Inferring species interactions in tropical forests

Igor Volkov[a,b], Jayanth R. Banavar[a,1], Stephen P. Hubbell[c,d], and Amos Maritan[e,1]

[a]Department of Physics, 104 Davey Laboratory, Pennsylvania State University, University Park, PA 16802; [b]Department of Biology, Muller Laboratory, Pennsylvania State University, University Park, PA 16802; [c]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095; [d]The Smithsonian Tropical Research Institute, P.O. Box 0843–03092, Balboa, Ancón Panama, Republic of Panama; [e]Dipartimento di Fisica "G. Galilei," Università di Padova, Consorzio Nazionale Interuniversitario per le Scienze Fisiche della Materia, Istituto Nazionale di Fisica Nucleare, Via Marzolo 8, 35131 Padua, Italy

We present 2 distinct and independent approaches to deduce the effective interaction strengths between species and apply it to the 20 most abundant species in the long-term 50-ha plot on Barro Colorado Island, Panama. The first approach employs the principle of maximum entropy, and the second uses a stochastic birth–death model. Both approaches yield very similar answers and show that the collective effects of the pairwise interspecific interaction strengths are weak compared with the intraspecific interactions. Our approaches can be applied to other ecological communities in steady state to evaluate the extent to which interactions need to be incorporated into theoretical explanations for their structure and dynamics.

ecology | maximum entropy principle | stochastic

How strong are the interactions among species in ecological communities, and what impact do they have on community structure and composition? This is perhaps the most fundamental problem in community ecology for understanding patterns of coexistence and the distribution and abundance of species, with important implications for conserving biodiversity. A long-standing debate persists in the literature over the extent to which communities are niche-assembled (1–4) versus dispersal-assembled (5, 6). Niche theory starts from the premise that niche differences among species are essential to species coexistence (7). In contrast, neutral theory (8), based on the simplifying assumption that all species within a trophic level are demographically equivalent, along with the hypothesis of noninteracting species, yields analytic predictions (9–11) that fit field data on static and dynamic patterns of species abundance remarkably well. However, patterns consistent with neutrality do not necessarily imply underlying neutral processes (10, 12, 13).

We present 2 independent and complementary methods for estimating interaction strengths and apply it to the tropical tree community on Barro Colorado Island, Panama (BCI). The effective interactions between species are emergent quantities that arise from the multiple interactions among the species and between the species and the environment.

The principle of maximum entropy, which can be used to derive equilibrium statistical mechanics (14–19) and has proved powerful in a variety of disciplines (20–26), is a powerful inference technique that can provide a measure of the effective interactions while faithfully encoding all available information and being unbiased with respect to missing information. Note that this approach follows the classic ecological prescription of analyzing the statics of the forest composition (27) reflected in the covariances of species abundances [the maximum-entropy approach that we present here goes beyond using only the covariances of species abundances and allows one to incorporate many-body interactions—see supporting information (SI) *Text*]. In contrast, the stochastic birth–death model provides a microscopic description of the dynamical evolution of the community in the vicinity of its present state. We study the interaction strengths among the top 20 species that make up >60% of all stems in the 50-ha plot. This dual analysis demonstrates that the collective effects of interspecific interactions among the common BCI tree species are weak

compared with the intraspecific ones. It is noteworthy that the results of both approaches are nearly identical.

Consider a steady-state community (8) composed of S coexisting species for each of which we have precise spatial information on the location of every individual. Our goal is to use these data to infer the effective interaction strengths between the species. In order to deduce various correlations from such data, we divide the community into equally sized quadrats large enough to contain many individuals yet small enough to have a sufficient number of quadrats to facilitate statistical averaging.

The analysis focuses on the probability distribution $P(\vec{n})$, where $\vec{n} = (n_1, n_2, \ldots, n_S)$, and n_i is the abundance of the ith species within a quadrat. Averaging over these quadrats, one can determine mean species abundances and higher-order correlations such as pair and triplet correlations of species abundances. The principle of maximum entropy provides a simple method to infer the probability distribution in an unbiased manner using these measured averages and correlation functions as an input. For a noninteracting system, this probability distribution would factorize into independent single-species probability distributions. Any deviation from a simple product of independent probability distributions is a measure of the interactions among species.

Operationally, the procedure is simple and straightforward (see *Materials and Methods*): One obtains the covariance matrix Σ from the data and inverts it to obtain the effective interaction matrix, $M \equiv -\Sigma^{-1}$, i.e., the component of this matrix M_{ij} may be associated with the interaction strength between species i and j. Indeed, the multivariate probability $P(\vec{n}) \propto \exp[1/2 \sum_{ij} M_{ij}(n_i - \langle n_i \rangle)(n_j - \langle n_j \rangle)]$, where the averages $\langle n_i \rangle$ are obtained by averaging the abundance data in the individual quadrats. M_{ij} serves as a coupling constant between species. Note that 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, 2 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 on a much more complex scale when one considers multiple species and the influence of the environment.

A negative interaction ($M_{ij} < 0$) is akin to "competition" between species, namely, the probability of the ith species having a given abundance decreases with the increase of abundance of the jth species. Likewise, positive interaction coefficients promote "mutualism" or similar patterns of environmental covariation between the species. The self-interaction components M_{ii} are necessarily negative—otherwise the normalization constraint, $\sum_{\vec{n}} P(\vec{n}) = 1$, cannot be satisfied. Note that the interactions deduced by using the maximum-entropy method provide a measure of the probability of occurrence of a given snapshot of species abundances. The maximum-entropy method, however, does not

Table 1. Intraspecific and interspecific covariances in abundance and interaction strengths

Comparison	Mean covariance	SE of mean covariance	Mean interaction	SE of mean interaction	No. of comparisons
Intraspecific	42.94	17.54	−0.062	0.014	20
Interspecific	0.63	0.32	0.0010	0.00051	190
Understory–Understory	1.97	0.93	0.00022	0.00065	45
Understory–Tree	0.11	0.41	0.0011	0.00079	100
Tree–Tree	0.42	0.29	0.0016	0.0011	45

Data were derived by using the method of stochastic dynamics for the 20 most abundant BCI species and for understory–understory, understory–tree, and tree–tree species pairs, not including intraspecific covariances and interactions. The distinction between tree and understory designations comes from the height of the adult plant being greater than or less than 10 m, respectively. The last column denotes the number of total number of species pairs in each category. SE, standard error.

provide information pertaining to the mechanism or strength of the direct interaction between individuals of a pair of species.

Many BCI tree species are rare to very rare, so we restrict the analysis to interactions among a subset of the most abundant species. The new effective interactions among the subset of species not only represent their direct interactions but also those mediated by the species not being considered. One can show (see *SI Text*) that the interaction matrix in this case is given by inverting a truncated covariance matrix consisting of just the rows and columns corresponding to the chosen subset of species.

The approach outlined above results in effective interactions characterizing the probability distribution $P(\vec{n})$ of the community in steady state. Our simplifying assumption that the ecological community is in the vicinity of its steady state has been validated in a recent analysis of the dynamics of the BCI forest (28). Analysis of systems far from equilibrium is much more challenging. Also, our analysis does not provide any information on the dynamics of birth and death processes. In particular, one might ask what the relationship is between the effective interactions deduced from the principle of maximum entropy and the dynamical interactions between species populations. In order to address this issue, in our second approach, we consider the dynamics of simple birth–death processes and infer the microscopic interactions that correspond to the observed steady state of the community.

Briefly, our approach (see *Materials and Methods*) is to write a master equation for the time-dependent probability $P(\vec{n}, t)$ governing the dynamics of the community. We postulate that the per-capita birth and death rates depend on the abundances of all species. In the simplest scenario, without loss of generality, the per-capita birth rate of the ith species is a linear function of the abundances of all the other species: $b_i + \sum_j c_{ij} n_j$; the per-capita death rate of the ith species is assumed to be a constant d_i. b_i is the per-capita birth rate of the ith species, and c_{ij} is a measure of the effective interaction between the ith and the jth species. If the community is at stationarity, then one can determine these coefficients by requiring that the appropriate moments of the probability distribution do not change with time.

We carried out our analysis on tree data from the BCI 50-ha plot divided into 1,250 quadrats each measuring 20×20 m. We considered all trees with a diameter >1 cm. We selected a quadrat size so that each quadrat had, on average, >180 trees, and yet the total number of quadrats allowed good statistical averaging. However, our results are essentially independent of quadrat size (see *SI Text*). We assigned to each quadrat a 20-component vector $\vec{x} \equiv (x_1, x_2, \ldots, x_{20})$ representing the abundances of the 20 most abundant species in the entire BCI plot in rank-descending order. The BCI data can be represented as a collection of 1,250 vectors (one for each quadrat) from which one can readily construct the covariance matrix for the maximum-entropy framework as well as the relevant averages used in the stochastic dynamics framework. We have studied how the effective interaction strengths depend on choosing fewer or more species in our analysis, and the results are robust to changes in the number of focal species (see *SI Text*). These effective interactions arise not only from the direct interactions between the focal species but also from the indirect

effects of all variables (environmental as well as other species) not considered explicitly.

The 20 most abundant BCI species include 10 main canopy and subcanopy species and 10 understory species, reflecting the mix of species growth forms in the BCI forest. The analysis incorporates interactions among these species at all stages of ontogeny from small saplings of 1-cm dbh (diameter at the breast height) to larger individuals. Note that canopy trees spend many years during their ontogeny in the understory interacting with understory species, and understory species have lifelong interactions with canopy tree species.

The covariance among these 20 species demonstrates nonrandom spatial structure in the BCI forest. Table 1 reports the mean covariances and their standard errors for various species groupings: intraspecific for within-species spatial covariances and interspecific for between-species covariances. The intraspecific covariances are all large and positive, reflecting the clumping of single-species populations due to a variety of factors such as habitat preferences and dispersal limitation. The interspecific covariances are generally much smaller, by approximately an order of magnitude, and can be both positive or negative. Table 1 also shows the covariances for understory–understory, understory–tree, and tree–tree species pairs.

Fig. 1 *Upper Left* is a graphical representation of the covariance matrix for the top 20 species in the BCI forest, with the intraspecific covariances, shown in Fig. 1 *Upper Right*, set to zero to make the smaller interspecific covariances more visually apparent. Fig. 1 *Upper Center* shows deviations of the covariance matrix from the average values of a set of 1,000 random realizations measured in units of their respective standard deviations. Note that the interactions obtained after the reshuffling are ≈2 orders of magnitude larger than the original interactions (see Fig. S1). The figure shows that, despite the generally small size of the interspecific covariances, they often deviate significantly from those of randomly shuffled configurations of the plants following a Poisson process.

To what extent do these covariances arise because of species interactions? Even if 2 species do not interact, they could have nonzero covariance, because the distributions of the 2 species are controlled independently by intraspecific processes and/or history. The fact that the intraspecific covariances are an order of magnitude larger than interspecific covariances suggests this might be the case for the main BCI species pairs. Of course, real ecological experiments are the best way to establish causal relationships between species.

Fig. 1 *Lower Left* is a graphical representation of the interaction matrix deduced by using the stochastic dynamics framework, and Fig. 1 *Lower Center* shows the deviations of these values from the randomly shuffled case measured in units of the respective standard deviation. These figures show that the deduced per-capita interactions are generally weak and especially weak among the most abundant species than among the less abundant species. Four of the 10 most abundant species and 10 of the 20 most abundant species are tree species. The tree–tree interactions are stronger than the understory–understory and the tree–understory
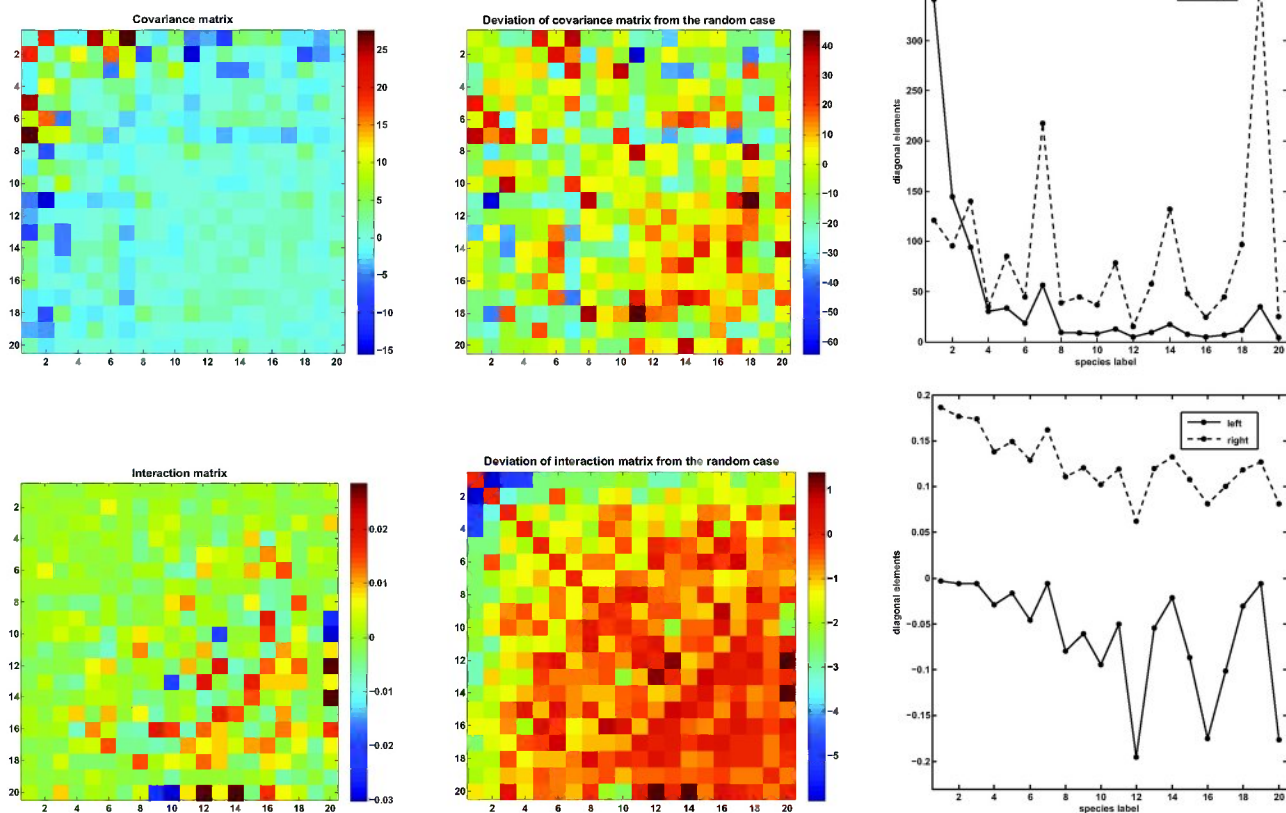
Fig. 1. Plots of covariance and interaction matrices for the top 20 species in the BCI forest. (*Left*) Plots of the covariance matrix Σ (*Upper*) and the interaction matrix M (*Lower*) deduced by using the stochastic dynamics framework. For ease of viewing, the diagonal elements have been set equal to zero because they are typically larger than the off-diagonal entries (see *SI Text*). (*Center*) Deviation of the covariance (*Upper*) and interaction (*Lower*) matrices from the randomized BCI dataset in units of standard deviations of the randomized plots, again with the diagonal entries set equal to zero. We generated 1,000 plots with randomly shuffled species labels—this was accomplished by repeatedly picking a pair of individuals randomly and interchanging their species labels. This procedure ensured that the total abundance of each of the species was held fixed. (*Right*) The diagonal values are shown with the values of the diagonal of the lower center matrix reduced by a factor of 100. There is a strong (88%) correlation between the intraspecific interactions and the sum of the absolute values of the interspecific interactions. The species with the overall strongest interactions are species 20 and 12 (*Virola sebifera* and *Sorocea affinis*, respectively). Even though the interactions are relatively weak, there are several interactions that deviate from those obtained with a random placement of trees and are thus sensitive to the spatial distribution of species in the BCI forest.

interactions (Table 1). Table 1 also shows that intraspecific interaction strengths are an order of magnitude larger than the interspecific interaction strengths. Our results for the strongest interactions estimated from the stochastic dynamics approach, along with the associated species covariances, are summarized in Tables 2 and 3. Species with large diagonal values have strong couplings with several other species, whereas species with smaller diagonal elements generally have couplings of lesser magnitude.

Plants are rooted organisms, and many plant species may not be able to interact because they do not grow in proximity to one another. The more species there are in the community, the less likely it is that 2 species are able to interact even if they are randomly intermingled, which they are not. Indeed, the mean number of the 20 most abundant species in each of the 1,250 quadrats is 15.7, and the fraction of quadrats with all 20 species is of the order of 1%. We repeated our analysis with 200 larger-sized quadrats. In

Table 2. Large positive or negative covariances and associated interaction strengths

Species A	Category	Species B	Category	Covariance	Interaction
Hybanthus prunifolius	U	*Psychotria horizontalis*	U	27.6	0.00073
Hybanthus prunifolius	U	*Alseis blackiana*	T	25.5	0.0014
Hybanthus prunifolius	U	*Faramea occidentalis*	U	18.9	0.000090
Mouriri myrtilloides	U	*Faramea occidentalis*	U	16.8	0.0057
Desmopsis panamensis	U	*Faramea occidentalis*	U	11.6	0.00089
Trichilia tuberculata	T	*Mouriri myrtilloides*	U	−5.4	−0.0026
Hirtella triandra	T	*Faramea occidentalis*	U	−6.9	−0.0038
Protium tenuifolium	T	*Faramea occidentalis*	U	−7.1	−0.0020
Hybanthus prunifolius	U	*Tetragastris panamensis*	T	−8.4	−0.0027
Poulsenia armata	T	*Faramea occidentalis*	U	−15.5	−0.0052

The 5 most positive and 5 most negative covariances are shown. T, tree; U, understory.

Table 3. Large positive or negative interaction strengths deduced by the method of stochastic dynamics and associated covariances

Species A	Category	Species B	Category	Interaction	Covariance
Virola sebifera	T	*Sorocea affinis*	U	0.0285	0.70
Virola sebifera	T	*Piper cordulatum*	U	0.0282	2.75
Tetragastris panamensis	T	*Sorocea affinis*	U	0.020	1.61
Garcinia intermedia	T	*Swartzia simplex*	U	0.019	0.86
Tetragastris panamensis	T	*Tachigali versicolor*	T	0.017	1.61
Mouriri myrtilloides	U	*Poulsenia armata*	T	−0.0075	−2.20
Piper cordulatum	U	*Sorocea affinis*	U	−0.0082	0.26
Tetragastris panamensis	T	*Capparis frondosa*	U	−0.021	−1.80
Virola sebifera	T	*Garcinia intermedia*	T	−0.025	−0.84
Virola sebifera	T	*Capparis frondosa*	U	−0.030	−0.93

T, tree; U, understory.

this case, >30% of the quadrats have all 20 species, and >85% have either 19 or 20 coexisting species. The scale of the interactions is approximately a factor of 10 smaller for the 200-quadrat case compared with 1,250-quadrat case. However, there is an excellent correlation of ≈96% between the interactions inferred in the 2 cases.

We also studied the sensitivity of the results to the range of interactions between trees. We considered the BCI data for trees with dbh >10 cm and chose the 20 most abundant species (11,766 trees in total). We performed a Voronoi tesselation (29) and identified the nearest neighbors for each tree within a cutoff distance of 7 m. One can regard each tree and its nearest neighbors as a "quadrat" in the same manner as above and construct covariance and interaction matrices. The correlation coefficients between covariance and interaction matrices for the "Voronoi partition" and for the regular partition into 1,250 quadrats are 0.97 and 0.99, respectively, which again highlight the robustness of our results.
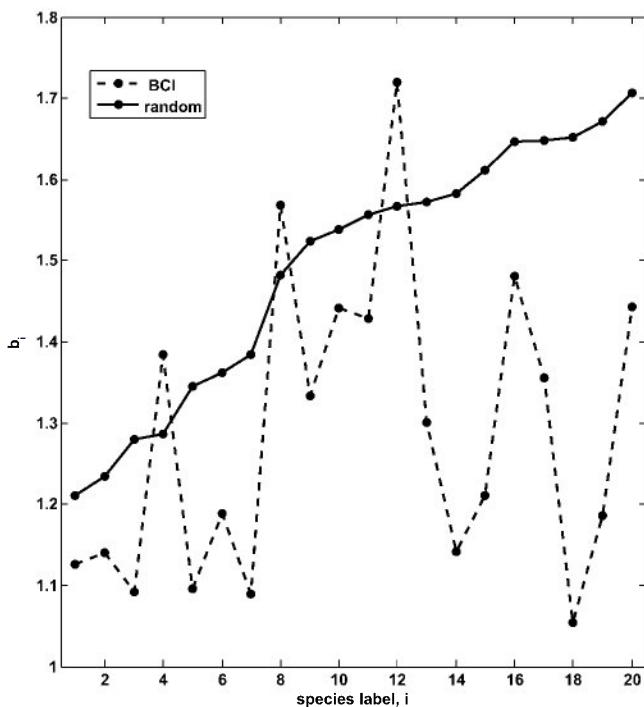
A key conclusion from our study is that there is remarkable consistency in the effective interactions deduced by using 2 quite dissimilar frameworks—one based on the principle of maximum

Fig. 2. Per-capita birth rates for the 20 most abundant species in the BCI forest and for the randomly reshuffled case using the stochastic dynamics framework. The per-capita birth rate b_i, measured in units of the death rate, is plotted versus species label. Species are labeled according to their ranking in terms of their abundance, with species 1 being the most abundant. For the random case, b_i increases monotonically as the species abundance decreases.

entropy and another based on stochastic dynamics (Fig. S1). The latter method provides key information pertaining to effective per-capita birth rates (Fig. 2). The agreement between the results of the 2 methods is not entirely surprising because the solution of the master equation in the stochastic dynamics framework can be approximated by a multivariate Gaussian distribution (30). However, both models have different underlying assumptions and complement each other, and the degree of accord is noteworthy. In general, from a quantitative perspective, both approaches lead to the conclusion that the effects of interspecific interactions are relatively weak and impose a relatively minor constraint on species dynamics in the BCI forest (Fig. 3). It is as though evolution has chosen weakly interacting species for proximal coexistence.
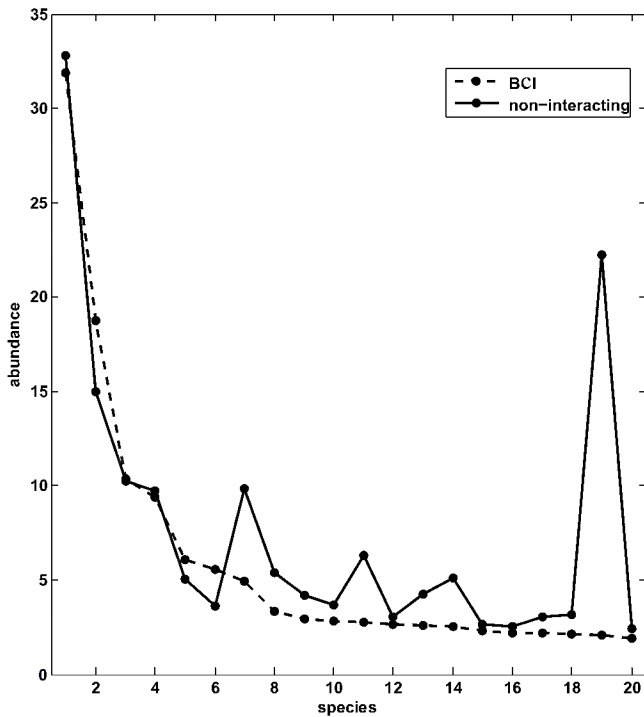
Why are these interaction strengths generally weak? It is useful to consider first the much simpler case of a standard physical system at temperature T and fixed density ρ whose N particles interact with a potential $V(\mathcal{C})$ which depends on the particle configuration $\mathcal{C}$. Quite generically, the configurational entropy, arising from this potential energy is given by $\mathcal{H}(T, N, \rho) \equiv -k_B \sum_{\mathcal{C}} P(\mathcal{C}) \ln P(\mathcal{C})$ with $P(\mathcal{C}) \propto e^{-\frac{V(\mathcal{C})}{k_B T}}$. $\mathcal{H}$ is an increasing function of T, $\exp\{\mathcal{H}\}$ is a measure of the configurational space that the system can visit for a particular choice of T, N and ρ and k_B is the Boltzmann constant. Imagine now that we allow particles to change their interactions. We can do this by constructing a reservoir of particles characterized according to the nature of their interactions and allow for particle exchanges between the system and the reservoir. The question then is what is the most probable state we observe at stationarity? The particles present at stationarity will be mostly noninteracting. This follows from the observations that: (*i*) $\mathcal{H}$ is an increasing function of T, at fixed density, and is largest when $T = \infty$; (*ii*) $1/T$ appears in the Boltzmann factor as a multiplicative factor of the potential V; and (*iii*) $\mathcal{H}$ is proportional to the system size and $\exp\{\mathcal{H}\}$ is a measure of the configuration space "occupied" by the system. This implies that the ratio of the measure of the configuration space occupied by the system in the absence of interaction, $\exp\{\mathcal{H}\}_{T=\infty}$ and in the presence of an interaction, $\exp\{\mathcal{H}\}_{T<\infty}$, diverges as the system size increases. As a consequence, if the dynamics of the system allows one to span the "space" of V completely, the state with $V = 0$ acts as an absorbing state due to its overwhelming entropy, leading to a phenomenon of entropy trapping.

One might likewise expect that a maximal-entropy state in an ecological community at steady state is one characterized by relatively weak interactions among proximally coexisting species. This can be understood in an intuitive manner: There are many more ways of arranging the individuals within an ecological community when they are weakly interacting than when they are strongly interacting. Furthermore, weak interactions are more likely in species-rich plant communities simply because the probability of 2 species encountering one another decreases with increasing number of species.

Fig. 3. Plots of the average abundances of the 20 most abundant species in a quadrat. Dashed line, actual abundances in the forest; solid line, abundances calculated by using the interactions deduced from the stochastic dynamics approach with the interspecific interactions being turned off. Here, we used the birth rates and intraspecific interaction coefficients obtained by the stochastic method described in the article and solved the master equation for the probability distribution of the abundances of each species (see *SI Text* for details). The noninteracting system yields abundances in good accord with data except for the 19th species, *Rinorea sylvatica*, which has much lower abundance than predicted by the noninteracting model. Indeed, unlike other species, the self-interaction coefficient for *R. sylvatica* is small and comparable with its interspecific interaction coefficients. Note also the unusual highly clustered spatial distribution of trees of this species (see *Fig. S2*) reflected in Fig. 1 *Upper Right*. The figure shows that even though the composition of the majority of species in the BCI forest can effectively be described through a noninteracting model, there do exist species for which it is necessary to consider interspecific interactions.

Summary

In summary, we have presented 2 independent approaches for estimating the effective interactions among species as mediated by all other influences and have illustrated their power with calculations on an extensive dataset in the BCI forest. Our results are valid in the vicinity of the present state of the BCI forest. We have considered species that are sufficiently similar in their requirements for limiting resources that growth and survival rates are not greatly affected by the species identity of nearby interacting individuals. Furthermore, the individuals of distinct species are not often in the vicinity of each other to have a direct effect on each other. In addition, it is logically and mathematically difficult to construct a community of many species, all or most of which have strong direct competitive interactions with the other species (31). Indeed, our studies show that, for the BCI forest, the effects of interspecific pairwise interactions between species are relatively weak compared with intraspecific pairwise interactions. These findings help explain why there is good accord of neutral models of noninteracting species with the BCI static and dynamic data.

Our findings, however, do not imply that all interactions are weak. Indeed, one would expect that the interactions between individuals, regardless of the species that they belong to, could be strong, e.g., a large canopy tree suppressing the growth and survival of smaller individuals in its vicinity.

The methodology developed here provides a way to systematically incorporate the most important species interactions into the development of theory beyond the purely noninteracting case. One may speculate that as a general rule, a noninteracting theory will be a better approximation the more species-rich communities become.

Materials and Methods

Maximum-Entropy Method. Entropy is conceptually linked to the amount of disorder or uncertainty in a system: The higher a system's entropy, the less certain one can be about its exact state. This connection is formalized in Shannon's theory of information (15). Our analysis focuses on the probability distribution $P(\vec{n})$, where $\vec{n} = (n_1, n_2, \ldots, n_S)$, and n_i is the abundance of the ith species within a quadrat. The basic idea is to maximize the entropy (15)

$$\mathcal{H}(P) = -\sum_{\vec{n}} P(\vec{n}) \ln P(\vec{n}) \tag{1}$$

while ensuring that all available information pertaining to the system is faithfully encoded. Note that because we are working at the most detailed level of description, the reference entropy can be taken to be constant. For example, one may have information about the mean abundance of a given species averaged over all quadrats or the covariance of abundances of a pair of species may be known a priori. Such information can be readily written in terms of $P(\vec{n})$ and imposed as constraints on the maximization of the entropy written as:

$$\langle Q_r \rangle \equiv \sum_{\vec{n}} P(\vec{n}) Q_r(\vec{n}) = \bar{Q}_r \quad r = 0, 1, 2, \ldots, \tag{2}$$

where the rth constraint requires that the mean value of a quantity Q_r is equal to $\bar{Q}_r$. $r = 0$ is a normalization condition which ensures that $\sum_{\vec{n}} P(\vec{n}) = 1$ and results from selecting $Q_0 = \bar{Q}_0 = 1$. For $r \geq 1$, $\bar{Q}_r$ is obtained from the partial knowledge one has about the system. The logic underlying the variational method follows from the link between information and entropy—the more information one has, the lower the entropy. The entropy is reduced as a result of the partial knowledge encoded in $\bar{Q}_r$. The entropy—maximization principle arises from the observation that the entropy must be the highest possible that includes the available information, because a lower entropy would imply that more information has been incorporated than is available. Using Lagrange multipliers, λ_r, to impose the constraints, one seeks to maximize

$$-\sum_{\vec{n}} P(\vec{n}) \ln P(\vec{n}) - \sum_{r, \vec{n}} \lambda_r P(\vec{n}) Q_r(\vec{n}). \tag{3}$$

Carrying out this procedure, the general solution is found to be (26)

$$P(\vec{n}) \propto e^{-\sum_r \lambda_r Q_r(\vec{n})}, \tag{4}$$

where the λs have to be chosen in order to satisfy the constraints (see Eq. 2). Upon choosing the constraints to be the first 2 moments of $P(\vec{n})$, one finds in the continuum limit the multivariate normal distribution (see *SI Text*):

$$P(\vec{n}) = \frac{1}{(2\pi)^{S/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} (\vec{n} - \langle \vec{n} \rangle)^T \Sigma^{-1} (\vec{n} - \langle \vec{n} \rangle) \right], \tag{5}$$

where Σ is the covariance matrix, $\Sigma_{ij} = \langle (n_i - \langle n_i \rangle)(n_j - \langle n_j \rangle) \rangle$ and S is the number of species. Note that $M \equiv -\Sigma^{-1}$ is a symmetric matrix by construction. Biologically, these constraints correspond to knowledge of the mean abundances of the species and the covariances of all pairs of species. $P(\vec{n})$, which faithfully encodes this knowledge, is deduced by using the maxent principle without incorporating any unwarranted additional information.

The species abundances n_i are treated as continuous variables and not just integers. This allows us to use integrals instead of sums. In order to facilitate the integration, we allow negative n_i. Negative values contribute negligibly to the integrals, and so this is an excellent approximation because large deviations from the average values lead to very small $P(\vec{n})$.

The power of the method lies in the fact that the effective interactions that one obtains are emergent quantities mediated by all the variables left out of the calculation such as the abundances of other species, unmeasured nutrients, and the amount of available sunlight.

Stochastic Approach. The transition probabilities for the birth and the death of an individual of the ith species, within a system with $\vec{n} = (n_1, n_2, \ldots, n_i, \ldots, n_S)$ individuals in species $(1, 2, \ldots, i, \ldots, S)$ respectively, can be written as:

$$T(\vec{n} + \vec{e}_i | \vec{n}) \equiv T_i^+ = b_i n_i + \sum_j c_{ij} n_i n_j, \qquad [6]$$

$$T(\vec{n} - \vec{e}_i | \vec{n}) \equiv T_i^- = d_i n_i. \qquad [7]$$

Here $\vec{e}_i$ is a unit vector whose ith component is 1 with all the other components being zero; b_i, d_i, and c_{ij} are the per-capita birth, death, and interaction rates, respectively. In order to avert a trivial stationary state of full extinction, we employ a reflecting boundary condition at the origin. Our results are insensitive to this for species with moderate-to-large abundances. In order to make a quantitative comparison with the results obtained from the maximum entropy framework, we set $c_{ij} = c_{ji}$. Note that in the maxent method, the term $M_{ij} n_i n_j$ in the exponent is dimensionless; in the stochastic approach, the c_{ij}s are also dimensionless because they are measured in units of d_is, which we have chosen to be 1 by fixing a suitable time unit. Due to fitness equivalence, the b/d ratios of the dominant BCI species are quite similar. We have carried out calculations in which we allowed the death rates to vary between those species by $\approx 10\%$ and found that our results are robust to this variation.

The master equation (32) is given by

$$
\begin{aligned}
\frac{\partial P(\vec{n}, t)}{\partial t} = \sum_i [& T(\vec{n} | \vec{n} + \vec{e}_i) P(\vec{n} + \vec{e}_i, t) \\
& + T(\vec{n} | \vec{n} - \vec{e}_i) P(\vec{n} - \vec{e}_i, t) \\
& - T(\vec{n} + \vec{e}_i | \vec{n}) P(\vec{n}, t) - T(\vec{n} - \vec{e}_i | \vec{n}) P(\vec{n}, t)].
\end{aligned}
\qquad [8]
$$

On multiplying the master equation by n_i (and $n_i n_j$) and summing over all values of $\vec{n}$, one obtains the following equations for the moments of $P(\vec{n}, t)$:

$$\frac{d\langle n_i \rangle}{dt} = \langle T_i^+ - T_i^- \rangle, \qquad [9]$$

$$\frac{d\langle n_i n_j \rangle}{dt} = \langle n_i (T_j^+ - T_j^-) + n_j (T_i^+ - T_i^-) \rangle, (i \neq j), \qquad [10]$$

$$\frac{d\langle n_i^2 \rangle}{dt} = \langle T_i^+ + T_i^- \rangle + 2\langle n_i (T_i^+ - T_i^-) \rangle, \qquad [11]$$

where the average of any function f is given by $\langle f \rangle \equiv \sum_{\vec{n}} f(\vec{n}) P(\vec{n})$. Leaving aside the demographic parameters that we seek to determine, the averages on the right-hand side of the above equations can be estimated directly from the data. For example, Eq. **9** can be rewritten as $\frac{d\langle n_i \rangle}{dt} = (b_i - d_i)\langle n_i \rangle + \sum_j c_{ij} \langle n_i n_j \rangle$, and the averages $\langle n_i \rangle$ and $\langle n_i n_j \rangle$ are obtained by averaging the abundance data in the individual quadrats.

In total we have $S + (S^2 - S)/2 + S$ linear equations and the same number of demographic parameters. Assuming steady state, we solve the equations and find the birth, death, and interaction coefficients. One might extend our approach to determine asymmetric interaction coefficients by adding equations for higher-order moments.

For additional information, see SI Figs. S2–S5.

1. Clements FE (1916) Plant succession: Analysis of the development of vegetation. *Carnegie Institution of Washington Publ* 242 (Carnegie Institution of Washington, Washington, DC), Chap 1.
2. Hutchinson GE (1958) Homage to Santa Rosalia, or Why are there so many kinds of animals? *Am Nat* 93:145–159.
3. Leibold MA, McPeek MA (2006) Coexistence of the niche and neutral perspectives in community ecology. *Ecology* 87:1399–1410.
4. Leibold MA (2008) Return of the niche. *Nature* 454:39–40.
5. Gleason HA (1922) On the relation between species and area. *Ecology* 3:158–162.
6. MacArthur RH, Wilson EO (1967) *The Theory of Island Biogeography* (Princeton Univ Press, Princeton).
7. Chesson P (1997) A need for niches. *Trends Ecol Evol* 6:26–28.
8. Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ Press, Princeton).
9. Volkov I, Banavar JR, Hubbell SP, Maritan A (2003) Neutral theory and relative species abundance in ecology. *Nature* 424:1035–1037.
10. Volkov I, Banavar JR, He F, Hubbell SP, Maritan A (2005) Density and frequency dependence explains tree species abundance and diversity in tropical forests. *Nature* 438:658–661.
11. Volkov I, Banavar JR, Hubbell SP, Maritan A (2007) Patterns of relative species abundance in rainforests and coral reefs. *Nature* 450:45–49.
12. Purvis DW, Pacala SW (2005) Ecological drift in niche-structured communities: Neutral pattern does not imply neutral process. *Biotic Interactions in Tropical Forests: Their Role in the Maintenance of Species Diversity*, eds Burslem D, Pinardand MA, Hartley SE (Cambridge Univ Press, Cambridge, UK), pp 106–138.
13. Chave J, Muller-Landau HC, Levin SA (2002) Comparing classical community models: Theoretical consequences for patterns of diversity. *Am Nat* 159:1–23.
14. Boltzmann L (1964) *Lectures on Gas Theory.* (Cambridge Univ Press, Cambridge, UK).
15. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423.
16. Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:620–630.
17. Jaynes ET (1957) Information theory and statistical mechanics. *II Phys Rev* 108:171–190.
18. Dewar RC (2003) Information theoretic explanation of maximum entropy production, the fluctuation theorem and self-organized criticality in non-equilibrium stationary states. *J Phys A* 36:631–641.
19. Dewar RC (2005) Maximum entropy production and the fluctuation theorem. *J Phys A* 38:L371–L381.
20. Shipley B, Vile D, Garnier E (2006) From plant traits to plant communities: A statistical mechanistic approach to biodiversity. *Science* 314:812–814.
21. Pueyo S, He F, Zillio T (2007) The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecol Lett* 10:1017–1028.
22. Harte J, Zillio T, Conlisk E, Smith AB (2008) Maximum entropy and the state-variable approach to macroecology. *Ecology* 89:2700–2711.
23. Sibisi S, Skilling J, Brereton RG, Laue E, Staunton J (1984) Maximum entropy signal processing in practical NMR spectroscopy. *Nature* 311:446–447.
24. Kitaura R, et al. (2002) Formation of a one-dimensional array of oxygen in a microporous metal-organic solid. *Science* 298:2358–2361.
25. Schneidman E, Berry MJ, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440:1007–1012.
26. Lezon T, Banavar JR, Cieplak M, Maritan A, Fedoroff NV (2006) Using entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc Natl Acad Sci USA* 103:19033–19038.
27. Pielou EC (1969) *An Introduction to Mathematical Ecology* (Wiley, New York).
28. Azaele S, Pigolotti S, Banavar JR, Maritan A (2006) Dynamical evolution of ecosystems. *Nature* 444:926–928.
29. Preparata FR, Shamos MI (2004) *Computational Geometry: An Introduction* (Springer, New York).
30. McKane AJ, Newman TJ (2004) Stochastic models in population biology and their deterministic analogs. *Phys Rev E* 70:041902.
31. Brown JH, Bedrick EJ, Ernest SKM, Cartron JLE, Kelly JF (2004) Constraints on negative relationships: Mathematical causes and ecological consequences. *The Nature of Scientific Evidence*, eds Taper ML, Lele S, Lewin-Koh N (Johns Hopkins Univ Press, Baltimore).
32. Van Kampen NG (2001) *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).