On the Assessment of Neutral Theory via Applying MaxEnt to the Analysis of Barro Colorado Island Species Abundance Data

Damian Musk
Stanford University Online High School
Palo Alto, CA 94305, USA
dmusk@ohs.stanford.edu

Abstract

This paper follows upon the Maximum Entropy (MaxEnt)-inspired ecophysical formulations of S. Azaele et al. and A. R. Rominger by constructing a computational model in conjunction with ecosystem recovery assessment models to construct physical hypothesis and computational trajectories of the Barro Colorado Island (BCI) based upon tree species abundance data collected by C. Richard et al. Through the development of quantitative methods for predicting the dynamics of multispecies communities of trees in tropical forests inspired by the BCI study area, a broad array of questions of particular importance to tropical systems—such as those pertaining to mechanisms responsible for large-scale patterns of species abundance and distribution, species coexistence, and the maintenance of vast species diversity—can be addressed, with this paper placing particular emphasis on implications relevant to those surrounding the neutral theory of molecular evolution.

Keywords

Ecophysics, MaxEnt, Neutral theory of molecular evolution, Classical statistical mechanics, Barro Colorado Island

1. Introduction

The first coherent proposal of neutral theory was presented in 1968 by M. Kimura, in which Kimura calculates a mutation rate function of $u(p) = p + 2N_e sp(1-p)$ for an initial frequency of the new allele p, effective population size N_e and selective advantage of the new allele over the pre-existing alleles s (M. Kimura 1968). Under the assumption that each enzyme in an organism with enzymatic molecular weights on the scale of $10^5 Da$—referred to as "higher organisms"—the mutation rate for the corresponding genetic site of 1,000 amino acid-enzymes of this weight (about 3,000 nucleotide pairs) is then $u = 3 \cdot 10^3 \cdot 5 \cdot 10^{-10} = 1.5 \cdot 10^{-6}$ per generation, a value so high Kimura argued that many of the mutations involved must be neutral ones. More broadly, the neutral theory of molecular evolution that comes of this observation holds that most evolutionary changes occur at the molecular level, and most of the variation within and between species is therefore due to random genetic drift of mutant alleles that are selectively neutral, "neutral" in this context referring to a mutation that does not affect an organism's fitness. In other words, the different observed outcomes for the number of species in a community and their differing abundances must therefore only result from stochastic birth, death, migration, and speciation events.

Given the resultant stochastic nature of evolution that neutral theory encourages, the application of statistical mechanics is a natural one. In particular, the maximum entropy principle of statistical mechanics (MaxEnt) was formulated by E. T. Jaynes (Jaynes 1957a, Jaynes 1957b), who emphasized the natural correspondence between statistical mechanics and Shannon's information theory, the theory that studies the quantification, storage, and communication of information. Jaynes argued that the entropy of statistical mechanics and the information entropy of information theory are effectively identical. Consequently, statistical mechanics should be seen just as a particular application of a general tool of logical inference and information theory, with MaxEnt being a useful method to obtain the least biased information from empirical measurements.

In its ecological applications, it can be seen as an inference method used to evaluate the effective strength of interactions among species on either species abundance data or simply the presence or absence of the researched species. This methodology provides a way to systematically incorporate the most important species interactions into the development of a theory beyond the purely non-interacting case.

In addition, the MaxEnt principle was implemented as a method to predict biodiversity patterns across different spatial scales using only the information on local interactions. This paper will discuss the extension of MaxEnt to the study of spatial biodiversity patterns, with the several fundamental parameters of neutral theory—a speciation rate, a measure of the size of the community (a carrying capacity), and what is effectively a time constant representing an inverse birth rate—being incorporated for the study of spatial biodiversity patterns in modern ecosystems.

Through these previous methodologies, this paper will operate on data from R. Condit et al. (R. Condit et al. 2019) comprised of species abundances of trees from eight censuses carried out from 1981 to 2015 in the 50 ha plot of Barro Colorado Island (BCI) every five years. The general goal is then to develop quantitative tools for predicting the dynamics of multispecies communities of trees in tropical forests, a method helpful to address a broad array of questions that are of particular importance to tropical systems. Including the mechanisms responsible for large-scale patterns of species abundance and distribution, species coexistence and the maintenance of the vast species diversity. These issues are not only important for the advancement of tropical ecology but are crucial for our overall understanding of basic ecology in any system. Ultimately, this research—ultimately founded on developments from the neutral theory of ecology—will come to assess neutral theory as an ecophysical perspective.

1.1 Objectives

This research primarily intends to perform further inquiry into the current landscape of quantitative methods for assessing the dynamical tools, and more importantly use this performance to assess the effectiveness of its underlying biological framework, the neutral theory of evolution. Preliminary past research into such methods has been performed, including such research with Barro Colorado Island data for data-driven applications, but the results of this report illustrate a deeper and more specifically framework-criticizing statistical approach than previous literature.

An alternate version of this report with all figures and data is available here.

2. Methods

Through applications of the described ecophysical formulations, analysis will be conducted as to show that intraspecific competition is a viable explanation for the fluctuations of the abundances of the 20 most abundant species in the BCI data. The time series of the eight available censuses is partitioned into a training period, used to estimate growth rates and carrying capacities of regional species, and a later validation period in which we check how well the estimated logsitic parameters actually predict these later observed abundances. Our approach, which works independently of whether the ecosystem is in a steady state or not, is able to predict the abundances of the 20 selected species with remarkable accuracy. A MaxEnt approach is then used to obtain the 20×20 interspecific interaction matrix for each census, as to support the assumption that pairwise interspecific interaction strengths are weak compared with the intraspecific interactions. Therefore, taking into account these interspecific interactions, through Lotka-Volterra equations, does not substantially improve the accuracy of the predicted abundances.

3. Statistical Overview of Barro Colorado Island

3.1. Conserved quantities

The BCI forest has exhibited considerable dynamism since the early 1980s (E. G. Leigh et al. 2004). However, some global quantities—defined for the whole population of trees—seem to have varied little along the 35 years across the eight censuses.

A closer look, however, shows that the stability of a quantity is dependent on the threshold dbh considered. A main threshold $dbh = 10cm$, which separates trees ($\geq 10cmdbh$) from saplings ($1 - 10cmdbh$) (R. Condit et al. 2017). As shown in Figure 2, the total numbers of trees, N, exhibits a drop of almost 15% from 1990 to 2005 and then the population seems to have more or less stabilized along the following three censuses. On the other hand, for dbh $\geq$ 10 cm (i.e. only including trees) the total abundance N oscillates in a narrow band of less than 4%.

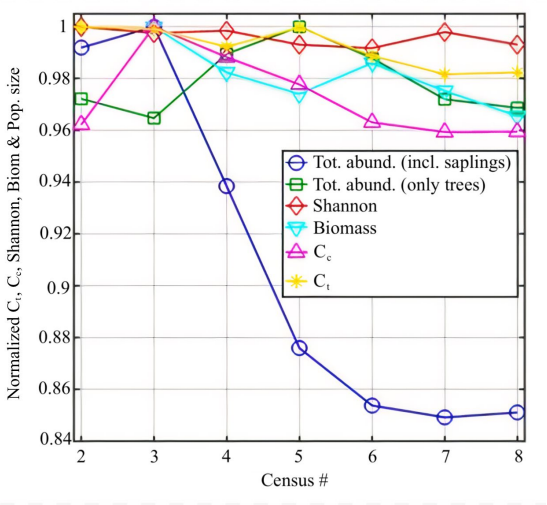

Figure 1. Empirical quantities for dbh $\geq$ 1 cm for the eight censuses, except squares which correspond to dbh $\geq$ 10 cm and includes only trees (no saplings). Census #2 corresponds to 1985 and the following censuses separated by 5 years, until census #8 in 2015.

In Figure 1, different quantities that appear to be conserved for dbh $\geq$ 1 cm are shown: $C_t = \textit{covering of trunks (discs of diameter = dbh)}$, $C_c = \textit{covering of crowns (discs of area = 0.235 dbh}^{1.43}\textit{)}$, total above-ground biomass (AGB) (proportional to the volume, which scales as $dbh^{8/3}$, for Shannon equitability of $H = -\sum_{i=1}^{S} p_i \frac{ln(p_i)}{ln(S)}$ with $p_i = N_i/N$.

Regarding species composition, a variety of population trajectories can be observed. Some species suffered steep population declines during the period 1982 - 1990 (likely as a result of severe droughts during that period), with most, but not all, recovering afterwards.

3.2. Covariance Matrices and Species Interactions

Let us consider the 20 most abundant species for trees of dbh $\geq$ 1 cm at the second census shown in Table 1. We can then divide the 50 ha BCI plot into equally sized quadrats of side L, in such a way that one can assign to each quadrat a vector of 20 abundances $\boldsymbol{n} = [n_1, \ldots, n_{20}]$, and averaging over these quadrats, determine mean species abundances and their covariances, a metric evaluating how variables change together.

From the principles of covariance, if the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values—variables tend to show similar behavior—the covariance is positive. In the opposite case, when the greater values of one variable mainly correspond to the lesser values of the other—variables tend to show opposite behavior—the covariance is negative. The diagonal elements of the covariance of abundances matrix $\boldsymbol{c}$, the intraspecific variances, are all large and positive, reflecting the clumping of single-species populations. The interspecific covariances are generally much smaller than the intraspecific variances, by approximately an order of magnitude, and can be both positive or negative.

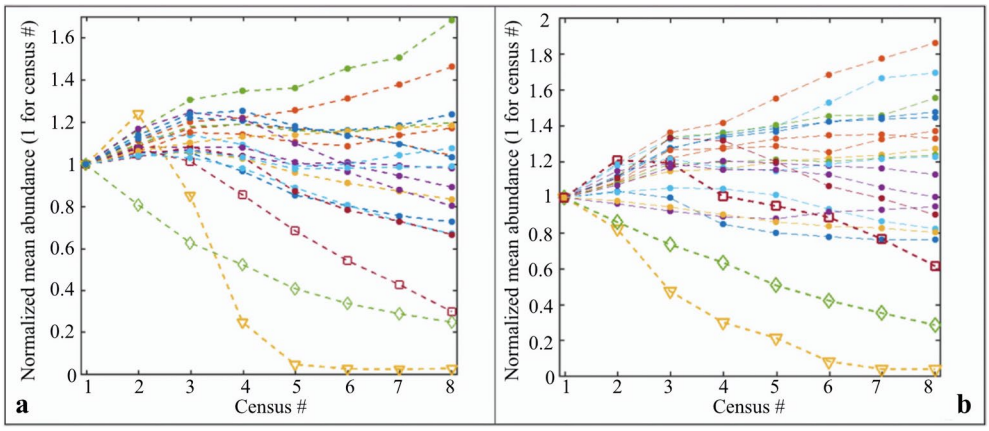

Figure 2. Normalized empirical mean abundances of the 20 most abundant species for the eight censuses. (a) dbh $\geq 1\,cm$. (b) $dbh \geq 2.5\,cm$.

Table 1. The 20 most abundant species at census #2.

Species #	Name	Abbrev.	Abundance
1	*Hybanthus prunifolius*	hybapr	42 041
2	*Faramea occidentalis*	faraoc	25 464
3	*Trichilia tuberculata*	tri2tu	13 436
4	*Desmopsis panamensis*	des2pa	12 362
5	*Alseis blackiana*	alsebl	8328
6	*Mouriri myrtilloides*	mourmy	7788
7	*Psychotria horizontalis*	psycho	6620
8	*Hirtella triandra*	hirttr	4720
9	*Garcinia intermedia*	gar2in	4064
10	*Piper cordulatum*	pipeco	3898
11	*Capparis frondosa*	cappfr	3823
12	*Tetragastris panamensis*	tet2pa	3816
13	*Sorocea affinis*	soroaf	3453
14	*Tachigali versicolor*	tachve	3028
15	*Protium tenuifolium*	protte	2917
16	*Protium panamense*	protpa	2911
17	*Swartzia simplex*	swars2	2882
18	*Beilschmiedia pendula*	belipe	2776
19	*Poulsenia armata*	poular	2766
20	*Rinorea sylvatica*	rinosy	285

4. Modeling

4.1. Inference of the effective interaction matrix from the covariance matrix via MaxEnt

We begin the modeling work by inferring the effective interaction matrix from the covariance matrix via the MaxEnt formulation. A two-dimensional square lattice Ising model relates mean abundances and the partition function through

$$\overline{x}_k = \frac{\int \prod_i dx_i x_k e^{-\sum_i h_i x_i - \frac{1}{2}\sum_{ij} J_{ij} x_i x_j}}{Z} = \frac{-\frac{\partial}{\partial h_k}\int \prod_i dx_i e^{-\sum_i h_i x_i - \frac{1}{2}\sum_{ij} J_{ij} x_i x_j}}{Z} = -\frac{\partial ln Z}{\partial h_k},$$

with the covariance matrix satisfying

$\overline{x_i x_j} - \overline{x_i x_j} = \frac{\partial^2 lnZ}{\partial h_i \partial h_j}$.

The Ising model in an electromagnetic context uses $\boldsymbol{J} = [J_{ij}]$ to describe a local magnetic field that aligns microscopic dipole moments with interaction coefficients between pairs of x_i and x_j such that

$$J_{ij} = \begin{cases} -J_{ij} & for s_i = s_j \\ +J_{ij} & for s_i = s_j \end{cases},$$

thereby making $\boldsymbol{J}$ the natural interpretation of an interaction matrix, with h_i being the magnitude of an external magnetic field such that the energy corresponding to these interactions is

$$\epsilon_i = -h_i s_i = \begin{cases} -h_i & for s_i = 1 \\ +h_i & for s_i = -1 \end{cases}.$$

For a partition function defined by the Gaussian integral

$$Z = \frac{1}{\sqrt{det\boldsymbol{J}/2\pi}} e^{-1/2 \sum_{ij} J_{ij}^{-1} h_i h_j} = \frac{1}{\sqrt{det\boldsymbol{J}/2\pi}} e^{-(1/2)\boldsymbol{h}^T \boldsymbol{J}^{-1} \boldsymbol{h}}.$$

Therefore, with $C_{ij} = \overline{(x_i - \overline{x}_i)(x_j - \overline{x}_j)} = \overline{x_i x_j} - \overline{x_i x_j}$, this partition function combined with the above covariance matrix condition gives

$$\boldsymbol{C} = -\boldsymbol{J}^{-1} \Rightarrow \boldsymbol{J} = -\boldsymbol{C}^{-1}.$$

Therefore, with interaction matrix $\boldsymbol{J}$ and covariance matrix $\boldsymbol{C}$, for each census, we must have

$$\boldsymbol{J}^{(c)} = -\boldsymbol{C}^{(c)^{-1}},$$

which has negative diagonal elements. To obtain the Lotka-Volterra interaction matrix $\mathcal{V}$ with -1 along the diagonal, we divide each row $\boldsymbol{i}$ of the $\boldsymbol{J}^{(c)}$ matrix by the absolute value of J_{ii}:

$$a_{ij}^{(c)} \equiv \frac{J_{ij}^{(c)}}{|J_{ii}^{(c)}|}.$$

Of particular importance is the effective nature of this interaction matrix, since there is no simple relationship in this many-body system between the covariance of abundances for a pair of species and their interaction.

For example, two species could have a weak interaction and yet a very strong correlation because they are both strongly interacting with a third species. Similar emergent behavior occurs when one considers more complex interactions of multiple species plus the influence of the environment they inhabit. A negative interaction with $\alpha_{ij} < 0$ is akin to "competition" between species, namely, the abundance of the ith species decreases with the increase of abundance of the jth species. Likewise, positive interactions, with $\alpha_{ij} > 0$, have coefficients that correspond to *facilitation*—the abundance of the ith species increases with the increase of the jth species' abundance.

Similarly, to what was observed for the covariance matrices, the diagonal elements of the $\boldsymbol{A}^{(c)}$ matrices, corresponding to intraspecific interactions strengths, are much greater than the off-diagonal elements, corresponding to interspecific interactions strengths.

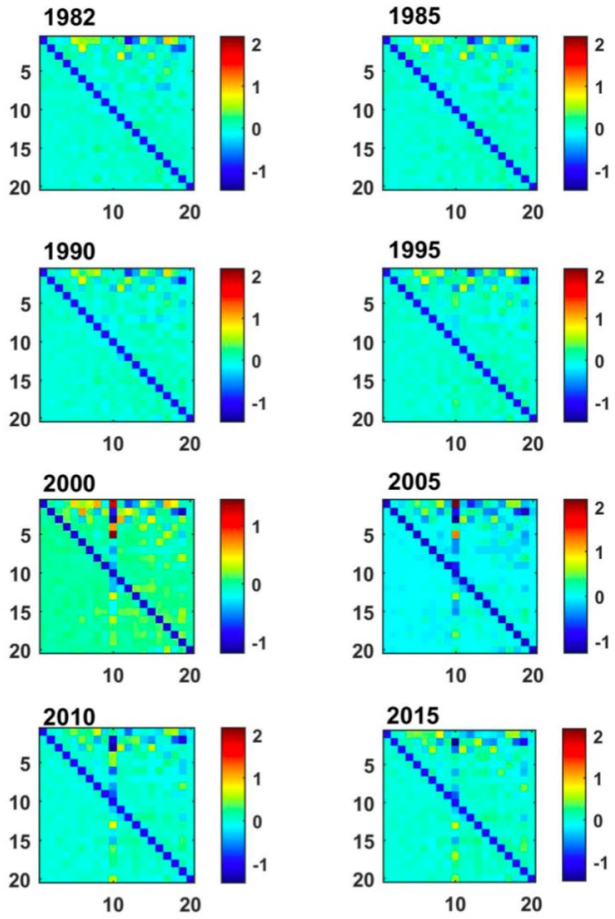

Figure 3. Interaction matrices $\boldsymbol{A}^{(c)}$ for the twenty most abundant species (dbh $\geq 1cm$) for each of the eight BCI censuses ($c = 1, \ldots, 8$). Procedure: first, compute $\boldsymbol{J}^{(c)} = -\boldsymbol{C}^{(c)^{-1}}$ and then, normalize the diagonal elements of $\boldsymbol{J}^{(c)}$ to -1 to get $\boldsymbol{A}^{(c)}$. Quadrats used to compute covariance matrices were of side $L = 20m$.

4.2. Differential and difference model equations

We are to predict the change in the abundances of the 20 most abundant tree species in the BCI plot. As previously determined, neglecting interspecific interactions implies considering a set of 20 logistic equations:

$$\frac{dn_i}{dt} = r_i n_i \left(1 - \frac{n_i}{\mathcal{K}_i}\right),$$

with tree species number i, intrinsic growth rate r_i with dimension $time^{-1}$, species i density $n_i = individuals/area$, and carrying capacity $\mathcal{K}_i > 0$ with dimensions $individuals \cdot area^{-1}$. We are not assuming equilibrium, rather we know the species composition is changing from census to census and this variation is precisely what we want to predict. Thus, besides carrying capacities $\mathcal{K}_i$ we need to estimate the intrinsic growth rates r_i, with model parameters being determined in the next subsection.

Since the time interval for our data is five years, we will use the corresponding finite difference equations $n_i(c+1) - n_i(c) = r_i n_i(c)(1 - \frac{n_i(c)}{\mathcal{K}_i})$.

5. Model Validation

5.1. Training period methodology

In time series forecasting, to assess the predictive performance of a model on new data, it is customary first to partition the data into an earlier training period ($t = 1,2, \dots, T_{tr}$) and a later validation period ($t = T_{tr} + 1, T_{tr} + 2, \dots$). 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.

Varying the training period T_{tr}, the extents of the training and validation partitions, is generally used to develop multiple models. The validation partition is used to asses the performance of each model so we can compare models and choose the optimal one.

Before attempting to forecast future values of the series, the training and validation periods must be recombined into one series, and the chosen model is rerun on the complete data. This final model is then used to forecast future values, with advantages in recombining—as suggested by Shmueli and Lichtendahl—being (G. Shmueli and K. C. Lichtendahl 2013):

1. The validation period, the most recent period, usually contains the most valuable information in terms of being the closest in time to the forecasted period;
2. With more data, some models can be estimated more accurately.

The general procedure for obtaining the intrinsic growth rates and the carrying capacities, arranged into 20×1 column vectors $\boldsymbol{r} = [\mathfrak{r}_i]$, $\boldsymbol{k} = [\mathcal{K}_i]$, is as follows:

1. Rewrite the finite difference equations as $\frac{n_i(c+1) - n_i(c)}{n_i(c)} = -\frac{\mathfrak{r}i}{\mathcal{K}_i} n_i(c) + \mathfrak{r}_i$.
2. For each species i, use empirical densities for the training period ($c = 1,2, \dots, T_{tr}$), which thereafter will be denoted as n_i^e to evaluate both sides of the difference equation. Then, define $y_i(c) \equiv \frac{n_i^e(c+1) - n_i^e(c)}{n_i^e(c)}$ and $x_i(c) \equiv n_i^e(c)$, in such a way that the previous rewritten finite difference equations become the equation of a straight line with slope $\mathfrak{r}_i/\mathcal{K}_i$ and intercept $\mathfrak{r}_i$.
3. For each species i, there are eight values of $n_i^e(c)$ with $c = 1, \dots, 8$. Use the first T_{tr} values as a training set to perform a least square fitting to obtain parameters $\mathfrak{r}_i$ and $\mathcal{K}_i$. With these four values, there are three points in the $x_i - y_i$ plane, thereby allowing for the calculation of vector parameters $\boldsymbol{r}$ and $\boldsymbol{k}$.

The results of this method are shown in Table 2, from which it can be observed that most of the estimated intrinsic growth rate coefficients r_i are positive—although this is not a necessary condition as in the case of $\mathcal{K}_i$. The carrying capacities estimated when using $T_{tr} = 4$ and 5 training censuses are very similar, with the one exception of *Poulsenia armata*. For this species, the method fails since it produces a negative carrying capacity when using $T_{tr} = 4$ training censuses. An additional concern for this species is that the estimated carrying capacity is very large and positive for $T_{tr} = 5$. Therefore, predictions of the abundance for this species are less reliable, although the relative errors are small.

6. Intrinsic Growth Rate Estimation

6.1. Generating predictions to be contrasted against a validation set of data

After obtaining the model parameters, vector parameters $\boldsymbol{r}$ and $\boldsymbol{k}$, from the training set of data we can compare the theoretical abundances generated by the model equations against the abundances of the validation set.

We begin with $n_i(T_{tr}) = n_i^e(T_{tr})$, the theoretical abundances generated through the modified finite difference equations are

$$n_i(c+1) = n_i(c) + r_i n_i(c)(1 - \frac{n_i(c)}{\mathcal{K}_i}), c = T_{tr}, \dots, 8.$$

Thus, the relative percentage errors between $n_i(T_{tr}+1), \ldots, n_i(8)$ and the empirical ones are given by $\epsilon_i(c) = 100 \times \frac{n_i(c)}{n_i^e(c)-1}, c = T_{tr+1}, \ldots, 8.$

Compiled error percentages for each species are shown in Table 3.

6.2. Model validation and predictions

As Table 3 demonstrates, increasing the training period from $T_{tr} = 4$ censuses—predictions for censuses #5 to #8—to T_{tr} for training causes the mean relative errors to decrease and converge to a value $< 5\%$ for five years ahead and $< 10\%$ for ten years ahead.

An increase of the average relative error of species abundance predictions of 1% per year is respectably accurate for a complex system like the tree community of a tropical forest, which is known to exhibit important variations in species composition.

The final model to forecast future values ultimately required recombining the training and validation periods into one series and estimating the model parameters for this complete dataset. Table 4 shows the resulting predictions for the species abundances for the future 2020 census. However, abundance data for the year of 2020 is yet to be fully published, so an assessment of the accuracies of the obtained values must wait.

7. Assessing Interspecific Interactions

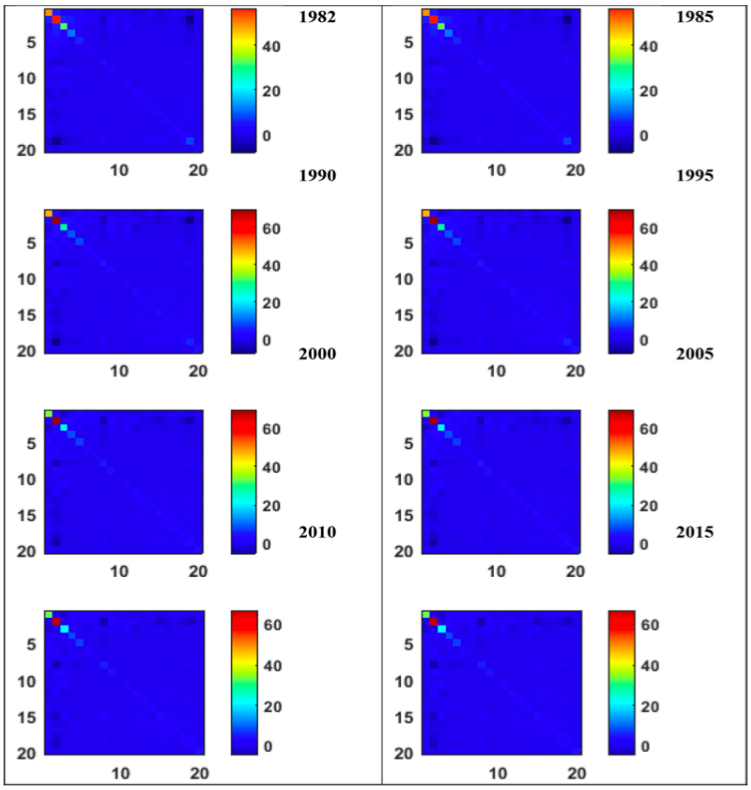

Figure 4. Covariance matrices for the eight censuses using dbh $\geq 2.5\,cm$ and $L = 10\,m$.

We recall that the diagonal elements of the Lotka-Volterra interaction matrices $\boldsymbol{A}^{(c)}$—corresponding to intraspecific interactions strengths—were much greater than the off-diagonal elements (corresponding to interspecific interactions strengths), as can be observed in Figure 3. Thus, as a first approximation, the off-diagonal terms were neglected. We

now check that considering the full matrix rather than minus the identity matrix does not produce much better results. Another issues to explore when considering the full interactions matrix is the sensitvitiy of the quality of predictions to varying the length of the quadrat side L and the dbh threshold. Thus, for efficiency, the calculations for $L = 10m$ and $dbh \geq 2.5cm$ will be repeated. The results can be seen in Figure 4 and Figure 5.

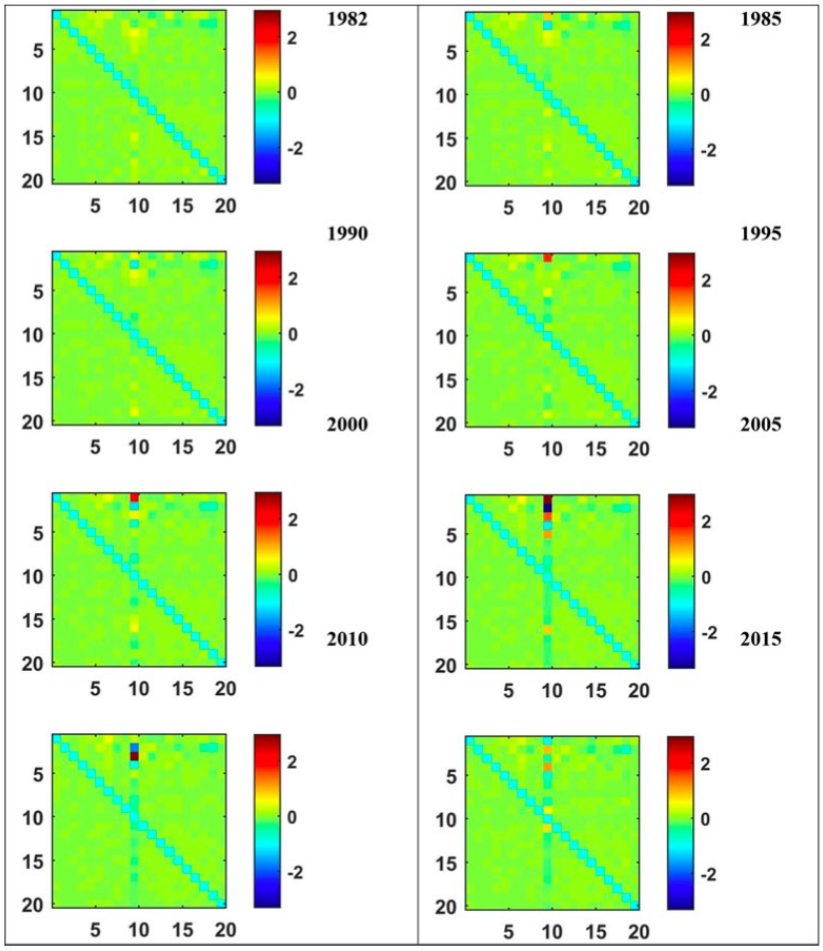

Figure 5. Interaction matrices for the eight censuses using dbh $\geq 2.5cm$ and $L = 10m$.

To make predictions that incorporate interspecific interactions, we will take the Lotka-Volterra interaction equations given by

$$n_i(c+1) - n_i(c) = r_i n_i(c)(1 + \frac{1}{\mathcal{K}_i} \sum_{j=1}^{S} \alpha_{ij}(c) N_j(c)) \Rightarrow$$
$$n_i(c+1) = n_i(c) + r_i n_i(c)(1 + \frac{1}{\mathcal{K}_i} \sum_{j=1}^{S} \alpha_{ij}(c) N_j(c)),$$

with $x_i(c) \equiv n_i^e(c)$ becoming $x_i(c) \equiv \sum_{j=1}^{S} \alpha_{ij}(c) n_j^e(c)$.

From Table 5, it becomes that there is no clear advantage regarding decreasing errors of predictions taking into account the interspecific interactions through the full Lotka-Volterra interaction matrix.

Table 2. Estimated values for intrinsic growth rates and carrying capacities when using $T_{tr} = 4$ censuses, $T_{tr} = 5$ censuses. The carrying capacity obtained for popular (*Poulsenia armata*) is clearly erroneous due being negative.

Abbrev.	$T_{tr} = 4$		$T_{tr} = 5$	
	r (y^{-1})	$\mathcal{K}$ (trees/ha)	r (y^{-1})	$\mathcal{K}$ (trees/ha)
hybapr	0.361	822	-0.078	896
faraoc	0.100	579	0.124	559
tri2tu	0.222	269	0.187	264
des2pa	0.225	247	0.221	244
alsebl	0.116	185	0.142	180
mourmy	0.238	155	0.243	151
psycho	0.147	119	-0.151	137
hirttr	0.133	106	0.168	102
gar2in	0.119	91	0.111	93
pipeco	-0.194	88	-0.194	88
cappfr	0.217	76	0.268	75
tet2pa	0.119	92	0.125	91
soroaf	0.323	68	-0.095	71
tachve	0.212	62	0.171	59
protte	0.143	64	0.164	62
protpa	0.114	64	0.146	62
swars2	0.086	63	0.094	63
belipe	0.185	59	0.220	57
poular	-0.028	-144	-0.040	7504
rinosy	0.210	54	0.206	53

Table 3. Relative errors ϵ as percentage values for the predicted abundances using $T_{tr} = 4$ censuses, $T_{tr} = 5$ censuses, $T_{tr} = 6$ censuses, and $T_{tr} = 7$ censuses for training data.

Abbrev.	$T_{tr} = 4$				$T_{tr} = 5$			$T_{tr} = 6$		$T_{tr} = 7$
	c #5	c #6	c #7	c #8	c #6	c #7	c #8	c #7	c #8	c #8
hybapr	26.0	23.4	40.3	38.5	3.6	9.9	20.6	1.5	6.3	3.5
faraoc	3.5	7.3	12.1	18.2	3.7	8.3	14.2	5.8	12.0	8.9
tri2tu	8.7	14.3	19.4	25.0	11.3	17.0	22.5	1.4	2.7	0.1
des2pa	5.6	4.3	6.1	7.3	3.3	4.8	6.1	3.8	5.4	3.4
alsebl	4.0	5.2	3.0	1.5	1.9	0.1	1.5	1.1	2.3	1.8
mourmy	13.3	11.6	7.1	3.9	9.9	4.2	1.3	0.8	0.7	1.1
psycho	35.6	76.0	125.3	221.3	10.3	34.4	59.2	10.1	19.1	2.0
hirttr	7.4	12.4	16.5	23.6	7.1	11.5	18.3	8.5	16.1	12.3
gar2in	1.3	5.0	9.3	14.4	3.7	8.1	13.2	4.5	9.9	5.1
pipeco	0.0	8.9	10.0	10.0	8.9	10.0	10.0	6.1	8.8	4.9
cappfr	6.7	12.1	23.0	34.6	11.8	20.6	32.7	15.0	28.2	4.9
tet2pa	1.4	4.1	7.0	16.5	5.3	8.3	17.7	3.8	13.9	10.7
soroaf	17.5	24.9	41.6	51.6	1.2	0.8	9.0	4.0	15.4	6.3
tachve	20.9	34.3	44.2	57.7	23.7	36.7	50.3	3.2	6.5	0.2
protte	4.7	5.6	3.6	0.8	2.6	1.0	3.3	0.1	4.0	4.0
protpa	4.6	6.8	2.3	0.7	3.3	0.8	3.7	2.4	4.8	3.9
swars2	0.6	0.2	1.0	1.6	1.0	1.9	2.6	0.8	1.5	0.7
belipe	11.8	23.5	30.3	38.2	19.3	25.7	33.4	17.3	28.3	12.4
poular	5.3	5.6	3.3	0.6	3.5	9.7	16.2	4.4	9.1	3.0
rinosy	7.0	6.5	6.1	5.8	4.7	4.4	4.1	1.6	2.2	0.9
Mean of $\lvert\epsilon\rvert$x	**9.3**	**14.6**	**20.6**	**28.6**	**7.0**	**10.9**	**17.0**	**4.8**	**9.9**	**4.5**

Table 4. Predicted species abundance values for 2020.

Abbrev.	Predicted abund. 2020
hybapr	27 262
faraoc	25 683
tri2tu	10 221
des2pa	11 697
alsebl	9028
mourmy	7448
psycho	1228
hirttr	4587
gar2in	5434
pipeco	71
cappfr	2570
tet2pa	5564
soroaf	1952
tachve	1762
protte	3142
protpa	3164
swars2	3216
belipe	2162
poular	729
rinosy	2573

Table 5. Mean relative errors as percentages for the predicted abundances using $T_{tr} = 4$ censuses for training for $L = 20m$ versus $L = 10m$ and for full Lotka-Volterra interaction matrix versus only intraspecific interactions.

	Full matrix				Only intraspecific			
	c #5	c #6	c #7	c #8	c #5	c #6	c #7	c #8
$L = 20$ m	9.4	23.4	24.3	31.4	8.0	16.7	26.7	98.1
$L = 10$ m	7.0	12.0	15.3	20.4	5.9	9.4	15.2	24.1

8. Conclusions

Thus, the tree species composition of the BCI forest is not stable and changing for reasons not yet entirely understood (R. Condit et al. 2019). However, we can predict these changes as a result of self-regulation. We have shown that most of the trajectories of the 20 most abundant species are well described by logistic difference equations with parameters estimated by species forecasting techniques. This makes us confident in the proposed method to predict the future abundances of these species. Through the Lotka-Volterra equations, the significance of interspecific competition was assessed, ultimately determined to be mostly negligible relative to intraspecific competition, which validates that initial assumption.

Moreover, we found the estimated species parameters, r_i and $\mathcal{K}_i$, vary strongly from one species to another. According to the dispersal-assembly hypothesis, which assumes that communities are open non-equilibrium assemblages of species that coexist only transiently through by chance, history, and random dispersal rather than by the stabilizing effects of niche differentiation, regarded as negligible (S. C. Thomas 2003), there are no important differences between species. Thus, the findings of this research contradict this neutrality, in agreement with large ecological differences found among tree species in tropical forests indicating that the central assumption of neutral theory that species are equivalent is simply incorrect (S. Hubbell 2008). Exact numerical predictions for the year of 2020 were determined, with overall mean relative errors of 22.1 and 13.7 for the quadrats of side length $L = 20m$ and $L = 10m$, respectively, indicating considerable accuracy for the $T_{tr} = 4$ analysis. C. Richard et al. is yet to such data, but it is likely conclusions pertaining to the neutral theory of evolution for older data will remain valid.

9. Acknowledgements
The author would like to acknowledge Dr. Lucia Foglia of Point Loma Nazarene University for her support of this paper's research proposal.

References
Kimura, Motoo. "Evolutionary rate at the molecular level". *Nature*. vol. 217, no. 5129, pp. 624–626, 1968.
Jaynes, E. T., Information Theory and Statistical Mechanics*, Physical Review*, vol. 106, no. 4, pp. 620–630, 1957. American Physical Society (APS).
Jaynes, E. T, Information Theory and Statistical Mechanics. II*, Physical* Review, vol. 108, no. 2, pp. 171–190, 1957. American Physical Society (APS).
Condit, R., Pérez, R., Aguilar, S., Lao, S., Foster, R., & Hubbell, S., Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years (Version 3) [Data set], 2019. Dryad.
Leigh, E. G., et al., in *Tropical Forest Diversity and Dynamism: Findings From a Large-Scale Plot Network*, edited by E. Losos, pp. 451-463. University of Chicago Press.
Condit, R., Pérez, R., Lao, S., Aguilar, S., & Hubbell, S. P, Demographic trends and climate over 35 years in the Barro Colorado 50 ha plot, *Forest Ecosystems*, vol. 4, no. 1, 2017. Springer Science and Business Media LLC.
Shmueli, G. and Lichtendahl, K. C., *Practical Time-Series Forecasting: A Hands-On Guide*, 3rd Edition, 2013.
Thomas, S. C., *Pasoh*, pp. 171-194, 2003. Springer Japan.
Hubbell, S., Approaching tropical forest complexity, and ecological complexity in general, from the perspective of symmetric neutral theory, *Tropical Forest Community Ecology*, 2008. New York: Wiley.

Biography
Damian Musk is a high school student at Stanford University Online High School. His scientific work specializes in physics, having conducted computationally driven research in applied statistical physics, the AdS/CFT correspondence, superconductor theory, polymer physics, and information-theoretic methods and simulation in quantum gauge theory. He currently works professionally as a theorist and phenomenologist at the Fermi National Accelerator Laboratory Theory and Simulation Department. He is a member of IEOM, APS, ACS, and IEEE.