemble of artificial data including the whole $(c_{tr}+k)!$ permutations. For example, in the case of the longest trajectories, $c_{tr}+k = 8$, we have $8! = 40{,}320$ permutations. Then the relative error for the prediction on the randomized series is calculated, and the results are compared to that obtained from the original time series. The p-Value of each species trajectory, for a given validation period, is computed as the number of times the error of model predictions for the randomly shuffled series is smaller than the error for predictions corresponding to the actually observed series. The prediction is selected as significant when the P-value is low, i.e. when the model is sensitive to the temporal trends (which are destroyed by permutation).

We did this test for the logistic method as well as for the three fits we considered. Thus, for each of these four models we obtained a different set of species with statistically significant trajectories at the 95 percent confidence level. Of course, in general, there is overlap of species among these sets, i.e. species whose trajectories were significant for more than one model.

3. Results

When we use the method without restriction to the whole tree species for a given abundance threshold, $n^e_{\min}$, its accuracy is not good. That is, we use as a null model the abundances kept constant at the value measured at c_{tr}; then the *MAPE* of predictions is comparable to the *MAPE* of the null model for the immediate next censuses but then it becomes much worse. E.g. for $n^e_{\min} = 0.4$, and $c_{tr}, = 4$, the *MAPE* of logistic predictions for censuses $c = 5$–8 are 13.0, 23.4, 44.4 and 93.5, respectively. This has to be compared with the percentage mean absolute variations of abundances respect to census c_{tr}, given by 11.0, 16.8, 22.0 and 29.3 for $c = 5$–8.

Table 1
Definitions of error/accuracy measures used in this study. $\overline{n^e}$ denotes the mean of the empirical densities.

Error/accuracy measure	Definition
Mean absolute percentage error (%)	$MAPE = \frac{100}{S}\sum_{i=1}^{S}\frac{\lvert n_i^e - n_i\rvert}{n_i^e}$
Modified coefficient of efficiency	$E_1 = 1 - \frac{\sum_{i=1}^{S}\lvert n_i^e - n_i\rvert}{\sum_{j=1}^{S}\lvert n_j^e - \overline{n^e}\rvert}$
Modified index of agreement	$d_1 = 1 - \frac{\sum_{i=1}^{S}\lvert n_i^e - n_i\rvert}{\sum_{j=1}^{S}(\lvert n_j - \overline{n^e}\rvert + \lvert n_j^e - \overline{n^e}\rvert)}$

Thus we analyzed what happens when, for each series of $c_{tr}+k$ (with k between 1 and 8− c_{tr}) censuses, we use the method restricted to the subset of species such that their trajectories along these $c_{tr}+k$ censuses are significant at the 95 percent confidence level. For these subsets the method works, as expected, well. The produced *MAPE* becomes much smaller than the one yielded by the null model. Table 2 shows the *MAPE* produced by the logistic method as well as for the three fits we considered and the null model, for different training periods c_{tr} and different values of the abundance threshold $n^e_{\min}$. Notice that in most of the cases the *MAPE* of all model predictions for the next census (i.e. five years ahead) are between 1 and 2%, compared with several times this percentage for the null model. As expected, A) for a given training period, the *MAPE* almost systematically increases for predictions from one census to the next (this is particularly the case for the three fits different from the logistic model); B) increasing the training period also generally reduces the *MAPE* (this is particularly apparent by looking at the last column for c_{tr}, = 7). The pattern when decreasing the threshold $n^e_{\min}$, i.e. moving downward along columns, is less clear. It seems that the best results are obtained for $n^e_{\min}$= 1, therefore we take this as the default case that we will study in more detail.

The logistic predictor fares better that the others in the sense that, for the set of species for which it gives significant predictions, it achieves a lower MAPE. Table 2 shows that in 16 out of 42 predictions, the logistic model is able to predict with lower MAPE with respect to the other procedures. However, note that the MAPEs in Table 2 for different predictors are computed over different sets of species (see Table 4). On the other hand, when considering predictions for individual species, there are instances where one of the other empirical fits leads to a better prediction (lower relative error) than the logistic growth model (Fig. 1).

The modified coefficient of efficiency and index of agreement are also consistent with high levels of accuracy converging toward values close to 1 as shown in Table 3. [We show results only for the logistic model, but the alternative three models produce similar results.] Obviously, the prediction performance also decreases as we increase the time window we are forecasting into the future.

Finally, Table 4 shows for each cell of Tables 2 and 3 the number of species whose trajectories were significant at the 95 confidence level for the three methods. For instance, the statistical significance test for individual species trajectories for $n^e_{\min}$= 1 ha^{-1} yielded for 50 out of the 170 tree species p-Value < 0.05 (using c_{tr}=7). Indeed, for some species more than one of the predictors −the logistic supplemented with the linear and quadratic fits− yielded a p-Value< 0.05. In these cases, to untie, we took the prediction of the predictor with smaller relative error. Therefore, significant predictions of trajectories for half of these species (23) are obtained through the logistic growth model (see Table 5 in next section).

4. Discussion

4.1. Method validation

To put our method in perspective the calculation by Volkov et al. (2009) can be helpful. The authors, assuming that the tree community of BCI plot is in steady state, obtained estimates for the abundances of the 20 most abundant tree species in 1982 at BCI, i.e. just a single time point. They were able to predict the abundances of some species with accuracy but failed badly for other species, on average the *MAPE* was 86%. The method we are proposing, in the case of BCI, also neglects the interspecific interactions. However, it has remarkable differences. Firstly, our approach doesn't require the assumption that the community is in a steady state (by default it is not). Rather populations seem to be in constant turmoil. Actually, there are several species whose abundances are clearly not in a steady state (Fig. 1, Fig. S1). Secondly, and most importantly, our method thus allows forecasting *future* abundances, i.e. species trajectories. And it works not just for the most abundant species but for many of them. Thirdly, we can predict the abundances of several species with small errors 15 to 20 years into the future for several

Table 2
Mean absolute percentage errors (*MAPE*) produced by logistic method and the different fits considered (see text) compared with the null method, consisting in keeping constant the abundances measured at c_{tr}, (highlighted in gray) for species (dbh ≥ 1 cm) with trajectories such that pVal < 0.05 (see Methods). Results along validation periods for different training periods c_{tr} and thresholds $n^e_{\min}$.

	Training period →	c_{tr} = 4 (1982, 1985, 1990, 1995)				c_{tr} = 5 (1982,…,2000)			c_{tr} = 6 (1982,…,2005)		c_{tr} = 7 (1982,…,2010)
$n^e_{\min}$	Predictions for census → Model ↓	2000	2005	2010	2015	2005	2010	2015	2010	2015	2015
0.4	logistic	**1.8**	**2.0**	*12.7*	*17.6*	**1.1**	1.7	*7.3*	1.9	2.1	**1.2**
	linear	2	9.5	19.2	**1.7**	1.3	**1.5**	1.6	**1.1**	**0.8**	1.5
	ewlinear	1.9	2.1	**2.2**	3.6	1.4	**1.5**	**1**	1.3	1.2	1.5
	quadratic	2.6	3.7	8.0	17.9	2.7	4.9	6.9	2.2	3.9	1.9
	null	15.6	22.3	32.5	41.7	10.0	19.5	29.8	9.2	14.9	9.1
1.0	logistic	**1.6**	**2.0**	*13.3*	*18.4*	**1.1**	*6.6*	*10.6*	2.1	2.0	1.3
	linear	1.9	9.4	18.1	26.4	8.8	**1.1**	5.5	**1.1**	**0.9**	**0.7**
	ewlinear	2.1	2.0	15.3	**2.4**	1.3	1.6	**1.0**	1.3	1.1	1.2
	quadratic	2.5	3.5	**6.9**	13.7	2.6	4.5	7.0	1.8	2.9	1.8
	null	15.3	20.5	29.7	37.3	9.4	16.4	23.5	9.0	13.6	8.1
2.0	logistic	**1.6**	**2.1**	*13.1*	*19.6*	**0.8**	*4.9*	*8.2*	2.3	*5.9*	1.2
	linear	1.9	9.6	19	27.9	9.1	18.4	27.6	1.1	5.7	**0.7**
	ewlinear	1.7	10.9	20.6	30.9	1.5	**1.6**	1.2	**1**	**1.2**	1.1
	quadratic	2.3	3.5	**6.9**	**12.3**	2.1	8.8	17.7	7.1	2.6	1.5
	null	11.3	21.1	29.1	36.9	8.2	14.6	22.3	7.4	14.1	7.7
10.0	logistic	**1.5**	4.5	*10.0*	*15.6*	**4.8**	***11.2***	***17.7***	**4**	***8.6***	1.0
	linear	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	ewlinear	NA	NA	NA	NA	NA	NA	NA	NA	NA	**0.6**
	quadratic	1.5	**2.7**	**4.4**	**9.1**	**5.6**	NA	NA	10.2	NA	9.1
	null	8.8	18.0	24.6	32.5	9.4	16.8	24.3	8.3	16.8	7.0
20.0	logistic	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.7
	linear	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	ewlinear	NA	NA	NA	NA	NA	NA	NA	NA	NA	**0.6**
	quadratic	**0.8**	**7.2**	NA	NA	NA	NA	NA	NA	NA	8.6
	null	6.2	12.5	17.7	24.1	5.9	10.7	16.9	4.7	10.9	7.1

*Bold face indicates that the MAPE of logistic method predictions are smaller than those for the null model.
**Numbers in Italic for the logistic model denote that there are less than five species with trajectories significant to 95 percent confidence level for the corresponding cell, therefore predictions were computed using the last set of species in this row fulfilling the condition that the set includes at least five species (with the same training period c_{tr}).

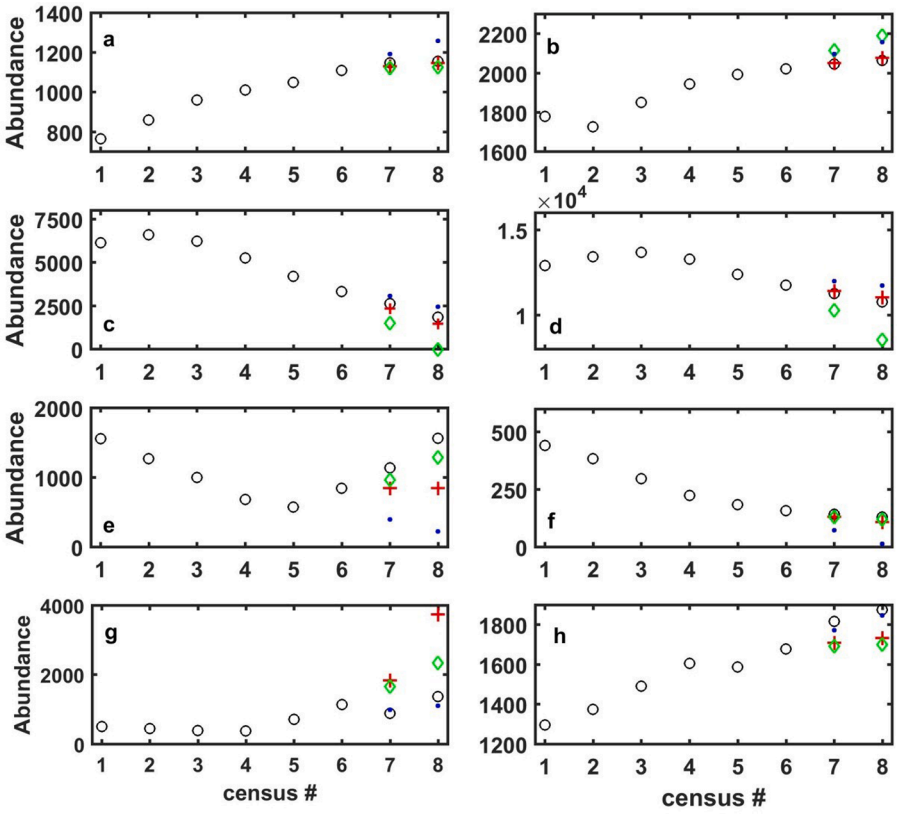


Fig. 1. Empirical abundances of eight selected species (dbh>1 cm) along the eight censuses (black circles) and predictions for census #7 and #8 (years 2010 and 2015) using six censuses for training, i.e. c_{tr}= 6.Logistic growth model predictions (red +), quadratic polynomial fit $a_2c^2+a_1c+a_0$ (green diamonds), linear least squares regression (blue dots).a) *Cassipourea elliptica*, b) *Oenocarpus mapora*, c) *Psychotria horizontalis*, d) *Trichilia tuberculata*, e) *Acalypha diversifolia*, f) *Olmedia aspera*, g) *Cecropia insignis*, h) *Tabernaemontana arborea.*

Table 3
Modified coefficient of efficiency E_1 and index of agreement d_1 for predictions from logistic method for species (dbh $\geq$ 1 cm) with trajectories such that pVal < 0.05 (see Methods). Results along validation periods for different training periods c_{tr}. and thresholds $n^e_{\min}$.

	Training period →	c_{tr} = 4 1982, 1985, 1990, 1995				c_{tr} = 5 1982,…,2000			c_{tr} = 6 1982,…,2005		c_{tr} = 7 1982/2010
$n^e_{\min}$	census → index ↓	2000	2005	2010	2015	2005	2010	2015	2010	2015	2015
0.4	E_1	0.987	0.940	*0.865*	*0.797*	0.989	0.997	*0.973*	0.988	0.973	0.993
	d_1	0.993	0.970	*0.933*	*0.898*	0.994	0.999	*0.986*	0.994	0.987	0.997
1.0	E_1	0.985	0.985	*0.932*	*0.881*	0.988	*0.955*	*0.911*	0.987	0.971	0.993
	d_1	0.993	0.992	*0.966*	*0.941*	0.994	*0.978*	*0.956*	0.994	0.986	0.996
2.0	E_1	0.980	0.967	*0.910*	*0.838*	0.980	*0.929*	*0.855*	0.987	0.954	0.992
	d_1	0.990	0.984	*0.955*	*0.940*	0.990	*0.965*	*0.928*	0.993	0.977	0.996
10	E_1	0.960	*0.896*	*0.758*	*0.619*	NA	NA	NA	NA	NA	0.991
	d_1	0.982	*0.948*	*0.885*	*0.813*	NA	NA	NA	NA	NA	0.995
20	E_1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.993
	d_1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.996

species. In particular, the MAPE for predictions five and ten years ahead is generally between 1 and 2.3%, (see Table 2 numbers in black bold face). These results are consistent with values of the modified coefficient of efficiency and index of agreement very close to 1 ($E_1 \geq 0.98$ and $d_1 \geq 0.99$). Tables 2 and 3 show that, as expected, our method yields the most accurate results for the longest possible training period of c_{tr}=7, namely a *MAPE* of at most 1.2% and E_1 and d_1 always above 0.99.

4.2. *Species consistent with a steady state are more the exception than the rule*

Only a small fraction of the 170 species with a minimum density of $n^e_{\min}$= 1 ha^{-1} might be considered in a steady state with variations not significantly different from 0 (see Table S4). This is in consonance with the observed compositional change of the BCI forest over the past 35 years (Condit et al., 2017).The estimated carrying capacities, for 155 out of these 170 species are > 0; and roughly half of the species (74) have abundances that in the last census (2015) differ in < 10% of their K_i (Table S4).

Interestingly, half of the species exhibit oscillations around their estimated carrying capacities along the 8 censuses (Table S5). It is worth noting that the species carrying capacities K_i are parameters that are far from trivial to estimate from empirical data for a community as complex as the set of trees of a tropical forest, including hundreds of species. A straightforward procedure to estimate r_i and K_i for each species is performing the corresponding *monoculture* experiment, in which the species is completely isolated from all the other species making up the community (Vandermer 1969). This seems not feasible for obvious practical reasons. It is encouraging that the estimated carrying capacities are of the order to the ones estimated by Fung et al. (2016) for the BCI forest. Nevertheless, the values provided by our method for the logistic parameters r_i and K_i, at most, serve to provide information on ongoing general trends. They should not to be considered as reliable estimates of their 'true' values, even when the abundances are accurately predicted

Table 4
Number of species (dbh ≥ 1 cm) with trajectories such that pVal < 0.05 (see Methods) for different training periods c_{tr} and thresholds.n^e_{min}.

	Training period →	c_{tr} = 4 1982, 1985, 1990, 1995				c_{tr} = 5 1982,…,2000			c_{tr} = 6 1982,…,2005		c_{tr} = 7 1982/2010
n^e_{min}	census →model ↓	2000	2005	2010	2015	2005	2010	2015	2010	2015	2015
0.4	logistic	37	20	8	6	28	11	4	22	12	34
	linear	19	7	6	12	12	14	12	18	11	15
	ewlinear	26	15	15	13	36	24	15	29	25	40
	quadrat	61	48	34	36	34	23	18	26	25	28
1.0	logistic	30	18	5	6	24	9	3	18	11	31
	linear	14	6	3	7	9	11	7	15	10	13
	ewlinear	21	11	8	10	30	20	14	25	23	35
	quadrat	56	44	31	29	30	18	15	17	22	24
2.0	logistic	21	11	3	5	16	5	3	12	8	24
	linear	11	3	1	4	7	7	6	12	9	13
	ewlinear	16	9	6	7	24	13	10	20	19	31
	quadrat	49	35	25	21	19	9	8	9	12	16
10.0	logistic	10	5	2	3	7	2	1	6	4	14
	linear	3	1	0	2	1	4	4	5	5	7
	ewlinear	3	1	1	2	9	5	5	9	9	13
	quadrat	23	18	14	12	7	3	2	3	2	7
20.0	logistic	4	2	1	2	2	2	0	3	0	10
	linear	1	0	0	1	0	3	4	2	2	4
	ewlinear	1	0	0	0	4	2	5	3	4	9
	quad	11	8	4	3	2	1	1	1	0	2

(since both estimates seem to be strongly correlated there can be compensations between them).

What drives species out from their steady state $n_i = K_i$? Two classes of stochasticity that operate in ecological communities and may cause fluctuations in species abundances are demographic variance and temporal environmental variance (Chisholm et al. 2014; Fung et al., 2016). Demographic variance comprises demographic stochasticity, which refers to variability in population growth arising from the discreteness of individuals and the random nature of birth and death processes in populations of finite size, and demographic heterogeneity, which arises when birth and death rates of individuals within a population vary by age, size, genotype, etc. Temporal environmental variance arises from the effect of random variation in environmental conditions over time on the demographic rates of species. These environmental changes include both fluctuations in abiotic factors (such as fluctuations in rainfall, temperature, fire) and in biotic factors (e.g. abundance of pests). At BCI, a major driver of tree abundance changes over the last 30 yr has been climate (Hubbell 2011. Such random perturbations could certainly trigger oscillations of n_i around K_i. There are several candidates among multiannual known climate oscillations (Baldwin et al. 2001; Bruun et al., 2017; Newman et al. 2016). Observed non-monotonically changes of the estimated K_i for many species, when varying the training period c_{tr} (Table S5), might reflect forced oscillations modulating the carrying capacities. Indeed, using time-dependent carrying capacities has long been recognized necessary to model population dynamics in environments that undergo change (Cushing 1986; Shepherd and Stojkov 2007). Such time-dependent carrying capacities have been used to model from fish population dynamics (Ikeda and Yokoi 1980) to seasonal variations of available grass in grassland-livestock production systems (Dieguez and Fort 2017) and microbial biomass (Safuan et al. 2011). The considerable compositional change in relative species abundances has been interpreted before (Hubbel 2011) as inconsistent with stabilizing forces, which would rather imply fluctuations around fixed carrying capacities. However, this new analysis, with the advantage of the availability of three new additional censuses, suggests the possibility that the carrying capacities K_i themselves could have been fluctuating strongly, painting a picture more consistent with stabilizing forces.

4.3. Comparison of the different predictors

We found p-Value < 0.05 for the predicted trajectories, by at least by one of the methods, of 77 out of 170 species for $n^e_{min} = 1.0$ (90 out of 204 species for $n^e_{min} = 0.4$). Interestingly, they can be classified into three classes, according to the type of trajectory followed by $n_i^e(c)$, which roughly coincide with the model producing the (best) prediction:

I For species such that n_i^e(c) grows or decreases more or less monotonically with a change in convexity, the logistic growth method is generally the best predictor. E.g.: Cassipourea elliptica, Psychotria horizontalis (Figs. 1.a & 1.c).

II For species such that n_i^e(c) shows, in addition to a global minimum (maximum), another local minimum (maximum), the SES or LLS turn out to be the best predictors. E.g.: Cecropia insignis, Tabernaemontana arborea Figs. (1.g & 1.h). The ability of SES to adapt to changing slope makes it often superior to LLS, and it can sometimes be slightly better than logistic or quadratic even in species for which these have P-values below threshold.

III For species such that n_i^e(c) is parabolic, as expected, the quadratic fit is the best predictor. E.g.: Acalypha diversifolia (Fig 1.e).

Which is the "best" predictor depends on how one chooses to measure the goodness of the predictions, according to intended use. In any case, it is clear that none of the four predictors can be applied blindly. The good (below null model) MAPEs we have obtained with the different predictors apply only to a subset of species, which we again emphasize is different for each predictor. Crucial in choosing this set is the use of the P-value, which can detect when the prediction is significantly better than pure chance. If one uses just one predictor on the whole set of species, the MAPEs obtained are rather poor (i.e. not better than the null model) for all predictors. The SES is rather successful in the sense that it is the predictor which has the largest number of species at a significant level of prediction for the 8th census using the first 7 as training, and for this reason dominates table 5 below. However, it can only remain successful over longer periods for species that show a sustained linear trend, since by construction it predicts a fixed Δn. It is probably more useful when trying to predict a few years ahead using a relatively long history. In contrast, the logistic method tends to give lower MAPEs over its own set of significant predictions. At variance with the other predictors, which are of empirical nature, the logistic predictor is based on theoretical expectations, and its fit parameters have a theoretical interpretation, although it must be stressed that it is harder to obtain good estimation of the logistic parameters than it is to obtain a good prediction with this form. This is due to interaction between the logistic

Table 5
Predictions for the 77 species with p-Value < 0.05 (and positive carrying capacity in the case of the Logistic method) to be compared against measurements in 2020 using the whole eight censuses as training set.

species name	pValue	Abundance		Method	Variation as%
		Observed 2015	Prediction 2020		
Acalypha diversifolia	0.037	1562	2122	quadratic	35.8
Adelia triloba	0.029	121	93	ewlinear	−22.8
Aegiphila panamensis	0.026	26	18	quadratic	−30.8
Alibertia edulis	0.029	473	501	ewlinear	5.9
Allophylus psilospermus	0.035	101	100	quadratic	−1.3
Ardisia standleyana	0.023	140	161	logistic	15.3
Aspidosperma spruceanum	0.021	542	552	linear	1.8
Astrocaryum standleyanum	0.046	133	116	logistic	−12.7
Bactris major	0.014	102	90	logistic	−11.9
Calophyllum longifolium	0.006	2237	2686	quadratic	20.1
Casearia arborea	0	108	84	ewlinear	−22.2
Cassipourea elliptica	0.015	1158	1165	quadratic	0.6
Cecropia insignis	0.009	1381	1717	quadratic	24.3
Cecropia obtusifolia	0.017	298	327	ewlinear	9.6
Chamguava schippii	0.012	600	656	linear	9.4
Chrysophyllum argenteum	0.037	819	868	ewlinear	6
Chrysophyllum cainito	0.025	188	194	logistic	3.2
Coccoloba coronata	0.046	189	195	ewlinear	3
Conostegia cinnamomea	0.039	94	82	logistic	−12.8
Cordia alliodora	0.04	223	278	quadratic	24.9
Cordia bicolor	0.007	736	717	linear	−2.6
Cordia lasiocalyx	0.011	1118	1050	ewlinear	−6.1
Croton billbergianus	0.035	634	644	ewlinear	1.5
Cupania seemannii	0.021	1649	1690	logistic	2.5
Diospyros artanthifolia	0.046	131	142	ewlinear	8.8
Drypetes standleyi	0.011	2233	2256	ewlinear	1
Erythrina costaricensis	0.025	60	56	quadratic	−6.8
Eugenia coloradoensis	0.026	629	610	logistic	−2.9
Garcinia intermedia	0.024	5340	5615	ewlinear	5.2
Garcinia madruno	0.02	425	470	quadratic	10.7
Genipa americana	0.01	65	62	ewlinear	−4.8
Guarea bullata	0.02	673	537	linear	−20.2
Guazuma ulmifolia	0.027	80	87	logistic	8.5
Guettarda foliacea	0.013	236	219	ewlinear	−7.2
Gustavia superba	0.004	713	673	quadratic	−5.6
Hybanthus prunifolius	0.019	28,960	26,829	linear	−7.4
Inga thibaudiana	0.008	344	399	logistic	16
Inga umbellifera	0.036	821	816	ewlinear	−0.6
Jacaranda copaia	0.049	309	301	logistic	−2.6
Macrocnemum roseum	0.04	94	94	ewlinear	0.5
Mosannona garwoodii	0.001	574	617	ewlinear	7.4
Miconia hondurensis	0.019	76	81	ewlinear	6.5
Miconia nervosa	0.018	442	515	quadratic	16.5
Mouriri myrtilloides	0.019	7459	7695	ewlinear	3.2
Nectandra cissiflora	0.024	249	238	ewlinear	−4.2
Nectandra lineata	0.006	277	303	ewlinear	9.3
Ocotea puberula	0.047	203	200	logistic	−1.3
Oenocarpus mapora	0.002	2065	2083	ewlinear	0.9
Trophis caucana	0.017	131	140	quadratic	6.6
Ormosia macrocalyx	0.001	154	155	quadratic	0.3
Palicourea guianensis	0.043	1326	1485	ewlinear	12
Perebea xanthochyma	0.013	223	218	ewlinear	−2.5
Piper arboreum	0.008	25	34	quadratic	35.1
Piper reticulatum	0.015	138	137	logistic	−1.1
Platypodium elegans	0.024	101	93	logistic	−8.1
Posoqueria latifolia	0.01	76	77	ewlinear	1.4
*Poulsenia armata**	*0.024*	*859*	*729*	*logistic*	*−15.1*
Protium panamense	0.047	3219	3302	ewlinear	2.6
Protium tenuifolium	0.018	3222	3274	linear	1.6
Psychotria grandis	0.028	49	50	quadratic	2.8
Psychotria horizontalis	0.025	1856	1228	logistic	−33.8
Quassia amara	0.006	113	105	linear	−6.7
Randia armata	0.003	935	918	ewlinear	−1.8
Attalea butyracea	0.001	32	32	logistic	0
Senna dariensis	0.015	248	348	quadratic	40.3
Siparuna grandiflora?	0.017	26	25	logistic	−3.7
Sloanea terniflora	0.03	447	428	ewlinear	−4.2
Socratea exorrhiza	0.003	453	404	linear	−10.9
Sterculia apetala	0.042	52	52	logistic	0.7
Stylogyne turbacensis	0.044	863	890	ewlinear	3.2
Swartzia simplex	0.042	3267	3311	logistic	1.3
Symphonia globulifera	0.034	155	153	logistic	−1.6

(*continued on next page*)

Table 5 (*continued*)

species name	pValue	Abundance Observed 2015	Prediction 2020	Method	Variation as%
Tabernaemontana arborea	0.003	1877	1957	linear	4.3
Tachigali versicolor	0.004	1950	1762	logistic	−9.7
Trichilia tuberculata	0.004	10,781	10,221	logistic	−5.2
Virola sebifera	0.004	1116	959	logistic	−14.1
Vochysia ferruginea	0.02	38	47	quadratic	22.4
Xylosma oligandra	0.043	31	11	ewlinear	−63.3

* Warning: negative *K*.

parameters, which can lead to two logistic functions with different parameters but making very similar abundance predictions (especially in the short term).

4.4. On the generality of the method and its limitations

Regarding the generality of the method, it is worth remarking that the requirement of negligibility of interspecific interactions is not mandatory. Indeed, in the general case of communities where interspecific interactions are important, the Lotka-Volterra predictor, in terms of the full interaction matrix, has to be used. That is, the pair of general equations (3-b) and (4) −rather than the simpler (3′-b) and (4′)− predict the species trajectories in the same way as we have done for BCI. The extra work is that one has to estimate the full matrix of interspecific coefficients, either from experiments −in cases in which this is feasible (Fort 2018; Vandermeer 1969)− or by other methods, like MaxEnt as we did here (see SI). Nevertheless, a limitation of the method (at least for the BCI dataset) is that it yields accurate predictions of future abundances only for the fraction of species whose trajectories during the training period are significant at the 95 percent confidence level (e.g. for a training period c_{tr}=7 for 50 out of the 170 species we have above a density threshold of $n^e_{\min}$= 1 ha^{-1}).

Increasing the number of censuses c_{tr} employed to fit the parameters r_i and K_i consistently improves the accuracy of our forecasting method. This suggests that having additional data from future censuses will probably enhance the method accuracy.

5. Conclusion

Let us conclude with our predictions for the above set of 77 species including some warnings about possible species crashes. The final model to forecast future values requires to recombine the training and validation periods into one series and estimate the model parameters r_i and K_i for these complete dataset (*i.e.* taking $c_{tr} = 8$). Table 5 shows predicted species abundances for the future # 9 census to be held in 2020 (the whole list for 170 species with a density of at least 1 tree ha^{-1} is in Table S6). Notice some yellow/red light signals in Table 5 of species for which we predict important losses in population by 2020. In particular for *Psychotria horizontalis* this drop is of one third (in consonance with a large negative growth coefficient $r = -0.45$ compared with the mean over species $|\overline{r}| = 0.34$ (Table S4). For one species in Table 5, *Poulsenia armata*, the fitted *K* parameter we obtained through the Logistic method is negative (SI). However, we included this species because it has exhibited a drastic declining in its abundance along the eight censuses and a negative *K* might be interpreted as a serious warning of this observed population decreasing trend.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank support from ANII through project ERANET-LAC R&I2016–1005422. We are also grateful to Andrea Cavagna, Massimo Cencini, Felicien Meunier, Francis Mumbanza and Hans Verbeck as well as to three anonymous referees for their criticism on previous versions of the manuscript.

We acknowledge the support of the Center for Tropical Forest Science for providing data for the BCI plot (Condit 2019).

Both authors contributed equally to this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2021.109504.

References

Adler, P.B., et al., 2018. Competition and coexistence in plant communities: intraspecific competition is stronger than interspecific competition. Ecol. Lett. 21, 1319–1329.
Baldwin, M.P., et al., 2001. The quasi-biennial oscillation. Reviews of Geophysics 39, 179–229.
Blackburn, T.M., Gaston, K.J., 1998. Some methodological issues in macroecology. American Naturalist 151, 68–83.
Brook, B.W., Sodhi, N.S., Bradshaw, C.J.A., 2008. Synergies among extinction drivers under global change. Trends in Ecology and Evolution 23, 453–460.
Brown, J.H., Maurer, B.A., 1989. Macroecology: the division of food and space among species on continents. Science 243, 1145–1150.
Bruun, J.T., Allen, J.I., Smyth, T.J., 2017. Heartbeat of the Southern Oscillation explains ENSO climatic resonances. Journal of Geophysical Research: Oceans 122, 6746–6772.
Chesson, P., 2000. Mechanisms of maintenance of species diversity. Annu. Rev. Ecol. Syst. 31, 343–366.
Clark, J.S., McLachlan, J.S., 2003. Stability of forest biodiversity. Nature 423, 635–638.
Condit, R., Perez, R., Lao, S., Aguilar, S., Hubbell, S.P., 2017. Demographic trends and climate over 35 years in the Barro Colorado 50 ha plot. Forest Ecosystems 4 art numb.17.
Chisholm, et al., 2014. Temporal variability of forest communities: empirical estimates of population change in 4000 tree species. Ecol. Lett. 17, 855–865.
Cushing, J., 1986. Oscillatory population growth in periodic environments. Theor Popul Biol 30, 289–308.
Dieguez, F., Fort, H., 2017. Towards scientifically based management of extensive livestock farming in terms of ecological predator-prey modeling. Agric Syst 153, 127–137.
Fort, H., 2020. Ecological Modelling and Ecophysics: Agricultural and Environmental Applications. IOP, Bristol, UK (IOP ebooks).
Fort, H., 2018. On predicting species yields in multispecies communities: quantifying the accuracy of the linear Lotka-Volterra generalized model. Ecol Modell 387, 154–162.
Fort, H., Segura, A., 2018. Competition across diverse taxa: quantitative integration of theory and empirical research using global indices of competition. Oikos 127, 392–402.
Fung, T., O'Dwyer, J.P., Rahman, K.A., Fletcher, C.D., Chisholm, R.A., 2016. Reproducing static and dynamic biodiversity patterns in tropical forests: the critical role of environmental variance. Ecology 97, 1207–1217.
Gaston, K.J., Blackburn, T.M., 2000. Pattern and Process in Macroecology. Blackwell Science, Oxford, UK.
Hubbell, S.P., 2008. Approaching tropical forest complexity, and ecological complexity in general, from the perspective of symmetric neutral theory. In: Carson, W., Schnitzer, S. (Eds.), Tropical Forest Community Ecology. Wiley-Blackwell, New York.
Hubbell, S.P., 2009. Neutral theory and the theory of island biogeography. In: Losos, J., Ricklefs, R.E. (Eds.), The Theory of Island Biogeography Revisited. Princeton University Press, Princeton, NJ.
Hubbell, S.P., 2011. To Know a Tropical Forest. What mechanisms maintain high treediversity on Barro Colorado island, Panama? In: Billick, I., Price, M.V. (Eds.), The Ecology of Place. The University of Chicago Press, Chicago.

Ikeda, S., Yokoi, T., 1980. Fish population dynamics under nutrient enrichment—A case of the east seto inland sea. Ecol Modell 10, 141–165.
Ishida, A., et al., 2003. Leaf physiological adjustments to changing lights: partitioning the heterogeneous resources across tree species. Pasoh: Ecology of a Lowland Rain Forest in Southeast Asia. Springer.
Jaynes, E.T., 1957. Information theory and statistical mechanics I. Phys. Rev. 106, 620–630.
Jaynes, E.T., 2003. Probability theory: the Logic of Science. Cambridge Univ. Press.
Newman, M., et al., 2016. The Pacific Decadal Oscillation, Revisited. J Clim 29, 4399–4427.
Nichols, J.D., MacKenzie, D.I., 2004. Abundance estimation and Conservation Biology. Anim Biodivers Conserv 27, 437–439.
Pastor, J., 2008. Mathematical Ecology of Populations and Ecosystems. Wiley-Blackwell.
Shmueli, G., Lichtendahl, K.C., 2016. Practical Time Series Forecasting With R: A hands-On Guide, 2nd Edition. Axelrod Schnall Publishers.
Shepherd, J., Stojkov, L., 2007. The logistic population model with slowly varying carrying capacity. ANZIAM Journal 47, 492–506.
Volkov, I., Banavar, J.R., Hubbell, S.P., Maritan, A., 2009. Inferring species interactions in tropical forests. PNAS 106, 13854–13859.
Vandermeer, J.H., 1969. The competitive structure of communities: an experimental approach with protozoa. Ecology 50, 362–371.
Whittaker, R.H., 1975. Communities and Ecosystems. MacMillan, New York.
Wilson, R.J., Thomas, C.D., Fox, R., Roy, D.B., Kunin, W.E., 2004. Spatial patterns in species distributions reveal biodiversity change. Nature 432, 393–396.