Modeling and Inferring Metacommunity Dynamics with Maximum Caliber

Zachary Jackson*[1], Mathew A. Leibold[2], Robert D. Holt[2], BingKan Xue[1]

[1] Department of Physics, [2] Department of Biology,
University of Florida, Gainesville, FL, United States

Abstract

A major challenge for community ecology is to use distribution patterns to infer basic parameters of dynamical models without conducting laborious experimental manipulations. We present a novel framework drawn from statistical physics – Maximum Caliber – for characterizing the temporal dynamics of complex ecological systems in spatially extended landscapes and inferring parameters from temporal data. As an extension of Maximum Entropy modeling, Maximum Caliber models the probability of possible trajectories of a stochastic system, rather than focusing on system states. We demonstrate the ability of the Maximum Caliber framework to capture ecological processes ranging from near- to far from- equilibrium, using an array of species interaction motifs including random interactions, apparent competition, intraguild competition, and non-transitive competition, along with dispersal among multiple patches. For spatio-temporal data of species occurrence in a metacommunity, the parameters of a Maximum Caliber model can be estimated through a simple logistic regression to reveal migration rates between patches, magnitudes of interactions between species, and effects of intrinsic local environmental suitabilities. We test the accuracy of the method over a range of system sizes and time periods, and find that these parameters can be estimated without bias. We introduce "entropy production" as a system-level measure of disequilibrium, and use "pseudo-R^2" to characterize the predictability of the system. We show that our model can predict the dynamics of metacommunities much better than steady state models, when the system is far from equilibrium. The capacity to estimate basic parameters of dynamical metacommunity models from spatio-temporal data represents an important breakthrough for the study of metacommunities with application to practical problems in conservation and restoration ecology.

Introduction

The basic processes that drive community changes are relatively simple (Vellend 2010; Vellend 2016), consisting of dispersal, "abiotic filtering" (density independent selection among species), "biotic filtering" (density/frequency dependent selection), stochasticity, and the generation of "novelty" (especially speciation). However, the resulting species distributions over the long term arise from complex dynamics that reflect the wide-spread "entanglement" of these causal processes (Darwin 1859; Kéfi et al. 2016). It is possible, using various theoretical approaches, to develop hypotheses and derive predictions about steady-state outcomes that result from particular model assumptions and parameterizations. However, the inverse problem, using data to infer the processes that determine these patterns, is much more difficult (Leibold et al. 2025). This is because different models often make convergent steady-state predictions, and also because disentangling causal effects is not always possible using current approaches based on data about spatial distributions alone.

Most prior work to derive predictions about community patterns has used population models based on the classic Lotka-Volterra formulation of species interactions (Lotka 1925; Volterra 1928), expanded to multi-species communities (May 1972; Novak et al. 2016) and multiple patches and trophic levels (Gravel, Massol, and Leibold 2016). These models are difficult to solve because they are non-linear in their formulation, especially when dispersal among communities occurs. A simpler modeling approach ignores variation in abundances and focuses on occupancy (presence or absence) instead. Such models examine how colonizations and extinctions affect community patterns (Levins and Culver 1971; Horn and MacArthur 1972; Holt 1997; Leibold, Govaert, et al. 2022), typically focusing on the eventual steady-state behavior of species distributions. Here we use occupancy models, but we use their spatio-temporal dynamics, both near steady state and far from equilibrium, to infer causal relations in metacommunities. With changes in data pipelines and development of statistical methods, there is an increasing access to spatio-temporal data in ecology (Hartig et al. 2024), permitting the use of methods that take advantage of temporal trajectories to better identify causal relations. Purely spatial models conventionally focus on a basic sample dataset (lists of species present at a particular patch at a particular time) that characterize the "states" of communities. However, an alternative approach is to focus on changes in these states from one time point to the next, consisting of "transitions". An accumulation of transitions over multiple time steps forms "trajectories", which can characterize both the equilibrial outcome of community dynamics and transient, far-from equilibrium behaviors. This is the basic approach we explore here.

We introduce a method from statistical physics that is focused on modeling such trajectories, known as Maximum Caliber (Jaynes 1980; Stock, Ghosh, and Dill 2008). It has been applied to stochastic processes in a variety of research areas, including thermal and chemical dynamics, protein folding, neural firing, network traffic flows, and more (Pressé et al. 2013; Ghosh et al. 2020; Wan, Zhou, and Voelz 2016; Firman, Balázsi, and Ghosh 2017; Mora, Deny, and Marre 2015). Here we apply Maximum Caliber to metacommunity ecology and find that this approach of modeling the behavior of metacommunities allows a straightforward way to infer parameters that can be interpreted as basic ecological processes. This method thus opens up previously unavailable, but critically important, linkages between observable patterns in metacommunities and the underlying processes that are responsible for them.

We first derive a general mathematical framework for using Maximum Caliber to study community assembly and dynamics in metacommunities. We then apply this approach to a representative,

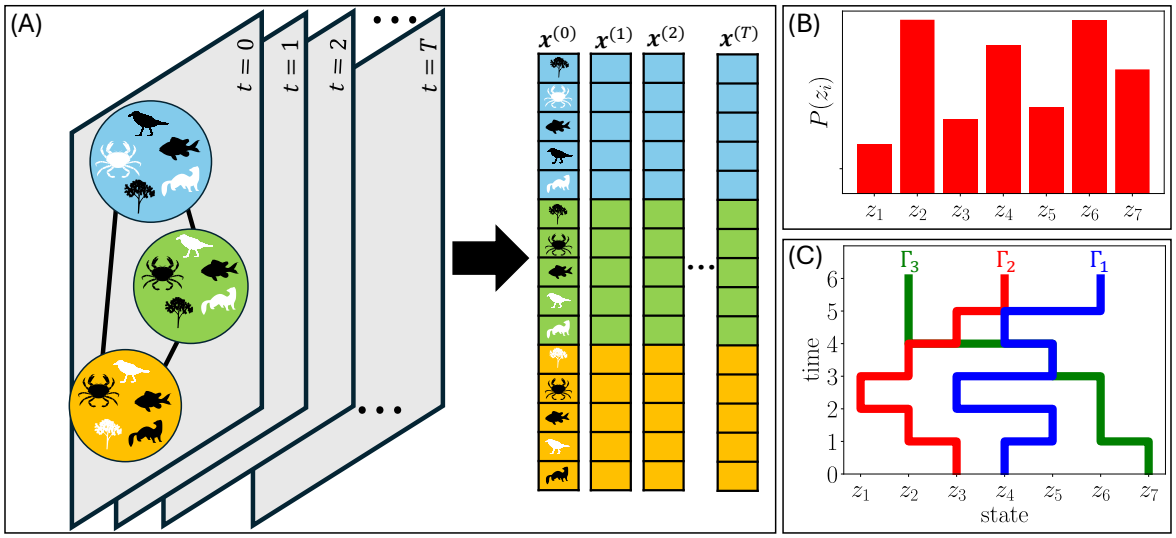

Figure 1: The Maximum Caliber framework for modeling spatio-temporal dynamics of a metacommunity. (A) Measurements of species presence/absence in a metacommunity are presented as a series of binary vectors, one for each time step. (B) Steady state models of the system (such as Maximum Entropy) calculates the probability $P(z_i)$ of each possible state z_i of the system. (C) The Maximum Caliber model calculates the probability of each possible trajectory Γ_n of the system through the state space.

albeit simple, set of models that involve three species in three discrete patches, ranging from fully random to highly deterministic interactions, so as to demonstrate the utility of Maximum Caliber in describing and analyzing metacommunity dynamics. We quantify the irreversibility of their dynamics by calculating the average rate of entropy production (Wang and Qian 2020). We then show that Maximum Caliber can make accurate predictions of data, as measured by the pseudo-R^2 value (described below), especially for systems far from equilibrium. An important advantage of our approach is that we can make unbiased estimation of model parameters from observations using simple logistic regression. Of course, the accuracy of the model crafted using Maximum Caliber depends on the quality and quantity of the data, but we expect that realistic results can often be obtained with the kind of data that are increasingly available to ecologists.

Methods

We consider a simplified view of metacommunity dynamics (Fig. 1) that has a venerable history in community ecology (Levins and Culver 1971; Horn and MacArthur 1972; Levin 1974; Hastings 1980; Hanski 1983; Polis, Anderson, and Holt 1997; Leibold and Loeuille 2015; Leibold, Govaert, et al. 2022). This perspective assumes that there are different discrete locations ("patches"), which can potentially be occupied by species that are subsets of a defined regional "biota". These patches can differ in local environmental conditions that define the "abiotic selection" affecting different species, modeled as the intrinsic capacity of species to colonize the patch (assuming the absence of other species), and the intrinsic vulnerability of species to go extinct (also assuming the absence of other species). These intrinsic capacities are modified by the presence of other species due to species interactions, because such other species alter either the colonization probability or the extinction

probability of any given species relative to what would be the case in patches not occupied by these other species. Here, we also assume that dispersal plays a role, either in relation to inter-patch connectivity (perhaps due to physical distance), or due to a more diffuse (typically much smaller) "global dispersal" that is uniform across all patches. Colonization, and extinction processes are all modeled as probabilities, so that there is a substantial amount of stochasticity in the model dynamics, the magnitude of which depends on the parameter values.

These processes when iterated generate spatio-temporal lists of species at different patches. These spatio-temporal datasets can be represented as state vectors with elements, either 0 or 1, describing the occupancy of patches by individual species at any given time point. However, the data can also be viewed as describing transitions among such states from one time point to the next. At a single patch with a potential pool of N species, the state of the local community at a given time can be one of 2^N possible states, depending on whether each species is present or not, and the possible number of transitions is $2^N \times 2^N$. The dynamics of the community can therefore be described by transitions of these states over time. If we have N species distributed among M patches, the number of states that describe the entire metacommunity is 2^{MN}, and the number of possible transitions is $2^{MN} \times 2^{MN}$. Deriving these probabilities by brute force is consequently a daunting task, except for cases with very few species and very few patches.

One approach in this situation is to estimate the marginal probability of each state and then use that probability to predict future states, regardless of the likelihood of state transitions. That is, we can assign probabilities to each of the possible states of the system according to their observed frequencies, regardless of the temporal context (Fig. 1B top panel). Maximum Entropy (Jaynes 1957; Shipley 2010; Harte and Newman 2014) is one modeling scheme that parametrizes the marginal probabilities for the states of a system, while attempting to bring in as few additional assumptions as possible. Maximum Entropy can be successful for describing and predicting the features of systems that are at or near equilibrium, but it is less informative otherwise. Yet, ecological systems are often far from equilibrium (DeAngelis and Waterhouse 1987; Hastings et al. 1993) and can have complex dynamics that are not captured by marginal probabilities.

To obtain a more effective model that accounts for non-equilibrium spatio-temporal community dynamics, we turn to the Maximum Caliber framework that has been developed for modeling random processes in physics (Jaynes 1980; Pressé et al. 2013; Ghosh et al. 2020). Maximum Caliber is a modeling scheme similar to Maximum Entropy, but it considers probabilities of possible trajectories of the system rather than probabilities of states. In the Maximum Caliber approach, predicting the future state of a system can be thought of as finding the most likely trajectory of the system in the state space over time. By focusing on the system trajectories rather than states, Maximum Caliber can capture dynamics of systems far from equilibrium as well as near or at steady states.

Maximum Caliber

The aim of this approach is to develop a model for system trajectories that is consistent with a given set of observational constraints, but is otherwise as unbiased as possible. If we choose the constraints to be the temporal correlations between consecutive states, then Maximum Caliber can generate a model parametrized by a matrix of effective species interactions, whose entries are interpretable as the rates of basic biological processes governing transitions among states. These parameter values are inferable by a simple logistic regression and thus provide direct inference of model parameters

from the data in a way that has not been possible for the classic Lotka-Volterra model.

We start with a single patch subject to colonization by an external pool of species, labeled by $i = 1, \cdots, N$. We then imagine that we have at hand the list of all species present at different time points. At each time point, the species list can be represented by a set of N binary variables, $x_i = 1$ or 0, according to whether species i is present or absent. The state of the system is thus represented by an N-dimensional binary vector $\mathbf{x}$ at a given time. The data from all time points then forms a time series of such binary vectors, i.e., a trajectory in N-dimensional space. We will label the community state at time t by $\mathbf{x}^{(t)}$, and denote a trajectory in the state space by $\Gamma = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(T)}\}$, where T is the length of the time series.

We seek to model the probability of different trajectories, $P(\Gamma)$, that are consistent with certain statistical properties of the observed data. We will focus on the frequency of occupancy of each species, f_i, and the cross-correlation between pairs of species at consecutive time points, c_{ij}. We thus look for a model of $P(\Gamma)$ that satisfies these constraints but is otherwise as unbiased as possible. This is done by maximizing the entropy of the distribution $P(\Gamma)$, under the constraints that the expected values of f_i and c_{ij} match the observed values from the data (see Box 1).

The Maximum Caliber solution takes the form:

$$P(\Gamma) = \prod_t \prod_i P\big(x_i^{(t+1)} \big| \mathbf{x}^{(t)}\big), \quad \text{where} \quad P\left(x_i^{(t+1)} = 1 \middle| \mathbf{x}^{(t)}\right) = \frac{1}{1 + \exp\left\{-\sum_j \eta_{ij}\, x_j^{(t)} - \lambda_i\right\}}. \quad (5)$$

This implies that the presence or absence of each species in the subsequent time point is conditional only on the community composition in the previous time point. Therefore, this is a special case of a Markov model — instead of being parametrized by a $2^N \times 2^N$ transition matrix $P\big(\mathbf{x}^{(t+1)} \big| \mathbf{x}^{(t)}\big)$, our model is parametrized by a much smaller $N \times N$ matrix η_{ij} and a N-dimensional vector λ_i. Here the parameters η_{ij} can be interpreted as interactions between pairs of species i and j (including effects of species i on itself that describe intrinsic relations to the abiotic environment), and the parameters λ_i that are related to the colonization of the patch by species i from an external pool.

Since the conditional probability in Eq. (5) takes the form of a logistic function, the parameters η_{ij} and λ_i can be estimated from a series of logistic regressions, one for each species i. Each time step provides a data point with $\mathbf{x}^{(t)}$ (the presence/absence of every species at time t) as the explanatory variables and $x_i^{(t+1)}$ (the presence/absence of species i at time $t+1$) as the outcome variable. The η_{ij}'s are the regression coefficients and λ_i's are the intercepts.

We can expand our model from characterizing a single patch to one that includes multiple patches that are interconnected by dispersal. That is, species can colonize individual patches not only from an external pool, but also from other patches. To describe such a metacommunity, the label i now represents a species-patch combination, so that $x_i^{(t)}$ represents whether a particular species is present in a given patch at a given time t. For practical purposes (to avoid complications implied by global extinctions), we also maintain the influx of species, independently of their presences in any of the local patches, as a regional immigration parameter λ_i. Meanwhile, the different entries in the parameters η_{ij} represent interactions between species within a patch, or rates of species-specific migration between patches (both can be asymmetric).

More specifically, the η_{ij} parameters form a matrix that is divided into blocks according to local communities, as illustrated in Fig. 2. The diagonal blocks represent within-community interactions as in the case of a single community described above. Among those, the diagonal entries of the

Box 1: Maximum Caliber Modeling

Here is a brief description of standard derivation of a Maximum Caliber model (Pressé et al. 2013). We start from the entropy of the probability distribution $P(\Gamma)$ of all possible trajectories:

$$\mathcal{H} = -\sum_{\Gamma} P(\Gamma) \log P(\Gamma) \,. \tag{1}$$

We maximize the entropy under constraints on the statistics of the trajectories, here chosen to be the occupancy frequency of each species, f_i, and the cross-correlation between pairs of species at consecutive time points, c_{ij}. For a given trajectory Γ, these quantities can be calculated as

$$f_i(\Gamma) = \frac{1}{T} \sum_{t=1}^{T} x_i^{(t)} \quad \text{and} \quad c_{ij}(\Gamma) = \frac{1}{T} \sum_{t=1}^{T-1} x_i^{(t+1)} x_j^{(t)} \,. \tag{2}$$

Suppose f_i^{obs} and c_{ij}^{obs} are the values calculated from the observed data. We assume that these are the expected values of f_i and c_{ij} according to the probability $P(\Gamma)$, i.e.,

$$\sum_{\Gamma} P(\Gamma) f_i(\Gamma) = f_i^{\text{obs}} \quad \text{and} \quad \sum_{\Gamma} P(\Gamma) c_{ij}(\Gamma) = c_{ij}^{\text{obs}} \,. \tag{3}$$

Introduce Lagrange multipliers to impose these constraints on $P(\Gamma)$, the function to be maximized becomes

$$\begin{aligned} \mathcal{L} = &-\sum_{\Gamma} P(\Gamma) \log P(\Gamma) + \sum_{ij} \eta_{ij} \left(\sum_{\Gamma} P(\Gamma) c_{ij}(\Gamma) - c_{ij}^{\text{obs}} \right) \\ &+ \sum_{i} \lambda_i \left(\sum_{\Gamma} P(\Gamma) f_i(\Gamma) - f_i^{\text{obs}} \right) + \mu \left(\sum_{\Gamma} P(\Gamma) - 1 \right) . \end{aligned} \tag{4}$$

Among the Lagrange multipliers, λ_i and η_{ij} constrain the frequencies and cross-correlations, and μ constrains the sum of the probabilities to 1. Maximizing this function (see more details in Appendix) leads to the solution in Eq. (5).

diagonal blocks (green) represent intraspecific effects that determine the persistence of a species over a time step at a given patch given that it is already present. If the entries corresponding to the same species are identical among the patches, there is no environmental heterogeneity; otherwise, the differences among patches would describe the effects of such heterogeneity on that species. As in the case of a single community, the non-diagonal entries of the diagonal blocks (blue) represent interspecific interactions, which can also vary among the patches. Finally, the diagonal entries of the non-diagonal blocks (red) represent possible dispersal of species between two communities. The non-diagonal entries of the non-diagonal blocks (gray) are expected to be zero (as otherwise they would represent dispersal that changes species identities). On the other hand, the λ_i parameters form a vector that also has multiple sections corresponding to different patches. These parameters represent the baseline colonization rates into the patches from the external species pool. If there are no differences among patches, then the entries of the vector λ that correspond to the same species at different patches would be equal. Otherwise, differences among these entries could represent en-

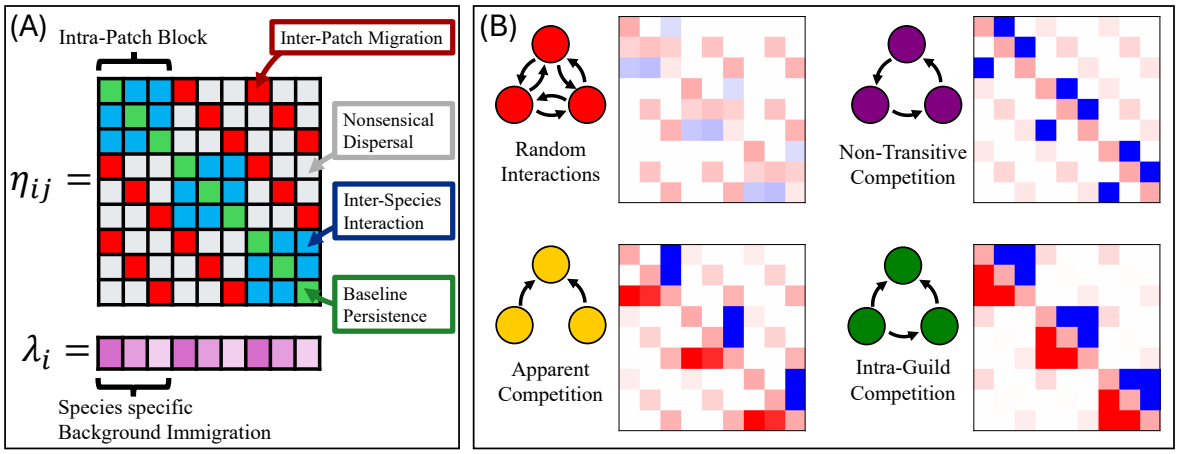

Figure 2: Structure of Maxiumum Caliber model parameters, interpretable as rates of ecological processes. (A) The η_{ij}'s represent the interactions between (blue) and within (green) species in the same patch, as well as rates of colonization among patches (red). The λ_i's correspond to the rate of background colonization of a species to a patch. (B) Four different community structures (motifs) used to test Maximum Caliber modeling. Each example system consists of three species found in three possible patches. The heatmaps show example η_{ij} values (red for positive and blue for negative) for each motif.

vironmental variation, such as variation in abiotic conditions among patches that affect colonization success, independent of species interactions.

Examples of community structures

To demonstrate the effectiveness of the Maximum Caliber approach, we consider four different community structures for three species in each of three patches, as illustrated in Fig. 2B. The first example is a "random interactions" model inspired by random Lotka-Volterra models (May 1972, etc.), where the within-community interspecific interactions are randomly chosen from a normal distribution with a given mean and variance. The second example is a trophic structure known as "apparent competition" (Holt 1977), where two prey species share the same predator. The third example is known as "intraguild predation" (Holt and Polis 1997), which includes a generalist consumer that can consume both a prey and its basal resource. The last example will be called "non-transitive competition", where each species is strongly excluded by one other species, like in a rock-paper-scissors game (May and Leonard 1975; Sinervo and Lively 1996). Fig. 2B shows the structure of the η_{ij} matrix for each type of metacommunity with three patches and randomly chosen dispersal parameters between them. (A graph representation of state transitions in each type of community is shown in Supp. Fig. S1)

Some of these examples may appear to have similar community structures, as one can be transformed into another by adding or flipping one arrow in their motif diagrams (Fig. 2B). However, they reveal qualitatively different dynamics in terms of how far they are from equilibrium (as will be quantified later). We use Maximum Caliber to model all these community structures and to assess the ability of Maximum Caliber to correctly infer the driving parameters of the model.

Results

Accuracy of parameter inference

Maximum Caliber is particularly useful because it provides a method for inferring the parameters of the occupancy model. An important concern is the degree to which such inferences are precise and unbiased. We tested the inference of parameters using logistic regression from simulation data over a broad array of simulations that varied the number of patches, the number of species, and the duration of the time series. To do so, we simulated data resulting from selecting the parameters η_{ij} and λ_i subject to the assumptions of each motif, and then we used logistic regression on the simulation data to estimate the parameter values $\hat{\eta}_{ij}$ and $\hat{\lambda}_i$. We next compared these estimated parameters with the original values η_{ij} and λ_i used to generate the simulations. We tested the accuracy of parameter inference using the random interactions model (the choice of parameters is described in the appendix); the other motifs yielded similar results.

To quantify the accuracy of our inferences, we compared the estimated values of the parameters against the original values by linear regression. The slope and intercept of the regression indicates whether there is bias in the parameter inference, and the residual variance measures how accurate we can expect any particular parameter estimation to be. Examples of such comparisons are shown in Fig. 3. The slopes and intercepts are very close to 1 and 0 respectively, and the residual variances are small. Crucially, these inferences appear equally good for all the parameters and indicate no bias in the estimates.

We expect the accuracy of our inference to depend positively on the length of the time series (T) and negatively on the system size (the number of interconnected patches M and the number of species N). Plots of how these measures of accuracy vary as the length of the time series, number of species, and number of patches change are shown in Fig. 3. These figures confirm our expectations and provide a guide as to how many data points are needed for estimating parameters in applications to real systems (see Discussion).

Irreversibility and entropy production

Maximum Caliber can model both dynamics that are at (or near) steady-state and possible transient dynamics that are not. By "steady-state" we mean that the *probability* of a species occurring at a given patch is stationary over time. This does not mean that the species has to be always present or always absent at the patch, only that its frequency of occurrence (over short periods) stops changing (over the long term). In contrast, "transient dynamics" are those during which the *probability* of some species occurring at a given patch changes over time. The dynamics of a system can be called "at equilibrium", or "time reversible", if the probability of any state transition is equal to that of the reverse transition (Li et al. 2024). We note that a steady-state is not necessarily at such equilibrium, whereas transient dynamics are even further from equilibrium, because the trajectories are clearly directional.

To capture how far the dynamics is from equilibrium, we borrow the definition of the average rate of "entropy production" from statistical physics (Wang and Qian 2020). This quantity has been used to measure the "irreversibility" of dynamical systems in physics and biology (GrandPre, Teza,

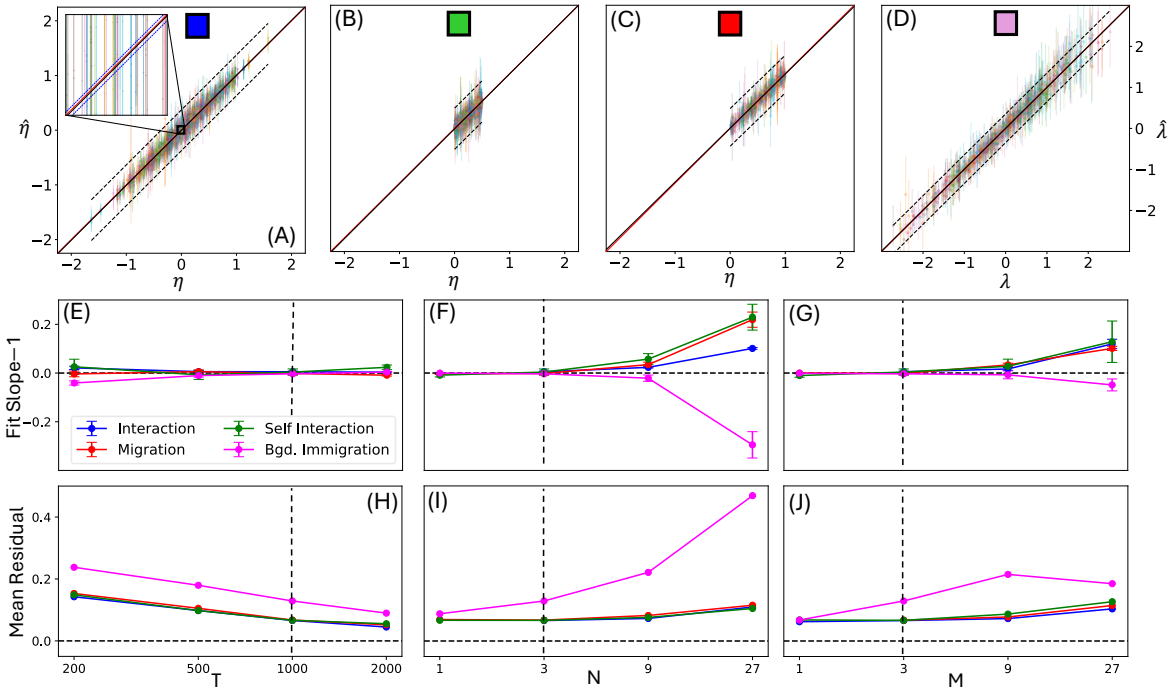

Figure 3: Accuracy of parameter inference using logistic regression. (A)-(D) are plots of the inferred versus true parameter values for a system with the number of species $N = 3$, the number of patches $M = 3$, and the length of time series $T = 1000$. The choice of random parameter values are described in the Appendix. The red line is slope 1 representing perfect inference. The black solid line is the line of fit (which largely overlaps with the red line). The dotted curves show the observational confidence interval for one standard deviation. Panel (A) shows the interspecific interactions, panel (B) shows the baseline persistence, panel (C) shows the migration rates, and panel (D) shows the background immigration rates (see Fig. 2A). The inset in panel (A) shows the mean confidence interval for the linear weighted regression (which is almost invisible in the main plot). Panels (D)-(F) show how the inference accuracy scales with the time period T and system size N and M. Panels (D), (E), and (F) show the error in the slope of weighted least squares fit of inferred parameter values against true values. Panels (H), (I), and (J) show the mean difference between the inferred parameter values and their true values. Vertical dashed lines in (E)-(J) correspond to the parameters used in panels (A)-(D).

and Bialek 2024). For a stationary process, entropy production can be calculated as:

$$E_{\mathrm{p}} = \sum_{\mathbf{x},\mathbf{x}'} P(\mathbf{x}) P(\mathbf{x}'|\mathbf{x}) \log \frac{P(\mathbf{x}'|\mathbf{x})}{P(\mathbf{x}|\mathbf{x}')} \tag{6}$$

where $P(\mathbf{x})$ is the marginal probability of the system being in the state $\mathbf{x}$, and $P(\mathbf{x}'|\mathbf{x})$ is the conditional probability of the system transitioning from $\mathbf{x}$ to $\mathbf{x}'$. Roughly speaking, the entropy production E_{p} measures the log-likelihood ratio of the trajectories forward and backward in time.

The entropy production is zero when the system is at equilibrium, and is positive when the system is out of equilibrium. Because its calculation involves the ratio of transition probabilities, if these probabilities were to be estimated directly from observed time series, it would take a prohibitive amount of data. However, in our model, the entropy production can be calculated from the estimated values of the η_{ij} and λ_i parameters by using them to calculate the transition probabilities according to Eq. (5). This means that once a set of parameters has been inferred for a system, the entropy production can be calculated even if not all possible state transitions have been observed in a finite dataset, as long as the parameter inference is accurate.

Examples of the distribution of the entropy production for different realizations of each type of system is shown in Fig. 4A. We can see that the Maximum Caliber framework can capture a wide range of qualitatively different dynamics, from the near-equilibrium of the random interactions model, to the near-deterministic dynamics of the non-transitive competition model, and intermediates in between.

Predictability and Pseudo-R^2

There is a distinction between how significant the estimated parameters of a model are, and how well the model can predict future outcomes of a system. In linear regression, for example, we can have an estimated value of the slope that is significantly nonzero if we have a large amount of data. However, we may still have a large variance of the residual, which would mean that there is a lot of uncertainty in our prediction even though our model parameters are accurate. One way to measure the predictability of a linear regression model is to use the coefficient of determination R^2, which is bounded between 0 and 1. The larger R^2 is, the greater the proportion of variation captured by the model, and the better the response variable is predicted by the explanatory variables.

Our Maximum Caliber model is equivalent to logistic regression; rather than continuous variables, we have categorical variables for both response and explanatory variables. To measure the variation captured by our model, we use a generalization of R^2 for logistic regression, called "pseudo-R^2" (Mittlböck and Schemper 1996). Several definitions of this quantity exist; a convenient one due to (Cox and Snell 1989) is calculated as:

$$R^2_{\mathrm{p}} = 1 - \left(\frac{L_M}{L_0}\right)^{-2/n} = 1 - \mathrm{e}^{-D/n} \tag{7}$$

where L_M is the likelihood of the model inferred by logistic regression (using maximum likelihood estimation), and L_0 is the likelihood of a null model that uses only the marginal probabilities of the response variables. Here n is the number of data points, and $D = 2\log(L_M/L_0)$ is known as the "deviance" of the inferred model. This definition of pseudo-R^2 is consistent with the standard R^2 in

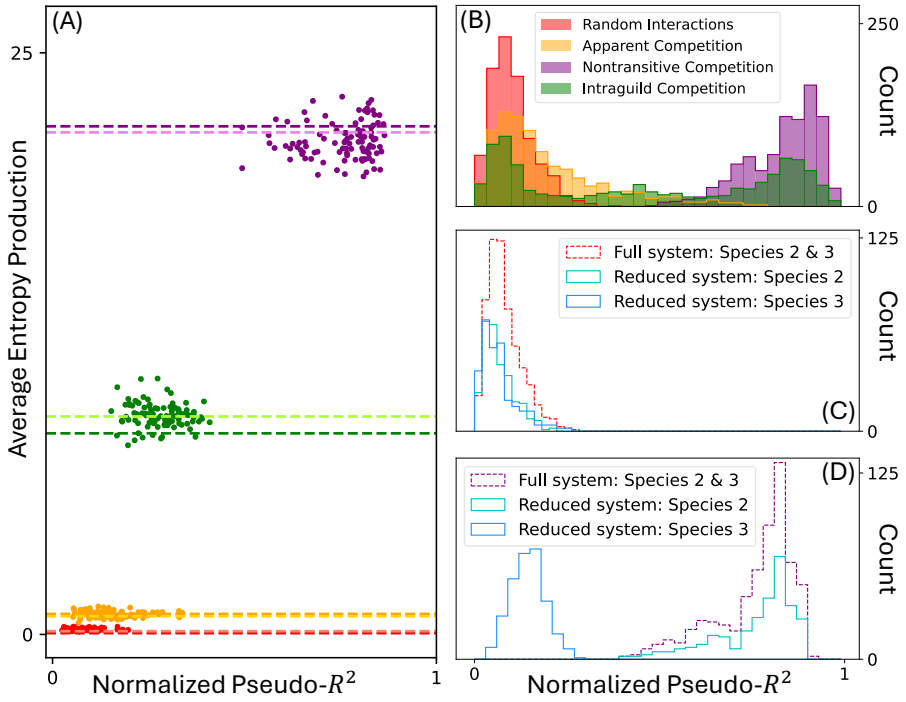

Figure 4: Statistical measures of irreversibility and predictability of the system dynamics. (A) The average rate of entropy production versus normalized pseudo-R^2. Colors represent different motifs in Fig. 2B; points represent replicate simulations with randomized parameter values. The larger the entropy production is, the further from equilibrium a system is. The larger the normalized pseudo-R^2 is, the bigger fraction of variation is captured by the Maximum Caliber model. (B) The distribution of R^2_{np} values for each species-patch pair in systems with different community structures. (C)-(D) Comparison of the distribution of R^2_{np} values between the full system and a reduced system that is missing one of the species (Species 1). In panel (D), Species 3 is the one strongly inhibited by Species 1, and its predictability is strongly affected if Species 1 is missing from the data.

the case of linear regression. However, a major difference between the standard R^2 and pseudo-R^2 is that pseudo-R^2 ranges between 0 and a maximum value that is less than 1. For logistic regression, the maximum value is given by $R^2_{\mathrm{p,max}} = 1 - L_0^{2/n} \leq 0.75$ (Nagelkerke 1991). A modified definition of pseudo-R^2, due to (Nagelkerke 1991), normalizes R^2_{p} by $R^2_{\mathrm{p,max}}$, so that it ranges between 0 and 1; we will call this the "normalized pseudo-R^2", or R^2_{np}.

We will use the normalized pseudo-R^2 to measure how well our inferred model makes predictions about the presence or absence of a species compared to using only its marginal probability. In Fig. 4B we show the distribution of the R^2_{np} values for each of the four types of our simple community interaction structures. Note that the quantity $R^2_{\mathrm{p,max}}$ measures how much variation there is in the data beyond the marginal probabilities, whereas R^2_{np} measures how much of that variation is captured by our Maximum Caliber model (for their distinction, see Supplementary Figure S2). In general, we found that our Maximum Caliber model can capture a larger fraction of the variation in systems that are far from equilibrium than it can in systems that are near equilibrium.

Moreover, the pseudo-R^2 value can be calculated for each species alone by using the likelihood function in Eq. (5) for a single species. In Fig. 4B we show the distributions of R^2_{np} for each species in different community structures. In the intraguild predation motif, for example, we see that not

all species within a system are equally predictable. The top predator (the leftmost peak) turns out to have lower R^2_{np} values than the other two and so is less predictable than the other two species. The variation in the R^2_{np} values observed for any particular system comes from a combination of system structure, true parameter differences, and data realization (see Supplementary Fig. S2).

Missing species

In empirical studies, there is no guarantee that all species involved in ecological interactions are sampled. Some species may be cryptic, or hard to identify, or may belong to taxonomic groups that are not sampled for pragmatic reasons. Such missing data may affect how well we can make predictions about the species that are actually observed. We can investigate the effect of such missing species using the simulations of the 3-species, 3-patch metacommunities explored above. To do so, we removed one species from the simulated data, then re-estimated the parameters of a model involving only the remaining two species. We examined the random interactions motif and the non-transitive competition motif, as they represent the two extremes of the community dynamics. Because there is a symmetry among all three species within each motif, we removed the first species without loss of generality. To compare the results with the full model studied above, we re-calculated the R^2_{np} values for each of the two remaining species. Shown in Fig. 4C,D is a comparison of the original distribution of R^2_{np} values for these two species in the full model and the recalculated R^2_{np} values after one species is removed. In the case of random interactions (Fig. 4C), we see only a slight shift of R^2_{np} to lower values for each of the remaining species. Thus, missing any particular species in a system of this type has a weak effect on the predictions for the remaining species. In the case of non-transitive competition (Fig. 4D), we see that missing species 1 in the data has almost no effect on the predictions made about species 2, but a strong effect on species 3. This is because species 1 is the one that strongly excludes species 3, making it an important factor in determining the presence of species 3, but less so for species 2. Our results suggest that, when species are missing or not included in empirical data, the effect on the predictability of the other species is not uniform.

Discussion

We have introduced the method of Maximum Caliber for community ecology, as illustrated by a metacommunity occupancy model that corresponds to previously well-studied scenarios. There are two important advances made possible by Maximum Caliber. The most important of these is that it allows for unbiased inference of dynamic parameters from observed spatio-temporal data using simple logistic regression. A second important advantage is that inference is not limited to steady-state or equilibrium situations, but that it can be carried out for transient dynamics. As the relevance of steady-state and equilibrium assumptions in (meta)community ecology are increasingly questioned (DeAngelis and Waterhouse 1987; McCann 2000; Wu and Loucks 1995), our method may provide a useful approach to study community ecology, such as in applications where human impacts are rapidly increasing. The use of simple logistic regression also means that calculations are more transparent than are other existing methods for metacommunity analysis, allowing the parameters to be easily interpretable in terms of biological processes involved in the (meta)communities.

We have also introduced two useful metrics to characterize the dynamics of metacommunities. One of these is the entropy production rate of the system, calculated from observed data. This can help quantify how far from equilibrium (or irreversible) a given system might be, and help identify systems whose inferred dynamics have no long-term steady state (such as in non-transitive competition). The other useful metric is the pseudo-R^2 that can be calculated for the system as a whole and for each species separately. This metric can quantify the predictability of future states, given the inferred dynamics of the system. Because we can calculate this for each species, it provides a metric that can reveal the "internal structure" of the metacommunity (*sensu* Leibold, Rudolph, et al. 2022). In conservation ecology, this can be useful by identifying species that are robust/sensitive to perturbations, thereby informing the crafting of specific conservation measures (Horvath et al. in progress), and it may be more broadly useful in the increasingly urgent issue of ecological forecasting (Petchey and Gaston 2006).

Nevertheless, several issues need to be considered when applying the method of Maximum Caliber. First is the amount of available data. Since we are inferring the *dynamics* of metacommunities, it is important to have long time series. As shown by our simulation results, at least tens or hundreds of time points are needed to infer parameters with any confidence. This number increases with the size of the species pool and the number of patches in the metacommunity. One way to gather a sufficient number of data points is to have replicate experiments or replicate time series, even if each time series does not run for a long time. The parameter inference method is such that we can compile data points from different trajectories together in a single logistic regression, as long as we believe that the interaction parameters are the same across these trajectories.

Another issue is missing species. The causal role of missing species might be folded into "effective parameters" between explicit species, absorbing some of the effects of the missing species. We have seen above that, even with missing species, the dynamics of some of the present species can still be predicted quite accurately, as measured by the normalized pseudo-R^2 metric for these species. In the examples we studied, the predictability of a species is significantly reduced if the missing species exert strong interactions on that species. The generality of this trend needs further studies. A related issue is when a species is always (or almost always) present or absent everywhere in the community. This would cause the regression algorithm to falter because of the lack of variation in the input value. In this case, it would be useful to remove that species from the pool. The effect of that species will then be absorbed into the background (i.e., the λ_i parameters of the other species). In practice, the species pool can be pre-screened to remove such "passive" species that act as a background for the other species. Note that we are not claiming that rare species do not play an important role in ecological processes, but that our method may not be sensitive to such species.

When using the Maximum Caliber approach to derive our model, we have chosen to use the species occurrence frequency f_i and the cross-correlation c_{ij} at consecutive time points as constraints. These are convenient choices in that they are readily calculated from temporal presence-absence data, and in that the resulting model has a simple parametric form and ecological interpretation. However, these choices are not unique, and different choices will lead to different parametric models. For instance, one could calculate the cross-correlation between two species at longer time lags than consecutive time points. Including such a new constraint in the maximization of trajectory entropy will lead to a new term in the logistic function (Eq. 5) that may represent species interactions over longer timescales. In practice, one could try including time-lagged variables in the logistic regression

and test whether they are significant, similar to testing for Granger causality (Tank et al. 2021). This will likely require a more substantial amount of data.

Similarly, because we used cross-correlation between *pairs* of species as constraints, the resulting Maximum Caliber model included only pairwise interactions between different species. Higher-order interactions, such as those induced by trait-mediated indirect effects (Ohgushi, Schmitz, and Holt 2012; Werner and Peacor 2003), can involve three or more species and may have important consequences in some ecosystems (Mickalide and Kuehn 2019; Goudard and Loreau 2008). Such interactions will result from imposing higher-order statistics (higher mixed moments) as constraints in building the Maximum Caliber model. It is possible to have many forms of such higher-order interactions, even among three species, by considering different combinations of present and time-lagged variables, such as $x_i^{(t+2)} x_j^{(t+1)} x_k^{(t)}$, $x_i^{(t+1)} x_j^{(t+1)} x_k^{(t)}$, or $x_i^{(t+1)} x_j^{(t)} x_k^{(t)}$. Models with such terms will become increasingly complex and require extensive amount of data to make reliable inferences; their interpretation in terms of ecological processes will also be more challenging. Yet, there are methods from Maximum Entropy models that can potentially be used to selectively introduce such terms based on their statistical significance (Cayco-Gajic, Zylberberg, and Shea-Brown 2018).

There are many areas for future work using Maximum Caliber that one can imagine. It can be used to study ecological succession, such as how secondary succession compares to primary succession, how multiple trophic levels and their interactions affect community assembly, and how the assembly of microbes in host-associated microbiomes is determined. Our method could be applied to any number of existing longitudinal community studies such as those from LTERs and NEON; and novel monitoring studies promise that such data will be increasingly available in the future. One such arena where, in the future, long time-series might be abundantly available is in microbiome research, where 'patches' are individual hosts harboring communities of symbionts.

An important consideration in preparing the data for analysis is sampling frequency. Although our method can be used to model even transient dynamics, it could miss features if sampling is too sparse, leaving out dynamical details in between time points. On the other hand, sampling too frequently and modeling interactions at only consecutive time points may miss out processes happening over longer time scales. Some systems may have clear intrinsic timescales, such as those with seasonal dynamics, in which case the choice of sampling frequency is more obvious.

In summary, Maximum Caliber provides a novel, versatile framework for the analysis of metacommunity dynamics involving diverse biotas interacting in complex landscapes that feature environmental heterogeneity, spatial organization, and stochasticity. It can provide both dynamical models that can make detailed predictions about patterns, and a tool for inferring parameters from a common type of metacommunity data. In doing so Maximum Caliber promises to substantially move the field of community ecology in ways that can provide better insights about nature and facilitate its wise management and conservation.

Acknowledgements

This research is supported by the NSF grants DMS-2330970 and 2025118, the University of Florida Foundation Fund, and the ;University of Florida Invasion Science Research Institute. We thank members of the Leibold lab for useful comments.

Appendix

Derivation of the Maximum Caliber solution

To maximize the function in Eq. (4), we set its derivative with respect to $P(\Gamma)$ equal to zero:

$$\frac{\partial \mathcal{L}}{\partial P(\Gamma)} = -\log P(\Gamma) - 1 + \sum_{ij} \eta_{ij} \frac{1}{T} \sum_t x_i^{(t+1)} x_j^{(t)} + \sum_i \lambda_i \frac{1}{T} \sum_t x_i^{(t)} + \mu = 0. \tag{8}$$

Solving for $P(\Gamma)$ gives

$$P(\Gamma) = \exp\left(\frac{1}{T} \sum_t \left(\sum_{ij} \eta_{ij}\, x_i^{(t+1)} x_j^{(t)} + \sum_i \lambda_i\, x_i^{(t)} \right) + \mu - 1 \right). \tag{9}$$

The parameter μ can be fixed by the normalization of probability, yielding:

$$\mu = 1 - \log\left(\sum_\Gamma \exp\left(\frac{1}{T} \sum_t \left(\sum_{ij} \eta_{ij}\, x_i^{(t+1)} x_j^{(t)} + \sum_i \lambda_i\, x_i^{(t)} \right) \right) \right). \tag{10}$$

Recall that $P(\Gamma)$ is the probability of a trajectory, $\Gamma = \{\mathbf{x}^{(1)} \cdots \mathbf{x}^{(T)}\}$. The form of Eq. (9) implies that the process is Markovian (Ge et al. 2012), i.e., $P\big(\mathbf{x}^{(t+1)} \big| \mathbf{x}^{(1)} \cdots \mathbf{x}^{(t)}\big) = P\big(\mathbf{x}^{(t+1)} \big| \mathbf{x}^{(t)}\big)$ (see detailed proof in Supplementary Material). Therefore, we can rewrite the probability of a trajectory as a product of conditional probabilities:

$$P(\Gamma) \equiv P\big(\mathbf{x}^{(1)} \cdots \mathbf{x}^{(T)}\big) = P\big(\mathbf{x}^{(1)}\big) \prod_{t=1}^{T-1} P\big(\mathbf{x}^{(t+1)} \big| \mathbf{x}^{(t)}\big). \tag{11}$$

It can be shown that the conditional probability for each time step is given by:

$$P\big(\mathbf{x}^{(t+1)} \big| \mathbf{x}^{(t)}\big) = \frac{1}{Z_t} \exp\left(\sum_{ij} \tilde{\eta}_{ij}\, x_i^{(t+1)} x_j^{(t)} + \sum_i \tilde{\lambda}_i\, x_i^{(t+1)} \right), \tag{12}$$

where $\tilde{\eta}_{ij}$ and $\tilde{\lambda}_i$ are parameters defined in terms of the original parameters η_{ij} and λ_i (see Supplementary Material), and Z_t is a normalization factor. We can further write Eq. (12) as a product of conditional probabilities for each species, i.e., $P\big(\mathbf{x}^{(t+1)} \big| \mathbf{x}^{(t)}\big) = \prod_i P\big(x_i^{(t+1)} \big| \mathbf{x}^{(t)}\big)$ with

$$P\big(x_i^{(t+1)} \big| \mathbf{x}^{(t)}\big) = \frac{1}{Z_{t,i}} \exp\left(x_i^{(t+1)} \left(\sum_j \tilde{\eta}_{ij}\, x_j^{(t)} + \tilde{\lambda}_i \right) \right). \tag{13}$$

Fixing the normalization factor $Z_{t,i}$ (and redefining the parameters by dropping the tilde) leads to the result in Eq. (5).

Choice of parameter values

In the production of data used to produce Fig.3 the η_{ij} on the main diagonal (green in Fig. 2A) are taken from a uniform distribution from 0 to 0.5. The diagonal of each on diagonal block is made identical so that the baseline persistence parameters are only species specific. This ensures that a species is more likely to be present at the next time step, if it is present in the current time step. The η_{ij} in the main diagonal blocks but away from the main diagonal (blue in Fig. 2A) are taken from a normal distribution with mean 0 and a standard deviation of .5. To assign the values of η_{ij} that are associated with migration (red in Fig. 2A), first, each pair of patches is either connected or not, with a probability of $1/M$. For each pair of patches that are connected, the diagonal of the block associated with that pair is filled with parameters chosen from a uniform distribution from 0 to 1. The λ_i are assigned per-species taken from a normal distribution with mean 0 and standard deviation of 1. Additionally , to include environmental variation between patches, an additional random factor is added to each λ_i that is taken from a normal distribution centered at 0 with standard deviation of .2. The extra factor means that the same species in different patches have similar but not identical interactions with the abiotic factors in their environment.

In the production of data used to produce Fig. 4, the parameters are chosen in accordance with each systems associated motif. Each system has 3 species and 3 patches and the patches are all connected. For the random interaction motif (red in Fig. 2B) the parameters are chosen in the same manner as above except without the environmental variation added to the λ_i. In the other motifs, the η_{ij} and the λ_i are set to 1. For the apparent competition motif (yellow in Fig. 2B) interspecific interaction parameters (blue in Fig. 2A), a strength of interaction s is chosen and then in each on diagonal block, $\eta_{02} = \eta_{12} = -s$, $\eta_{20} = s$, and $\eta_{21} = s/2$. For the intraguild competition motif (green in Fig. 2B) three strengths are chosen a, b, and c and then, in each diagonal block $\eta_{20} = -\eta_{02} = a$, $\eta_{21} = -\eta_{12} = b$, and $\eta_{10} = -\eta_{01} = c$. For the non-transitive competition motif (purple in Fig. 2B) a strength s is chosen and then in each diagonal block $\eta_{01} = \eta_{12} = \eta_{20}$=-s. All of theses parameter choices ensure that each predator has a negative relationship with the presence of its prey and each prey has a positive effect on the presence of its predator.

References

Cayco-Gajic, N. Alex, Joel Zylberberg, and Eric Shea-Brown (2018). "A Moment-Based Maximum Entropy Model for Fitting Higher-Order Interactions in Neural Data". *Entropy* 20.7, p. 489. ISSN: 1099-4300. DOI: 10.3390/e20070489.

Cox, D. R. and E. J. Snell (1989). *Analysis of Binary Data*. 2nd. CRC Press. ISBN: 978-0-412-30620-4.

Darwin, Charles (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

DeAngelis, D. L. and J. C. Waterhouse (1987). "Equilibrium and Nonequilibrium Concepts in Ecological Models". *Ecological Monographs* 57.1, pp. 1–21. ISSN: 1557-7015. DOI: 10.2307/1942636.

Firman, Taylor, Gábor Balázsi, and Kingshuk Ghosh (2017). "Building Predictive Models of Genetic Circuits Using the Principle of Maximum Caliber". *Biophysical Journal* 113.9, pp. 2121–2130. ISSN: 00063495. DOI: 10.1016/j.bpj.2017.08.057.

Ge, Hao et al. (2012). "Markov processes follow from the principle of maximum caliber". *The Journal of Chemical Physics* 136.6, p. 064108. ISSN: 0021-9606. DOI: 10.1063/1.3681941.

Ghosh, Kingshuk et al. (2020). "The Maximum Caliber Variational Principle for Nonequilibria". *Annual Review of Physical Chemistry* 71, pp. 213–238. ISSN: 0066-426X, 1545-1593. DOI: 10.1146/annurev-physchem-071119-040206.

Goudard, Alexandra and Michel Loreau (2008). "Nontrophic Interactions, Biodiversity, and Ecosystem Functioning: An Interaction Web Model." *The American Naturalist* 171.1, pp. 91–106. ISSN: 0003-0147. DOI: 10.1086/523945.

GrandPre, Trevor, Gianluca Teza, and William Bialek (2024). *Direct estimates of irreversibility from time series*. DOI: 10.48550/arXiv.2412.19772. URL: http://arxiv.org/abs/2412.19772.

Gravel, Dominique, François Massol, and Mathew A. Leibold (2016). "Stability and complexity in model meta-ecosystems". *Nature Communications* 7.1, p. 12457. ISSN: 2041-1723. DOI: 10.1038/ncomms12457.

Hanski, Ilkka (1983). "Coexistence of Competitors in Patchy Environment". *Ecology* 64.3, pp. 493–500. ISSN: 1939-9170. DOI: 10.2307/1939969.

Harte, John and Erica A. Newman (2014). "Maximum information entropy: a foundation for ecological theory". *Trends in Ecology & Evolution* 29.7, pp. 384–389. ISSN: 01695347. DOI: 10.1016/j.tree.2014.04.009.

Hartig, Florian et al. (2024). "Novel community data in ecology-properties and prospects". *Trends in Ecology & Evolution* 39.3, pp. 280–293. ISSN: 0169-5347. DOI: 10.1016/j.tree.2023.09.017.

Hastings, Alan (1980). "Disturbance, coexistence, history, and competition for space". *Theoretical Population Biology* 18.3, pp. 363–373. ISSN: 00405809. DOI: 10.1016/0040-5809(80)90059-3.

Hastings, Alan et al. (1993). "Chaos in Ecology: Is Mother Nature a Strange Attractor?*". *Annual Review of Ecology, Evolution, and Systematics* 24.Volume 24, 1993, pp. 1–33. ISSN: 1543-592X, 1545-2069. DOI: 10.1146/annurev.es.24.110193.000245.

Holt, Robert D. (1977). "Predation, apparent competition, and the structure of prey communities". *Theoretical Population Biology* 12.2, pp. 197–229. ISSN: 00405809. DOI: 10.1016/0040-5809(77)90042-9.

— (1997). "From Metapopulation Dynamics to Community Structure: Some Consequences of Spatial Heterogeneity". *Metapopulation Biology*. Ed. by Ilkka Hanski and Michael E. Gilpin. San Diego: Academic Press, pp. 149–164. ISBN: 978-0-12-323445-2. DOI: 10.1016/B978-012323445-2/50010-9.

Holt, Robert D. and Gary A. Polis (1997). "A Theoretical Framework for Intraguild Predation". *The American Naturalist* 149.4, pp. 745–764. ISSN: 0003-0147. DOI: 10.1086/286018.

Horn, Henry S. and Robert H. MacArthur (1972). "Competition among Fugitive Species in a Harlequin Environment". *Ecology* 53.4, pp. 749–752. ISSN: 1939-9170. DOI: 10.2307/1934797.

Jaynes, E. T. (1957). "Information Theory and Statistical Mechanics". *Physical Review* 106.4, pp. 620–630. DOI: 10.1103/PhysRev.106.620.

— (1980). "The Minimum Entropy Production Principle". *Annual Review of Physical Chemistry* 31.Volume 31, pp. 579–601. ISSN: 0066-426X, 1545-1593. DOI: 10.1146/annurev.pc.31.100180.003051.

Kéfi, Sonia et al. (2016). "How Structured Is the Entangled Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased Persistence and Resilience". *PLOS Biology* 14.8, e1002527. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1002527.

Leibold, Mathew A., Lynn Govaert, et al. (2022). "Evolution and Community Assembly Across Spatial Scales". *Annual Review of Ecology, Evolution, and Systematics* 53, pp. 299–326. ISSN: 1543-592X, 1545-2069. DOI: 10.1146/annurev-ecolsys-102220-024934.

Leibold, Mathew A. and Nicolas Loeuille (2015). "Species sorting and patch dynamics in harlequin metacommunities affect the relative importance of environment and space". *Ecology* 96.12, pp. 3227–3233. ISSN: 1939-9170. DOI: 10.1890/14-2354.1.

Leibold, Mathew A., F. Javiera Rudolph, et al. (2022). "The internal structure of metacommunities". *Oikos* 2022.1. ISSN: 1600-0706. DOI: 10.1111/oik.08618.

Levin, Simon A. (1974). "Dispersion and Population Interactions". *The American Naturalist* 108.960, pp. 207–228. ISSN: 0003-0147. DOI: 10.1086/282900.

Levins, Richard and David Culver (1971). "Regional Coexistence of Species and Competition between Rare Species". *Proceedings of the National Academy of Sciences* 68.6, pp. 1246–1248. DOI: 10.1073/pnas.68.6.1246.

Li, Junang et al. (2024). *Quantifying irreversibility of ecological systems*. DOI: 10.1101/2024.08.04.606544. URL: https://www.biorxiv.org/content/10.1101/2024.08.04.606544v1.

Lotka, Alfred J. (1925). "Elements of Physical Biology". *Nature* 116.2917, pp. 461–461. ISSN: 1476-4687. DOI: 10.1038/116461b0.

May, Robert M. (1972). "Will a Large Complex System be Stable?" *Nature* 238.5364, pp. 413–414. ISSN: 1476-4687. DOI: 10.1038/238413a0.

May, Robert M. and Warren J. Leonard (1975). "Nonlinear Aspects of Competition Between Three Species". *SIAM Journal on Applied Mathematics* 29.2, pp. 243–253. ISSN: 0036-1399. DOI: 10.1137/0129022.

McCann, Kevin Shear (2000). "The diversity–stability debate". *Nature* 405.6783, pp. 228–233. ISSN: 1476-4687. DOI: 10.1038/35012234.

Mickalide, Harry and Seppe Kuehn (2019). "Higher-Order Interaction between Species Inhibits Bacterial Invasion of a Phototroph-Predator Microbial Community". *Cell Systems* 9.6, 521–533.e10. ISSN: 24054712. DOI: 10.1016/j.cels.2019.11.004.

Mittlböck, Martina and Michael Schemper (1996). "Explained Variation for Logistic Regression". *Statistics in Medicine* 15.19, pp. 1987–1997. ISSN: 1097-0258. DOI: 10.1002/(SICI)1097-0258(19961015)15:19<1987::AID-SIM318>3.0.CO;2-9.

Mora, Thierry, Stéphane Deny, and Olivier Marre (2015). "Dynamical Criticality in the Collective Activity of a Population of Retinal Neurons". *Physical Review Letters* 114.7, p. 078105. DOI: 10.1103/PhysRevLett.114.078105.

Nagelkerke, N. J. D. (1991). "A note on a general definition of the coefficient of determination". *Biometrika* 78.3, pp. 691–692. ISSN: 0006-3444. DOI: 10.1093/biomet/78.3.691.

Novak, Mark et al. (2016). "Characterizing Species Interactions to Understand Press Perturbations: What Is the Community Matrix?" *Annual Review of Ecology, Evolution, and Systematics* 47, pp. 409–432. ISSN: 1543-592X, 1545-2069. DOI: 10.1146/annurev-ecolsys-032416-010215.

Ohgushi, Takayuki, Oswald J. Schmitz, and Robert D. Holt (2012). *Trait-mediated indirect interactions: ecological and evolutionary perspectives*. Ecological reviews. New York: Cambridge University Press. ISBN: 978-1-139-85371-2.

Petchey, Owen L. and Kevin J. Gaston (2006). "Functional diversity: back to basics and looking forward". *Ecology Letters* 9.6, pp. 741–758. ISSN: 1461-0248. DOI: 10.1111/j.1461-0248.2006.00924.x.

Polis, Gary A., Wendy B. Anderson, and Robert D. Holt (1997). "Toward an Integration of Landscape and Food Web Ecology: The Dynamics of Spatially Subsidized Food Webs". *Annual Review of Ecology and Systematics* 28.1, pp. 289–316. ISSN: 0066-4162. DOI: 10.1146/annurev.ecolsys.28.1.289.

Pressé, Steve et al. (2013). "Principles of maximum entropy and maximum caliber in statistical physics". *Reviews of Modern Physics* 85.3, pp. 1115–1141. DOI: 10.1103/RevModPhys.85.1115.

Shipley, Bill (2010). "Inferential permutation tests for maximum entropy models in ecology". *Ecology* 91.9, pp. 2794–2805. ISSN: 1939-9170. DOI: 10.1890/09-1255.1.

Sinervo, B. and C. M. Lively (1996). "The rock–paper–scissors game and the evolution of alternative male strategies". *Nature* 380.6571, pp. 240–243. ISSN: 1476-4687. DOI: 10.1038/380240a0.

Stock, Gerhard, Kingshuk Ghosh, and Ken A. Dill (2008). "Maximum Caliber: A variational approach applied to two-state dynamics". *The Journal of Chemical Physics* 128.19, p. 194102. ISSN: 0021-9606. DOI: 10.1063/1.2918345.

Tank, Alex et al. (2021). "The Convex Mixture Distribution: Granger Causality for Categorical Time Series". *SIAM Journal on Mathematics of Data Science* 3.1, pp. 83–112. DOI: 10.1137/20M133097X.

Vellend, Mark (2010). "Conceptual Synthesis in Community Ecology". *The Quarterly Review of Biology* 85.2, pp. 183–206. ISSN: 0033-5770. DOI: 10.1086/652373.

— (2016). *The Theory of Ecological Communities*. Princeton University Press. ISBN: 978-1-4008-8379-0. DOI: 10.1515/9781400883790.

Volterra, Vito (1928). "Variations and Fluctuations of the Number of Individuals in Animal Species living together". *Journal du Conseil* 3.1, pp. 3–51. ISSN: 0020-6466. DOI: 10.1093/icesjms/3.1.3.

Wan, Hongbin, Guangfeng Zhou, and Vincent A. Voelz (2016). "A Maximum-Caliber Approach to Predicting Perturbed Folding Kinetics Due to Mutations". *Journal of Chemical Theory and Computation* 12.12, pp. 5768–5776. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.6b00938.

Wang, Yue and Hong Qian (2020). "Mathematical Representation of Clausius' and Kelvin's Statements of the Second Law and Irreversibility". *Journal of Statistical Physics* 179.3, pp. 808–837. ISSN: 1572-9613. DOI: 10.1007/s10955-020-02556-6.

Werner, Earl E. and Scott D. Peacor (2003). "A Review of Trait-Mediated Indirect Interactions in Ecological Communities". *Ecology* 84.5, pp. 1083–1100. ISSN: 1939-9170. DOI: 10.1890/0012-9658(2003)084[1083:AROTII]2.0.CO;2.

Wu, Jianguo and Orie L. Loucks (1995). "From Balance of Nature to Hierarchical Patch Dynamics: A Paradigm Shift in Ecology". *The Quarterly Review of Biology* 70.4, pp. 439–466. ISSN: 0033-5770.

Supplementary Material

Detailed derivation of the Maximum Caliber model

For simplicity and clarity, we will derive the Maximum Caliber solution for a single species, which means dropping the subscript i. We start from the loss function:

$$L = -\sum_{\{x^{(t)}\}} P(\{x^{(t)}\}) \log P(\{x^{(t)}\}) + \eta \left(\sum_{\{x^{(t)}\}} P(\{x^{(t)}\}) \frac{1}{T} \sum_{t=1}^{T-1} x^{(t+1)} x^{(t)} - c \right)$$
$$+ \lambda \left(\sum_{\{x^{(t)}\}} P(\{x^{(t)}\}) \frac{1}{T} \sum_{t=1}^{T-1} x^{(t)} - f \right) + \mu \left(\sum_{\{x^{(t)}\}} P(\{x^{(t)}\}) - 1 \right) \tag{S1}$$

maximizing with respect to $P(\{x^{(t)}\})$ gives

$$\frac{\partial L}{\partial P(\{x^{(t)}\})} = -\log P(\{x^{(t)}\}) - 1 + \frac{\eta}{T} \sum_{t=1}^{T-1} x^{(t+1)} x^{(t)} + \frac{\lambda}{T} \sum_{t=1}^{T-1} x^{(t)} + \mu = 0 \tag{S2}$$

if we renormalize our parameters with $\eta \leftarrow \eta/T$, $\lambda \leftarrow \lambda/T$, and $Z \leftarrow e^{1-\mu}$ and solve for $P(\{x^{(t)}\})$,

$$P(\{x^{(t)}\}) = \frac{1}{Z} \exp\left(\eta \sum_{t=1}^{T-1} x^{(t+1)} x^{(t)} + \lambda \sum_{t=1}^{T-1} x^{(t)} \right) = \frac{1}{Z} \prod_{t=1}^{T-1} \exp\left(\eta x^{(t+1)} x^{(t)} + \lambda x^{(t)} \right) \tag{S3}$$

To see why this is a Markovian process, we will calculate the conditional probability starting from the end of the time series and working backwards. The joint probability of the full trajectory from $t = 1$ to T is given by a product of terms for each time step:

$$P\left(x^{(1)} \cdots x^{(T)}\right) = \frac{1}{Z} \exp\left(\eta x^T x^{T-1} + \lambda x^{T-1}\right) \exp\left(\eta x^{T-1} x^{T-2} + \lambda x^{T-2}\right) \cdots \exp\left(\eta x^1 x^0 + \lambda x^0\right) \tag{S4}$$

We then calculate the joint probability of a partial trajectory from $t = 1$ to $T - 1$ by summing over the possible states of $x^{(T)}$

$$P\left(x^{(1)} \cdots x^{(T-1)}\right) = \sum_{x^{(T)}=0}^{1} P\left(x^{(1)} \cdots x^{(T)}\right)$$
$$= \frac{1}{Z} \left(1 + \exp\left(\eta x^{T-1}\right)\right) \exp\left(\lambda x^{T-1}\right) \exp\left(\eta x^{T-1} x^{T-2} + \lambda x^{T-2}\right) \cdots \tag{S5}$$

We can use these to calculate the conditional probability of $x^{(T)}$ given the trajectory up to $T - 1$.

$$P\left(x^{(T)} | x^{(1)} \cdots x^{(T-1)}\right) = \frac{P\left(x^{(1)} \cdots x^{(T)}\right)}{P\left(x^{(1)} \cdots x^{(T-1)}\right)} = \frac{\exp(\eta x^{(T)} x^{(T-1)})}{1 + \exp(\eta x^{(T-1)})} \tag{S6}$$

From this expression we see that the conditional probability depends only on $x^{(T-1)}$, which means this step is Markovian. We can rewrite this conditional probability as $P\left(x^{(T)} | x^{(T-1)}\right)$ and, for

$x^{(T)} = 1$,:

$$P\left(x^{(T)}=1|x^{(T-1)}\right) = \frac{\exp(\eta x^{(T-1)})}{1+\exp(\eta x^{(T-1)})} = \frac{1}{1+\exp(-\eta x^{(T-1)})} \tag{S7}$$

Through a similar process, we can calculate the joint probability of the trajectory from $t = 1$ to $T - 2$:

$$P\left(x^{(T)}|x^{(1)}\cdots x^{(T-1)}\right) = ... \tag{S8}$$

Then we calculate the conditional probability of $x^{(T-1)}$ given the trajectory up to $t = T - 2$:

$$\begin{aligned} P\left(x^{(T-1)}|x^{(1)}\cdots x^{(T-2)}\right) &= \frac{P\left(x^{(1)}\cdots x^{(T-1)}\right)}{P\left(x^{(1)}\cdots x^{(T-2)}\right)} \\ &= \frac{\left(1+\exp(\eta x^{(T-1)})\right)\exp(\lambda x^{(T-1)} + \eta x^{(T-1)} x^{(T-2)})}{2+(1+\exp(\eta))\exp(\lambda + \eta x^{(T-2)})} \end{aligned} \tag{S9}$$

Again we see that this conditional probability depends only on $x^{(T-2)}$, which means it is Markovian. This conditional probability can be simplified as

$$P\left(x^{(T-1)}=1|x^{(1)}\cdots x^{(T-2)}\right) = \frac{1}{1+\exp(-\eta x^{(T-2)} - \lambda')} \tag{S10}$$

where $\lambda' = \lambda + \log(1 + \exp(\eta)/2)$. By continuing this recursion we obtain

$$P\left(x^{(T-2)} = 1|x^{(T-3)}\right) = \frac{1}{1+\exp(-\eta x^{T-3} - \lambda'')} \tag{S11}$$

with $\lambda'' = \lambda + \log\left(\frac{1+\exp(\eta+\lambda')}{1+\exp(\lambda')}\right)$. As we move further away from the end of the time series, the effective parameter value of λ quickly approaches a constant. To find this limiting value, $\tilde{\lambda}$, we can set $\lambda' = \lambda''$ in the recursive relation:

$$\exp(\tilde{\lambda}) = \exp(\lambda)\frac{1+\exp(\eta+\tilde{\lambda})}{1+\exp(\tilde{\lambda})} \tag{S12}$$

Solving this for $\tilde{\lambda}$ gives

$$\exp\left(\tilde{\lambda}\right) = \frac{1}{2}\exp(\eta+\lambda) - 1 + \sqrt{(1-\exp(\eta+\lambda))^2 + 4exp(\lambda)} \tag{S13}$$

Therefore, for large T, the conditional probability at each time step can be taken as:

$$P\left(x^{(t+1)} = 1|x^{(t)}\right) = \frac{1}{1+\exp(-\eta x^{(t)} - \tilde{\lambda})} \tag{S14}$$

Note that when calculating the average occupancy of a species f_i to use in the Maximum Caliber calculation, there is a choice of whether to sum from $x^{(1)}$ to $x^{(T-1)}$, or from $x^{(2)}$ to $x^{(T)}$. More

generally, we can use

$$f_i = \frac{1}{T}\sum_{t=1}^{T-1} \alpha x_i^{(t)} + (1-\alpha)x_i^{(t+1)} \tag{S15}$$

for any value of α between $0 \leq \alpha \leq 1$. Our choice above corresponds to $\alpha = 1$. Following a similar scheme for $\alpha = 0$, we find that λ converges to the same value. As such, for long time series (large T), the resulting model will be the same regardless of the value of α used.

Maximum Caliber versus Logistic Regression

Here we give a comparison between the parameter inference in Maximum Caliber and in logistic regression. In Maximum Caliber, strictly speaking, we need to find parameter values by solving the constraint equations:

$$f_i^{\text{obs}} = \sum_{\{x_i^{(t)}\}} P(\{x_i^{(t)}\})\frac{1}{T}\sum_t x_i^{(t+1)} \tag{S16}$$

$$= \frac{1}{T}\sum_t \sum_{x_i^{(t+1)}} P(x_i^{(t+1)})x_i^{(t+1)} \tag{S17}$$

$$= \frac{1}{T}\sum_t \sum_{x_i^{(t+1)}} \sum_{\mathbf{x}^{(t)}} P(x_i^{(t+1)}|\mathbf{x}^{(t)})P(\mathbf{x}^{(t)})x_i^{(t+1)} \tag{S18}$$

If we assume that the process is stationary, then each t term would give the same result, which means:

$$f_i^{\text{obs}} = \sum_{x_i^{(t+1)}} \sum_{\mathbf{x}^{(t)}} P(x_i^{(t+1)}|\mathbf{x}^{(t)})P(\mathbf{x}^{(t)})x_i^{(t+1)} \tag{S19}$$

$$= \sum_{\mathbf{x}^{(t)}} P(\mathbf{x}^{(t)})P(x_i^{(t+1)}=1|\mathbf{x}^{(t)}) \tag{S20}$$

Similarly for c_{ij}:

$$c_{ij}^{\text{obs}} = \sum_{\{x_i^{(t)}\}} P(\{x_i^{(t)}\}) \frac{1}{T} \sum_t x_i^{(t+1)} x_j^{(t)} \tag{S21}$$

$$= \frac{1}{T} \sum_t \sum_{x_i^{(t+1)}} \sum_{\mathbf{x}^{(t)}} P(x_i^{(t+1)}, x_j^{(t)}) x_i^{(t+1)} x_j^{(t)} \tag{S22}$$

$$= \frac{1}{T} \sum_t \sum_{x_i^{(t+1)}} \sum_{\mathbf{x}^{(t)}} P(x_i^{(t+1)} | \mathbf{x}^{(t)}) P(\mathbf{x}^{(t)}) x_i^{(t+1)} x_j^{(t)} \tag{S23}$$

$$= \sum_{x_i^{(t+1)}} \sum_{\mathbf{x}^{(t)}} P(x_i^{(t+1)} | \mathbf{x}^{(t)}) P(\mathbf{x}^{(t)}) x_i^{(t+1)} x_j^{(t)} \tag{S24}$$

$$= \sum_{\mathbf{x}^{(t)}} P(x_i^{(t+1)}{=}1 | \mathbf{x}^{(t)}) P(\mathbf{x}^{(t)}) x_j^{(t)} \tag{S25}$$

$$= \sum_{\mathbf{x}^{(t)}} P(\mathbf{x}^{(t)}) P(x_i^{(t+1)}{=}1 | \mathbf{x}^{(t)}) x_j^{(t)} \tag{S26}$$

In logistic regression, we minimize the cross entropy between the predicted probabilities P_t and the realized probabilities y_t.

$$L = -\sum_t y_t \log P_t + (1 - y_t) \log(1 - P_t) \tag{S27}$$

where y_t is the presence or absence of a species in a patch at the next time step $x_i^{(t+1)}$, and P_t is the probability of that species being present in that patch given the configuration of the system at the previous time step.

$$P_t = \prod_i P(x_i^{(t+1)}{=}1 | \mathbf{x}^{(t)}) = \prod_i \frac{1}{1 + e^{-\sum_j \eta_{ij} x_j^{(t)} - \lambda_i}} \tag{S28}$$

With this expression, the log loss becomes:

$$L = -\sum_t \sum_i \left(x_i^{(t+1)} \log P(x_i^{(t+1)} | \mathbf{x}^{(t)}) + (1 - x_i^{(t+1)}) \log(1 - P(x_i^{(t+1)} | \mathbf{x}^{(t)})) \right) \tag{S29}$$

Differentiating the loss with respect to λ_i gives

$$\frac{\partial L}{\partial \lambda_i} = \sum_t x_i^{(t+1)} \left(1 - P\left(x_i^{(t+1)} | \mathbf{x}^{(t)} \right) \right) - \left(1 - x_i^{(t+1)} P\left(x_i^{(t+1)} | \mathbf{x}^{(t)} \right) \right) \tag{S30}$$

$$= -\sum_t x_t^{(t+1)} - P\left(x_i^{(t+1)}{=}1 | \mathbf{x}^{(t)} \right) \tag{S31}$$

$$= -T \left(f_i^{\text{obs}} - \frac{1}{T} \sum_t P\left(x_i^{(t+1)}{=}1 | \mathbf{x}^{(t)} \right) \right) \tag{S32}$$

Differentiating with respect to η_{ij} gives

$$\frac{\partial L}{\partial \eta_{ij}} = -\sum_t x_i^{(t+1)} x_j^{(t)} (1 - P\left(x_i^{(t+1)} | \mathbf{x}^{(t)}\right)) - \left(1 - x_i^{(t+1)}\right) x_j^{(t)} P\left(x_i^{(t+1)} | \mathbf{x}^{(t)}\right) \tag{S33}$$

$$= -\sum_t \left(x_i^{(t+1)} - P\left(x_i^{(t+1)}{=}1 | \mathbf{x}^{(t)}\right)\right) x_j^{(t)} \tag{S34}$$

$$= -T \left(c_{ij}^{\text{obs}} - \frac{1}{T} \sum_t P\left(x_i^{(t+1)}{=}1 | \mathbf{x}^{(t)}\right) x_j^{(t)}\right) \tag{S35}$$

After maximizing the log likelihood, we should get

$$f_i^{\text{obs}} = \frac{1}{T} \sum_t P\left(x_i^{(t+1)}{=}1 | \mathbf{x}^{(t)}\right) \tag{S36}$$

$$c_{ij}^{\text{obs}} = \frac{1}{T} \sum_t P\left(x_i^{(t+1)}{=}1 | \mathbf{x}^{(t)}\right) x_j^{(t)} \tag{S37}$$

Note that in these equations the $x_i^{(t)}$'s are the observed values of species presence or absence, rather than dummy indices that are summed over in Eqs. (S20) & (S26).

If we compare Eqs. (S36) & (S37) to Eqs. (S20) & (S26), the difference between the logistic regression calculation and the strict Maximum Caliber calculation is that logistic regression takes an average over time of observed data, whereas Maximum Caliber takes an average over the probability of system trajectories. When the time series are sufficiently long, the two calculations will give equal results.

Supplementary Figures

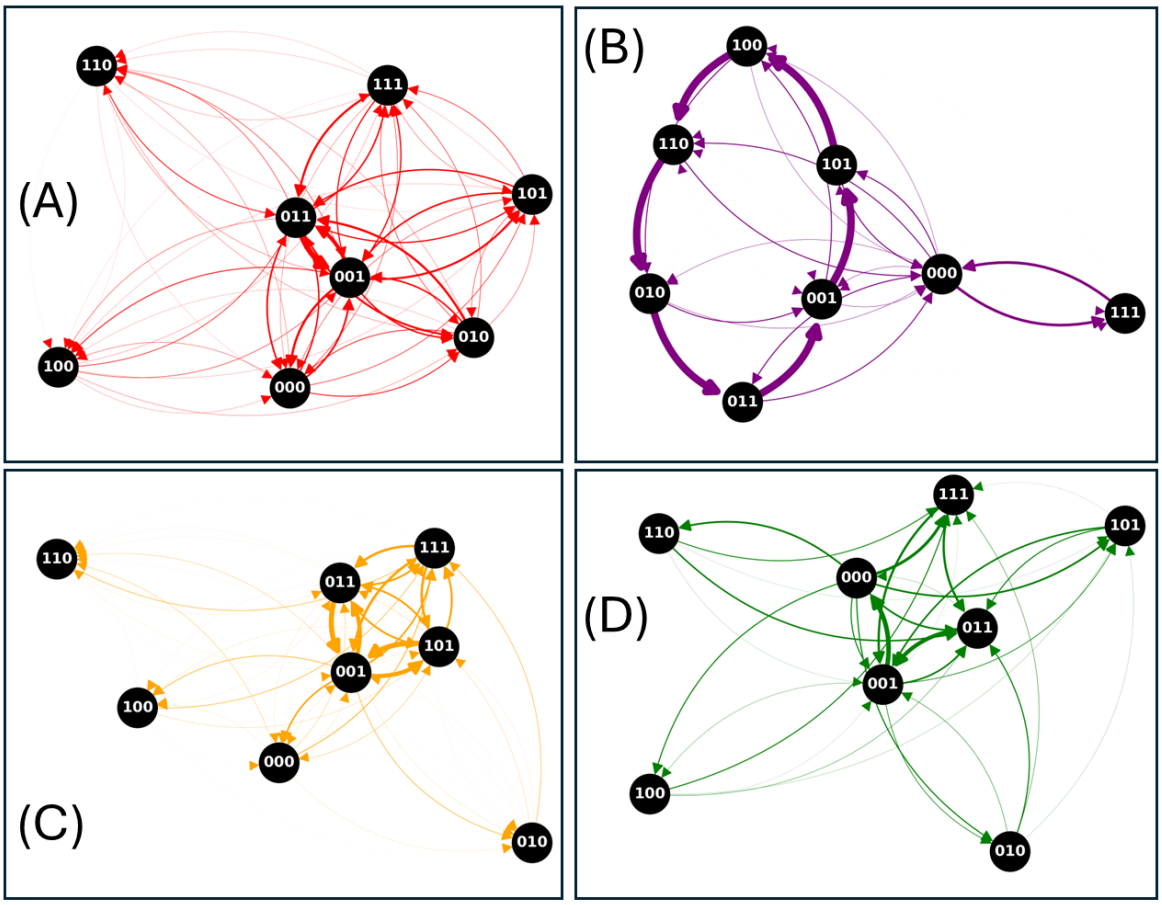

Figure S1: Graph representations of the state transitions in the four motifs studied in the main text: (A) random interactions, (B) non-transitive competition, (C) apparent competition, and (D) intraguild competition. Each of the systems is a single patch with three species. Each node represents a state of the system labeled by a binary vector that denotes whether each species is present. The thickness of an arrow between two nodes corresponds to the rate of transition between the states, calculated from simulations of over 10000 time steps. The thickness is relatively homogeneous in the random interactions motif, showing that the system is near equilibrium. In contrast, there is a dominant cycle in the non-transitive competition motif, showing that the system is far from equilibrium.

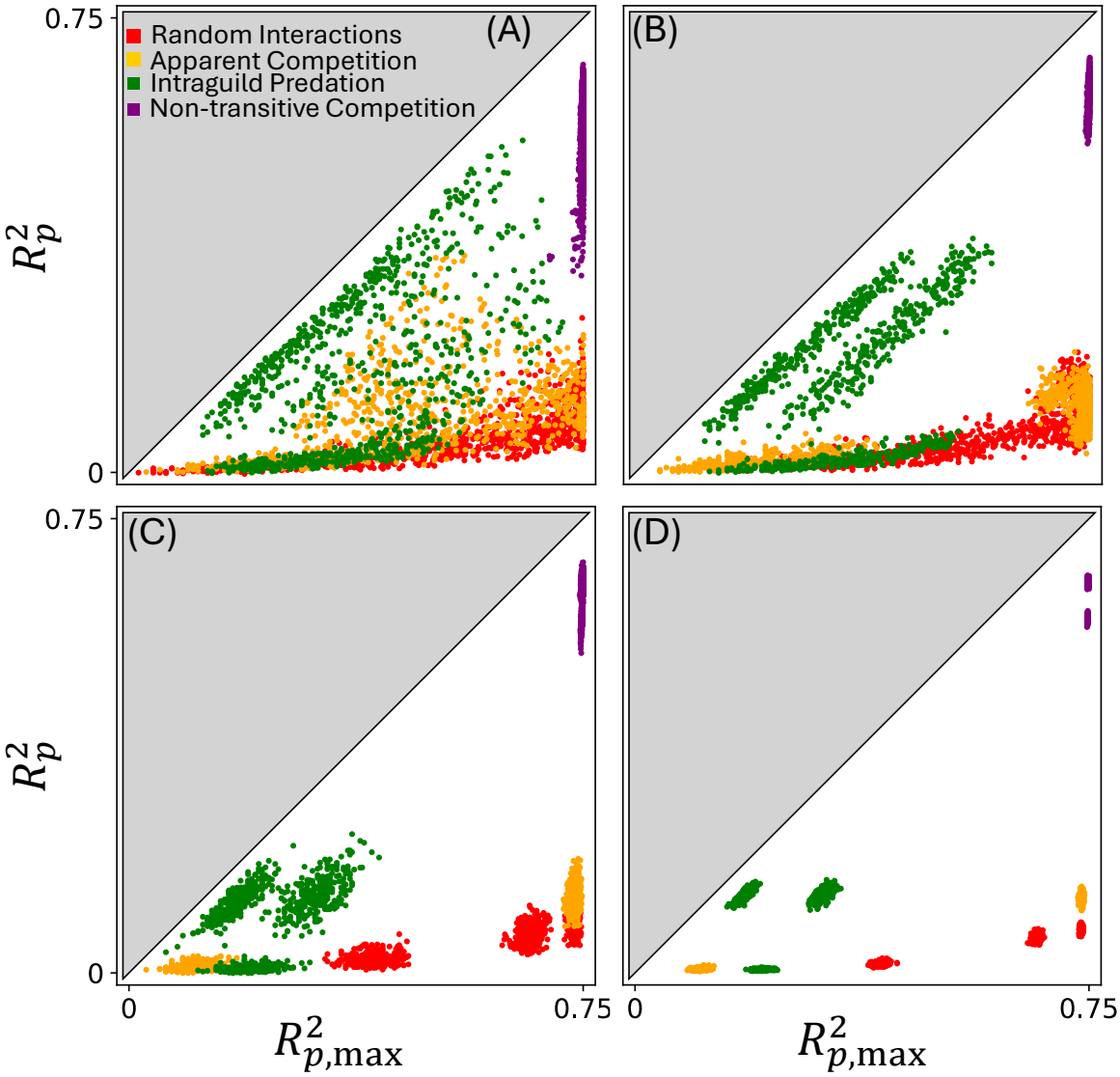

Figure S2: A plot of R_{p}^2 versus $R_{\mathrm{p,max}}^2$ for each of the system motifs from the main text. All points are in the lower triangle because $R_{\mathrm{p}}^2 \leq R_{\mathrm{p,max}}^2$, and for logistic regression we have $0 \leq R_{\mathrm{p,max}}^2 \leq 0.75$. Each point represents a species-patch combination in a particular simulation. The differences in the spread of the points illustrate the different sources of variability in parameter inference. Panels (A)-(C) are produced from parameter inferences from time series of length $T = 1000$. (A) shows a collection of inferences from sets of data, each of which was produced from a new randomization of η_{ij} and λ_i values. (B) shows the results from simulations in which the diagonal blocks of η_{ij} are all identical, but the migration parameters are randomized. (C) shows the results from simulations in which all of the η_{ij} and λ_i values are identical between simulations, but each point is inferred from a different realization of time series data. (D) is similar to (C) except that the time series used for inference are of length $T = 10000$.