Linking Pattern to Process in Metacommunities: Challenges and Opportunities

Mathew A. Leibold, Department of Biology, University of Florida, Gainesville, Florida 32661, USA and Institut Natura e Teoria en Pireneus, Surba, France. mleibold@ufl.edu

Matthieu Barbier, CIRAD, UMR PHIM, 34090 Montpellier, France. PHIM Plant Health Institute, University of Montpellier, CIRAD, INRAE, Institut Agro, IRD, 34090 Montpellier, France. ORCID: 0000-0002-0669-8927. contact@mrcbarbier.org

Leonora S. Bittleston, Department of Biological Sciences, Boise State University, Boise, Idaho 83725. ORCID: 0000-0003-4007-5405. leonorabittleston@boisestate.edu

Adam T. Clark, Department of Biology, University of Graz, Holteigasse 6, 8010, Graz, Austria. ORCID: 0000-0002-8843-3278. adam.clark@uni-graz.at

Catalina Cuellar-Gempeler, Department of Biological Sciences, California Polytechnic University, Humboldt. 1st Harpst Street, Arcata, CA. ORCID: 0000-0002-2634-9599. ccg@humboldt.edu

Rafael D'Andrea, Department of Ecology and Evolution, Stony Brook University, Stony Brook, 11790, USA. rafael.dandrea@stonybrook.edu

Veronica F. Frans, Hopkins Marine Station, Stanford University, Pacific Grove, CA, USA. verofrans@gmail.com ORCID: 0000-0002-5634-3956

Gabriel Khattar, Department of Biology, Concordia University, Montreal, Quebec, Canada. 0000-0002-1703-4210. gabriel.khattar@concordia.ca

Zachary Miller, Department of Earth and Planetary Sciences, Yale University, New Haven CT 06511, USA. ORCID: 0000-0001-9762-8301. zachary.miller@yale.edu

Pedro Peres-Neto, Department of Biology, Concordia University, Montreal, Quebec, Canada. ORCID: 0000-0002-5629-8067. pedro.peres-neto@concordia.ca

Nathan I. Wisnoski, Department of Biological Sciences, Mississippi State University, Mississippi State, MS 39762, USA. nathan.wisnoski@msstate.edu

Abstract

Ecological communities, and especially metacommunities, are complex and dynamic entities. Resolving the processes and mechanisms that shape these systems remains a central challenge in ecology. This challenge is compounded by the increasing entanglement of mechanisms, processes, and emergent patterns of biodiversity as scales of space, time, and biological organization expand. Here, we define and contextualize key issues, recent progress, and remaining challenges in interpreting basic metacommunity data and using predictive models to link processes to patterns.

We find substantial progress in connecting pattern and process through improved data repeatability and scaling, enhanced analytical tools to quantify patterns, and increasingly sophisticated theoretical models that address ecological complexity. However, accurately matching observable patterns with process-oriented theory remains a persistent challenge. We identify potential pipelines connecting process and pattern and highlight areas for future progress.

Key words: Communities, metacommunities, patterns, metrics, processes, mechanisms, disordered systems models, joint species distribution models.

Introduction:

Ecological communities are shaped by a complex interplay of a limited number of basic ecological processes—or 'forces'—that influence species colonization, growth when rare, and persistence after establishment. While the number of fundamental **processes** is considered small (Vellend 2010, 2015), their combinations generate a wide variety of recognized **mechanisms** that organize community dynamics and structure (a glossary of terms in bold is provided in Table S1). This complexity is especially pronounced in systems with many species and when local communities interact across space as metacommunities.

The interplay among ecological processes (see below) gives rise to the patterns observed in metacommunity **data**. Recent advances in data collection and analytical methods have significantly improved our ability to detect, describe, and quantify metacommunity **patterns**. Yet, inferring the underlying processes and their relative importance from these patterns remains a major challenge in community ecology (Schaffer 1981; Sanderson and Pimm 2015; Leibold et al. 2022).

Vellend (2010, 2015), drawing an analogy with evolutionary theory, proposed that community dynamics are shaped by a handful of basic processes that shape community assembly and long-term dynamics . As in Thompson et al. (2020) we reframe Vellend's original processes to better align with contemporary ecological theory:

a) **Density-independent selection**: Species growth and persistence depend on their responses to environmental conditions, often (though not exclusively) abiotic. We assume no feedback from the biota to these environmental factors over time scales of interest.

b) **Density-dependent selection**: Biotic interactions with conspecifics or heterospecifics, including direct interactions and indirect effects mediated by other species and environmental factors that aren't measured.

c) **Dispersal**: Movement of organisms among local communities, influencing colonization ability and potentially leading to source–sink dynamics. Dispersal limitation restricts species to subsets of suitable sites, while dispersal excess allows persistence in otherwise unsuitable habitats. This could also include various forms of dormancy, which can be thought of as dispersal in time.

d) **Novelty and trait diversification**: Includes speciation, evolutionary change, and species introductions due to biogeographic shifts or anthropogenic influence (e.g., introductions). This component remains underexplored in metacommunity studies (but see Borregaard et al. 2014; Germain et al. 2021; Leibold et al. 2023).

e) **Stochasticity**: While stochasticity permeates all the above processes, we specifically highlight demographic stochasticity (random birth and death events, especially in small populations) and temporal environmental variation. These sources of variability are often conflated with measurement error (Shoemaker et al. 2020). Grouping them as variance components provides a practical way to account for their effects in ecological analyses of metacommunity data without the need for explicit mechanistic models.

In nature, these processes interact to generate various mechanisms (e.g., resource partitioning, species sorting, mass effects, trophic cascades) with distinct effects on community patterns. **Models** often assume specific relationships among processes—e.g., trade-offs or context dependence—to represent ecological mechanisms, typically in simplified systems involving few species. Classic models like Lotka–Volterra (Lotka 1925; Volterra 1927) or resource competition models (MacArthur 1974; Tilman 1982) illustrate how abiotic trade-offs and species interactions shape coexistence. Experimental validation of such models has occurred in simple plant, animal, and microbial communities (e.g., Gause 1932, 1934; Crombie 1944; Vandermeer 1965; Tilman 1980; see Kneitel and Chase 2004).

However, as the number of species increases, the potential combinations of **parameters** grow exponentially (and sometimes even factorially!), and multiple mechanisms often operate simultaneously. Most natural patterns emerge not from isolated species interactions but from the intertwined web of biotic interactions and environmental effects—Darwin's "entangled bank" (Darwin 1859; Schaffer 1981; Kéfi et al. 2016).

While we understand how specific processes can generate different distributional patterns in metacommunities, reliably inferring the underlying processes and mechanisms from observed patterns remains a major challenge. Multiple distinct process-based models can predict similar patterns (e.g., Barbier et al. 2018). This many-to-one mapping is both an opportunity and a challenge for ecology. It means that our models may effectively forecast responses to change (e.g., under climate change scenarios) without explaining the causal mechanisms involved. Such predictive capacity is highly valuable for ecological applications, including policy and management.

Yet, as René Thom noted, “To predict is not to explain” (Thom et al. 2016). As environmental changes push systems beyond the conditions that informed past predictions, explanatory understanding becomes essential. Identifying the processes underlying observed patterns is therefore critical for addressing contemporary environmental issues, grounding empirical ecology, and advancing ecological theory.

We structure this essay as follows:

1. We begin by defining “patterns” in metacommunities, contrasting those found in individual communities with those specific to metacommunities.
2. We then explore “processes” along with associated concepts such as mechanisms and models. We contrast two modeling approaches: a reductionist, bottom-up strategy and a top-down approach inspired by statistical mechanics.
3. We evaluate whether basic ecological processes can be robustly linked to patterns to test or validate models.
4. We assess the inverse—how patterns might be used to infer underlying processes.
5. Finally, we contextualize these discussions within current trends in community and metacommunity ecology, identifying opportunities to overcome outstanding challenges.

Taken together, these points offer a roadmap for unifying process-based and pattern-based approaches in community and metacommunity ecology. Despite the difficulty of linking pattern and process, we argue that there are significant reasons for optimism. Progress in data generation, modeling, and theoretical frameworks positions community and metacommunity ecology for continued advancement.

Data and Patterns

We begin by clarifying what we mean by "**data**" in the context of metacommunities. Data are direct observations of community variation across localities within a region and are typically structured as site-by-species matrices. These matrices may represent presence/absence, abundance, relative abundance, biomass, or other descriptors of community composition. Data can originate from natural systems, experimental studies, or simulations of an ecological model. From these matrices, we derive "patterns": simplified, often quantitative, summaries of ecological structure or variation.

Patterns range from local community attributes—such as species richness, diversity indices, species abundance distributions, and compositional turnover—to aggregated summaries across sites, such as beta-diversity or variability and synchrony in temporal fluctuations. Comparisons among sites often incorporate ancillary variables like environmental gradients, spatial structure, or spatial isolation.

Metacommunity-level patterns consider spatially-structured assemblages of multiple, interconnected communities. This broader focus enables exploration of how spatial dynamics, such as dispersal, interact with other structuring forces, such as internal patch dynamics. For example, Leibold et al. (2002) proposed the "Elements of Metacommunity Structure" to characterize coexistence patterns in spatially explicit metacommunities. Similarly, Cottenie (2005) applied variation partitioning methods (Borcard et al. 1992; Peres-Neto et al. 2006) to separate spatial and environmental components underlying community variation, aiming to diagnose dominant ecological processes.

Together, these methods yield a diverse menu of pattern metrics for characterizing communities and metacommunities (e.g., Ovaskainen et al. 2019; Guzman et al. 2022). Each metric serves as a potential clue about the processes and mechanisms shaping communities, though few are diagnostic on their own.

For example, researchers frequently examine how observed patterns deviate from null models—baseline expectations assuming the absence of particular ecological processes. While null models can become complex and contentious (e.g., Peres-Neto et al. 2001), most patterns deviate significantly from null expectations in at least some systems (Gotelli and McCabe 2002; Cottenie 2005).

More recent tools, such as Joint Species Distribution Models (JSDMs; Ovaskainen et al. 2017), allow for explicit modeling of co-occurrence structures while accounting for environmental and spatial variation. JSDMs provide a powerful framework for inferring latent (unmeasured) influences beyond measured environmental and spatial variables or species trait variation, including potential effects of biotic interactions.

However, such deviations alone do not identify the specific processes underlying metacommunity structure, as multiple models based on different mechanisms can produce convergent predictions. To improve inference, ecologists increasingly use cross-validation across multiple patterns (Holling and Allen 2002; Yanco et al. 2020). For instance, May et al. (2015) showed that while a neutral model (similar to Hubbell 2001) could individually match several empirical patterns from Barro Colorado Island, it could not do so with consistent parameter estimates. As a result, the model was rejected due to inconsistencies across patterns.

This strategy—testing models against suites of independent metrics—helps identify which mechanisms are more plausible. While not definitive, such multi-pattern approaches (e.g., Ovaskainen et al. 2019; Thompson et al. 2020; Guzman et al. 2021) represent a practical path toward more robust ecological inference.

Processes, Models, and Mechanisms

While patterns arise from data, we define "processes" as the underlying forces and interactions that generate those patterns. Unlike patterns, processes are not directly observable and must be inferred, often through modeling. This distinction—between what is observed and what is hypothesized to underlie the observations—has long posed a challenge in ecology, and definitions of processes vary widely in the literature (see S2 in the Supplementary Information).

Although Vellend's five-process framework remains comprehensive (Vellend 2010, 2016), the ways in which these processes can interact to shape ecological patterns remain unclear. In theory, processes operate in a quasi-linear fashion in short-term models—such as metacommunity Lotka–Volterra models (Gravel et al. 2016) or colonization–extinction models (Leibold et al. 2022). Yet, over longer timescales or under more complex ecological dynamics, the effects of these processes become entangled in ways that obscure clear causal links. Understanding this entanglement is central to resolving how pattern and process are related in community and metacommunity ecology.

Historically, ecologists have used mechanistic theories to explain patterns in ecological communities. One common strategy involves identifying a plausible mechanism—such as interspecific competition, predation, or mutualism—and modeling it with systems of differential (or difference) equations. For example, Lotka (1925) and Volterra (1927) modeled interspecific competition and predator–prey dynamics. These foundational models were empirically validated in microcosm experiments by Gause (1932, 1934) and others (e.g., Crombie 1945, 1946).

Such experiments were instrumental in testing theories involving biotic interactions—processes that are density-dependent. However, they often overlooked density-independent processes such as environmental filtering (e.g., Choler et al. 2001; Qi et al. 2018) or the role of spatial dynamics and dispersal (Amarasekare 2003; Peres-Neto et al. 2012).

Alternative models have focused on community patterns without assuming strong species interactions. These include spatial models assuming weak or no interspecific interactions (MacArthur and Wilson 1967; Connor and Simberloff 1979; Hart and Newman 2014) or neutral models in which all individuals are ecologically equivalent (Hubbell 2001). More comprehensive models that integrate species interactions, environmental filtering, and spatial dynamics are still relatively rare and often rely on complex simulations that are difficult to interpret, especially in species-rich or spatially structured systems.

Bottom-Up versus Top-Down Perspectives

As more species are included in models, their complexity increases dramatically (see Appendix 1). A bottom-up approach to managing this complexity is to model small sets of species using mechanistic **modules** (Holt and Hochberg 1999). These modules allow for detailed analysis of interactions, including indirect and higher-order effects, often in relation to environmental or spatial contexts (e.g., Tilman 1982; Holt et al. 1995; Leibold 1996). However, such models become analytically intractable with more than a few species—even three-species systems can be difficult to analyze (e.g., Ranjan et al. 2024). While they serve as an informative starting point, these modules inevitably oversimplify the dynamics of natural communities, which involve complex webs of direct and indirect interactions (Schaffer 1981; Kéfi et al. 2016).

An alternative is a top-down approach that seeks to explain patterns at the aggregate level, using simplified assumptions about species interactions. As a classic example, May (1972) used random matrix theory to analyze stability in large communities. He assumed that interaction coefficients among species were randomly distributed, and showed that community stability occurs when the number of species (S), the connectance (c), and the standard deviation of interaction strengths (σ) obey the inequality:

$$\sigma\sqrt{c\,(S-1)} < m$$

where m is the mean intraspecific self-limitation term. According to this inequality, communities with too many strongly connected species as compared with self-limitation are unlikely to be stable. Gravel et al. (2016) extended this model to metacommunities, showing that when dispersal and environmental heterogeneity are added, the stability condition becomes:

$$\sigma\sqrt{c\,(S-1)/N} < m$$

where N represents the effective number of uncorrelated local interaction matrices—a measure of environmental complexity that is always ≥ 1. This model suggests that metacommunities may support more species than local communities due to source–sink dynamics and spatial heterogeneity.

These "disordered systems" models rely on minimal assumptions and treat interaction networks as random objects, summarized by statistical moments (means, variances). However, many ecologists are uncomfortable with their oversimplified assumptions. For instance, Yodzis (1981) noted that trophic structure was a missing component in May's original formulation.

To address this, researchers have developed "partially structured" models (Ahmadian et al. 2015; Barbier et al. 2018; Carugno et al. 2022; Servan et al. 2025) that incorporate limited structure into interaction matrices—such as distinct subgroups (guilds), body size scaling, trophic levels, or evolutionary relationships. Reviews by Akjouj et al. (2024) and Cui et al. (2024) explore how such partial structure can be introduced into community matrices to increase realism without full complexity.

A clear question arises: how much structure should we impose? While including trophic structure might seem essential, further structure (e.g., spatial correlations or functional traits) risks reverting to highly constrained, bottom-up models. If over-structured, these models may become as complex and difficult to analyze as the bottom-up models they were intended to complement.

In sum, we can approach ecological modeling from two strategic directions (Figure 1). The bottom-up approach starts with simple population models and incrementally adds complexity. Predicted patterns are closely linked to mechanisms, but this approach is limited to small systems. The top-down approach, inspired by statistical mechanics, uses randomized or partially structured models (e.g., interaction matrices) to make robust but coarse-grained predictions in large systems. Each strategy has limitations: bottom-up models lack scalability, while top-down models lack mechanistic detail. A promising direction may lie in hybrid approaches that blend these strategies, enabling cross-validation of predictions and mechanisms. For instance, Miller et al. (2024) showed that specific mechanistic models (in their case, patch colonization–extinction dynamics) can be embedded within disordered systems models to explore how pairwise processes scale up in complex, realistic communities.

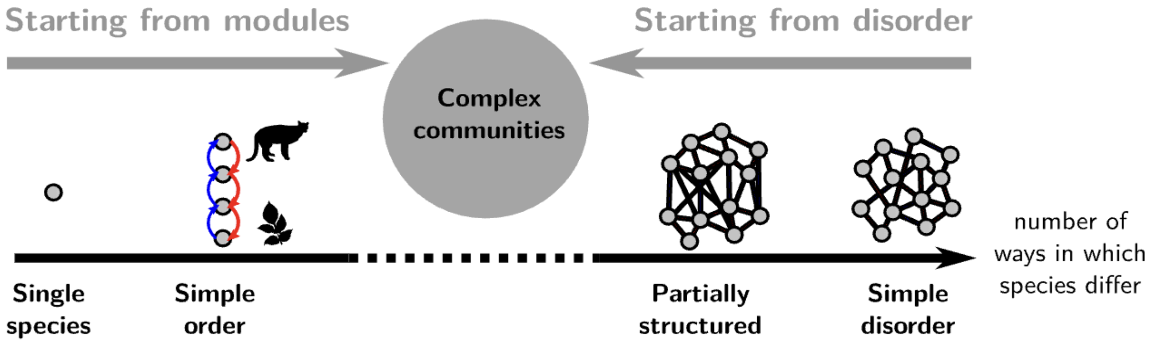

Figure 1: Two “strategic” approaches to modeling in community ecology. The more conventional “bottom-up” approach starts from the left and moves to the right. This is based on starting from single species population models, which can be arbitrarily complex to begin with, and adding species interactions in a structured way. For example, pairwise interactions may be combined to build up small sets of species as “modules” (Holt 1997). In principle, this could eventually lead to the analysis of realistically complex communities (center) that better represent those we find in nature, but progress seems to be exponentially harder as diversity increases. Alternatively, it is possible to think about communities as much more diffuse in nature. This approach is usually embodied by using random matrices to model species interactions (e.g., Random Lotka-Volterra Matrix or “RLVM” models), allowing the application of powerful mathematical methods. By progressively adding structural constraints on these models (e.g. distinct submatrices within the RLVM), they are converted into “partially structured models” that may approach realistic complex communities from another direction (right to left).

Linking Process to Patterns:

Both modeling approaches can be used to explore how a limited set of processes act together to shape metacommunity patterns. One important application of metacommunity models is to identify patterns that are associated with specific sets of mechanisms, as a key step toward inferring relevant processes from metacommunity patterns. By comparing predicted patterns – linked to specific mechanisms – against empirical patterns, it may be possible to diagnose the mechanisms that are acting in the metacommunity.

The bottom-up approach attempts to link process to pattern in a relatively direct way. Using bottom-up models to link processes to patterns is typically most successful when few species are involved, and when systems are relatively simple. Models of such systems frequently produce specific predictions that match empirical patterns both qualitatively and, in some cases, quantitatively (e.g., Vandermeer 1969; Friedman et al. 2017; Saavedra et al. 2017).

As previously noted, it is difficult to extend such models to systems with many species Aside from the challenge of building and analyzing models with many species, increasing complexity also makes it hard to identify a simple mapping between mechanisms and diagnostic patterns. This is an issue of convergence between models: different process-based models, even those grounded in distinct mechanisms, can produce similar or indistinguishable patterns. For example, McGill (2010) demonstrated that six mechanistically-distinct models could generate similar predictions for five widely used community patterns. None of the patterns were uniquely diagnostic of any specific model.

Given these challenges, one possible way to proceed is a "brute force" approach: associating modeled mechanisms with a larger suite of metrics or patterns (an illustrative list is shown in Table 1) to find unique combinations of predictions.This approach makes it possible to extend the one-to-one mapping between processes and patterns further into complex settings. Holling and Allen (2002) and Yanco et al. (2020) argue that evaluating models in this way can help eliminate less plausible candidates. A more targeted variation on this approach would identify a smaller subset of particularly informative or complementary metrics, which might enhance diagnostic power and increase efficiency (ruling out more mechanisms with less data).

Table 1: A representative set of possible metrics and patterns that can be derived from the site-by-species data matrix. Illustrative examples adapted and extended from Guzman et al. 2022. These represent a limited selection of a very large number of statistics and derived patterns that ecologists have explored and are meant to assist in our narrative of linking patterns to processes.

Metric or Pattern	Examples
Simple Descriptive Ecological Statistics	
Summaries	Means or variances of abundance, biomass, relative and absolute density, functional traits.
Distributions	Of abundances, incidences, functional traits, biomass, spatial or temporal occupancy, species-abundance distributions.
Diversity	Alpha and gamma, Hill numbers, functional and phylogenetic diversity.
Simple Derived Patterns	
Turnover	Spatial or temporal beta diversity; distance-decay, environmental-decay (with ancillary data).

Scaling relationships	Species-area, Taylor power, rarefaction or sampling curves.
Network structure	Nestedness, co-occurrence, centrality.
Second Order Derived Patterns	
Species network structure that varies as a function of environmental variation.	Metacommunity model with landscape fragmentation; time-varying trophic interactions.
Model Derived Outputs	
Variation partitioning only in space and environment	Variation partitioning of space and environment. Latent variables for each and their interactions.
Variation partitioning in time and space	Same as above but also with time.
Results from distribution models incorporating species covariances	Variation partitioning of space, environment, time, and species co-distributions; e.g., from JSDMs and similar approaches.

Ultimately, the brute force approach functions as a model selection procedure: models that fail to explain the full pattern set are rejected. However, this approach does not necessarily confirm the remaining models, since new and untested alternatives may perform as well or better. For example, Ovaskainen et al. (2019), Thompson et al. (2020), and Guzman et al. (2021) applied sets of idealized and alternative models to test against a panel of pattern metrics. They found that while brute force filtering could efficiently rule out some models, it was often inconclusive in narrowing down to a single best-fit model. Including more or better metrics may help, but data will often become limiting in empirical systems.

A complementary alternative is inspired by disordered systems (top-down) modeling. Instead of seeking a one-to-one correspondence between detailed process-based models and observed patterns by enlarging the set of patterns to keep pace with the larger and larger parameter space of models, this approach seeks specific features of models that remain uniquely associated with specific patterns even as the system of interest becomes very complex. For example, Barbier et al. (2021) used data from grassland plots to predict relative yield distributions based on mean and variance in interspecific interaction strengths— linking summary statistics of interactions to a specific pattern without attempting to parameterize specific pairwise interactions. Similarly, in another study, Barbier et al. (2023) found that pairwise correlations in species abundances could reflect interaction variances and carrying capacity heterogeneity, even when actual interaction coefficients were poorly known.

The key to this approach is to identify high-level parameters or parameter combinations that characterize specific ecological mechanisms and relate them to specific, diagnostic patterns. In contrast to brute force approaches, this top-down strategy targets robust, emergent features of community structure as signatures of underlying processes. However, by design, this approach cannot resolve low-level mechanistic details of metacommunity dynamics.

While this dichotomy between bottom-up and top-down approaches is simplified, it highlights contrasting philosophies in ecological modeling, and how both can be applied to link process to pattern. Hybrid strategies that integrate these perspectives—particularly through partially structured models—may offer another path forward. Appendix 1 further explores some nuances in combining these frameworks.

Linking Patterns to Processes:

In the section above, we considered how metacommunity models can be used to link processes to patterns, providing, in turn, a way to link observed patterns back to underlying processes. This approach to inference relies on first building out a map between processes and patterns, but this is complicated by the breakdown of a one-to-one mapping as the systems of interest become more complex. The bottom-up and top-down modeling strategies suggest two distinct approaches to coping with this issue: a brute force approach, where larger sets of patterns are used to maintain the one-to-one mapping, and a disordered systems approach where one seeks higher-level mappings between summary statistics or other emergent features of models and data. Both of these approaches are inherently limited – by data availability or by the ability to identify diagnostic patterns.

Is there an alternative approach that avoids these limitations? One possibility is to use methods that more directly decompose the sources of variation in community patterns. Ovaskainen et al. (2017) proposed a method for doing this using Joint Species Distribution Models (JSDMs), which decompose observed variation in species distributions into components attributable to environment, space, species co-distribution (potentially indicative of interactions), and stochasticity.

Following the logic of traditional species distribution models, “classic” JSDMs apply a sequential partitioning of variation: first accounting for environment, then for spatial effects, and finally for residual co-distribution among species. If environmental predictors are comprehensive, any remaining co-distribution may reflect biotic interactions. Figure 2a illustrates this sequential variation partitioning.

However, JSDMs can also be used in a non-sequential framework, in which each component—environment, space, co-distribution—is estimated simultaneously. This leads to a more complex pattern of shared and unique contributions (Figure 2b), where the interpretation of overlaps becomes ambiguous. For instance, two species might be mutually exclusive along an environmental gradient because of direct environmental filtering, competitive exclusion, or both. Without further assumptions, these effects cannot be disentangled (Dormann et al., Blanchet et al. 2018, Poggiatto et al. 2020).

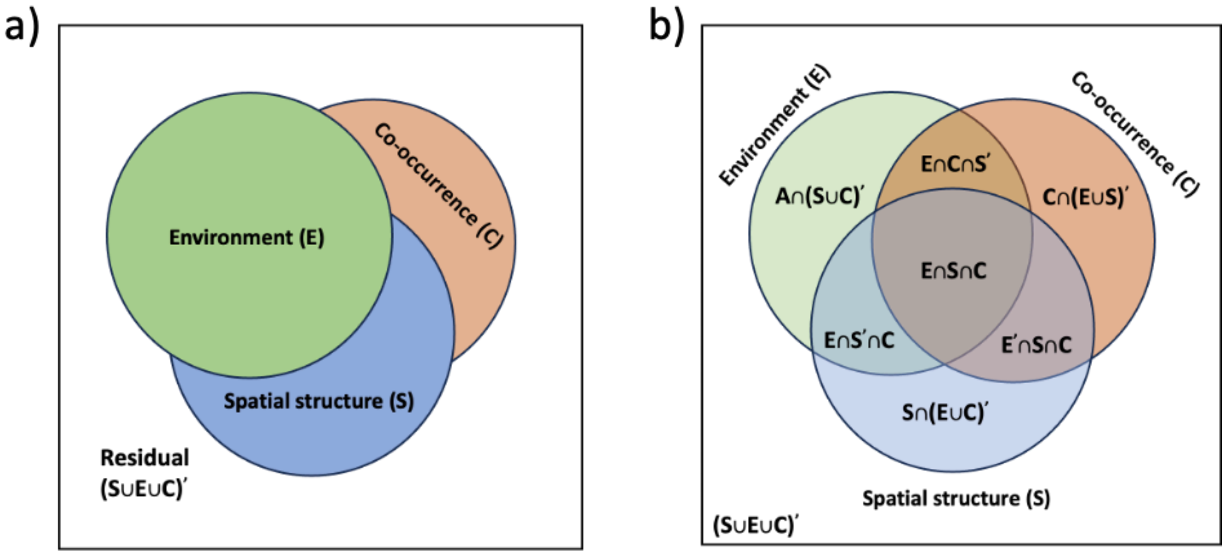

Figure 2: Variation partitioning alternatives for JSDM represented by Venn diagrams. The total variation in community composition in the metacommunity is encompassed by the outer square and is equal to 1. The white part of the figure is the unexplained or residual variation that is not accounted for by any of the predictors in the model. a) Accounting for predictors in a sequential order: First the measured environmental predictors are used and their contribution to the community variation is quantified as represented by the green area. Then spatial predictors are used to further account for community variation and their marginal contribution is shown in blue. Finally, the non-random latent correlations among the species are described, and their marginal contribution is quantified by the orange area. b) A non-sequential variation partitioning approach. Here, unique contributions from environmental, spatial, and codistribution components are represented (i.e., fractions without overlaps). Additionally there are a number of components that account for community variation that cannot be uniquely described (fractions representing predictor intersections). An interesting possibility is to compare the results of JSDM or other related methods across different metacommunities (e.g. Khattar and Peres-Neto 2024, Peak et al. 2024)

Thus, while JSDMs offer powerful tools for detecting and predicting community patterns, they do not necessarily provide unambiguous inference about underlying ecological processes. Nevertheless, they represent one of the most promising methods currently available for linking pattern to process.

A key direction for future progress may involve refining process models to better align with the data structures used in pattern-based inference. For example, Leibold et al. (2022) proposed analyzing the internal structure of metacommunities by parsing JSDM components across species and sites. This decomposition reveals how different species respond to environmental and spatial variation and how their distributions co-vary.

Unlike traditional pattern metrics, which aggregate across species, this approach allows process inference to operate at the species level. In doing so, it opens the possibility of aligning species-specific effects with bottom-up models that operate on subsets of interacting taxa. This could eventually allow for a true integration of bottom-up and top-down perspectives, as outlined in Figure 3.

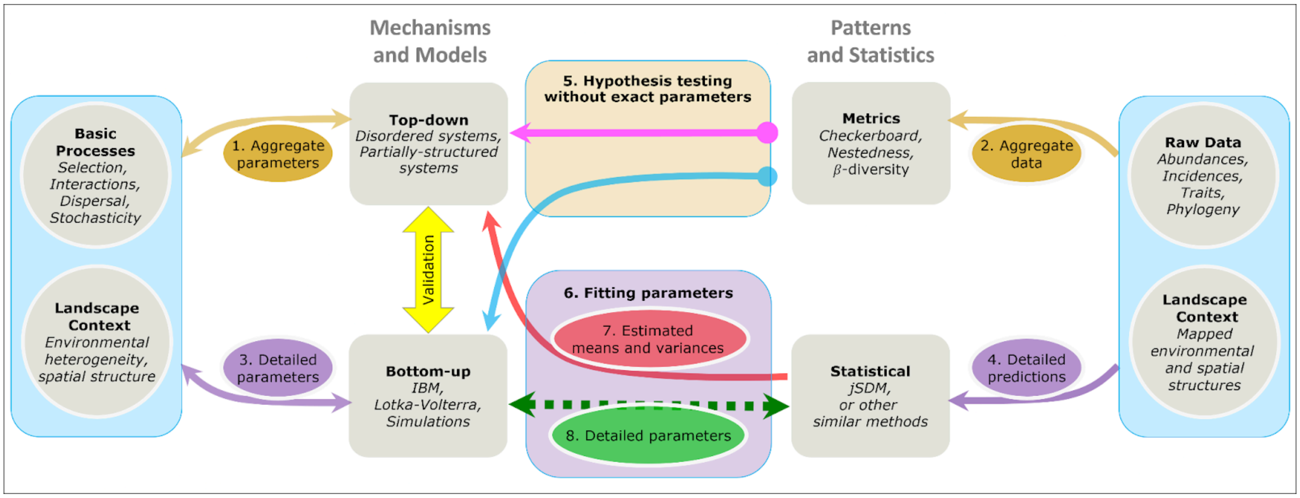

Figure 3: Pipelines between ‘Process’ (left) and ‘Pattern’ (right). We distinguish between approaches that aggregate parameters and/or data (upper pipeline) and those that focus on maintaining detailed parameters and predictors (lower pipeline). We defined process-based approaches as those that combine basic processes to make predictions about resulting patterns within landscapes that consider levels of heterogeneity and spatial structure (left side of the figure). There is a ‘top-down approach’ that uses ‘disordered’ or ‘partially structured’ models by aggregating parameters (typically, into mean, variances and covariances of these parameters) to predict robust patterns (but that ignore ‘fragile’ detailed components of those patterns). Alternatively, there is a ‘bottom-up’ approach that specifies parameters, typically in models with relatively few species, and uses detailed model specifications (e.g., using individually based models (IBMS), Lotka-Volterra models with specified parameters, or other simulation models). Such models can also predict some of the same general patterns as the top-down models but aim to be able to make more detailed predictions that may produce detailed fit to data. There is a parallel structure for approaches to deriving patterns from data (typically the site-by-species matrix, along with ancillary data on species traits/interrelations and/or explicit landscape features). One approach derives aggregated pattern descriptors. These include a wide array of ‘metrics’ such as species area relations, diversity and related methods, etc. A more recent effort has been directed at identifying parameter values from data. These metrics provide means of rejecting particular hypotheses (5. and 6., magenta and blue lines) but do not normally parameterize the processes. A prominent example is the use of JSDMs to infer metacommunity mechanisms (e.g. Ovaskainen et al. 2017, Leibold et al. 2022). While JSDMs are powerful methods to identify patterns, neither they nor related methods are able to resolve the entanglement of the five processes to clearly parameterize mechanistic models (dashed green line). In the absence of such detailed inference of models, Barbier et al. 2023 proposed that the detailed output of such studies can be broken down into the predictions of means, variances, and other moments, and thus providing another way to link top-down models to data (7. red arrow). An intriguing point is to explore how bottom-up and top-down methods might be cross-validated (yellow arrow).

Figure 3 illustrates potential inference pipelines that connect process-based models and data-based metrics. The top half of the figure emphasizes aggregated descriptors (e.g., means, variances), while the bottom half highlights parameter-specific models. Arrows show how models and metrics can reject or support hypotheses, and where gaps in inference still remain—particularly the persistent entanglement of processes such as dispersal, environmental filtering, and species interactions.

In summary, the mapping from patterns to processes remains complex and often ambiguous. However, methodological advances—particularly JSDMs and species-level decomposition—offer a promising route forward. Further refinement of both models and inference tools may help to close the gap between observed community patterns and the underlying processes that generate them.

Prospectus

We have outlined the conceptual and methodological challenges involved in linking processes and patterns in ecological communities and metacommunities. These issues trace back to the origins of community ecology, yet recent decades have seen tremendous progress. The development of metacommunity theory has notably reshaped how we interpret spatial biodiversity patterns (Leibold and Chase 2018). On the analytical side, species distribution models have evolved from environmentally focused predictors to those incorporating spatial effects and species co-distributions, thereby increasing predictive power from roughly 20–30% to 70–80% (Leibold and Peres-Neto, in prep).

Despite these advancements, substantial challenges remain. Chief among them is the uncertainty of current inference approaches. This problem is especially clear in JSDMs, where the effects of density-independent environmental filtering and density-dependent biotic interactions are often conflated. Similar ambiguity arises in multi-metric brute force approaches, as illustrated by McGill (2010), and in simulation studies (Chave et al. 2002, Ovaskainen et al. 2019; Thompson et al. 2020; Guzman et al. 2021) that failed to uniquely support any single mechanistic model.

It is important to acknowledge that many existing methods were not designed to infer mechanisms with high specificity. Often, the goal has been to reject null models or generate predictive models without mechanistic interpretation. While such approaches are valuable—particularly in applied contexts—the need for mechanistically grounded prediction grows as we face novel environmental conditions.

Encouragingly, the field is evolving. Developments in data collection, modeling, and computation suggest a promising trajectory rather than a fixed endpoint. We anticipate major progress in the coming years, driven by:

- **Improved data streams** that enhance sampling coverage, temporal resolution, and accuracy.
- **Greater computational capacity** to analyze complex datasets and run more sophisticated models—including those incorporating machine learning or artificial intelligence.
- **Advances in process modeling**, both in modular bottom-up approaches (e.g., for eco-evolutionary dynamics) and in statistical-mechanics-based top-down approaches, particularly through partially structured models.
- **Increased emphasis on validation and prediction**, to better link model outputs with empirical data and strengthen the integration of metrics and models.

To accelerate progress, we suggest several promising directions:

1. **Integrate temporal dynamics and multi-scale data**: Most metacommunity data remain spatial, but temporal metrics offer powerful insights. For example, Guzman et al. (2022) showed how time-based metrics improve interpretation of dynamics. Temporal data also facilitate causal inference via concepts like Granger causality, which can help infer directionality in species responses—something not possible from spatial data alone.
2. **Design and incorporate experimental manipulations**: Experiments can validate observational inferences and provide stronger process-based insights (Werner 2001; Grace 2024). Effective experimental designs—such as manipulating species presence/absence across environments—could directly test candidate mechanisms identified in models. Excitingly, advances in sequencing and high-throughput culture methods have fostered an explosive growth in experimental microbial ecology; the unique tractability of microbial (meta)communities makes them a promising platform for such experimental manipulations.
3. **Bridge top-down and bottom-up approaches**: Rather than treating these as oppositional approaches, hybrid strategies can combine mechanistic insights from modules with the robustness of statistical aggregation. Partially structured models are one way to integrate these perspectives. For instance, Barbier et al. (2018) showed how modest structure imposed on random matrices could shift predictions significantly, combining both realism and analytical tractability.

Community ecology, and particularly metacommunity theory, is undergoing a methodological and conceptual transformation. Understanding the links between processes and patterns remains a central goal. Doing so not only enhances ecological theory but also improves prediction under environmental change. Purely predictive models will struggle as conditions diverge from historical baselines. Mechanistic understanding, while harder to achieve, provides a necessary complement to maintain predictive power. By reviewing current approaches and their limitations, while highlighting promising future directions, we aim to support and accelerate progress in this critical area. Ultimately, bridging the gap between pattern and process is not only an heuristic ambition but also an essential step toward predictive and causal ecology. By leveraging the complementary strengths of diverse approaches and fostering integration across empirical, theoretical, and computational fronts, we can move closer to a unified framework for understanding the dynamics of biodiversity.

Acknowledgements

This study was conceived and executed by a working group of participants of the National Science Foundation-funded workshop, “Merging Statistical Theory and Analyses at the Interface of Microbial and Macrobial Ecology” (Award Number: NSF2224331, Montreal, Canada, 2022). VFF was supported by the NSF Postdoctoral

Research Fellowship in Biology (NSF Grant no. 2410350) and the Stanford Science Fellows program (Stanford University).

Literature Cited:

Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N. M., Wang, J., Yu, D. W., & Ovaskainen, O. (2021). Accounting for species interactions is necessary for predicting how arctic arthropod communities respond to climate change. *Ecography*, *44*(6), 885–896. https://doi.org/10.1111/ecog.05547

Ahmadian, Y., Fumarola, F., & Miller, K. D. (2015). Properties of networks with partially structured and partially random connectivity. *Physical Review E*, *91*(1), 012820. https://doi.org/10.1103/PhysRevE.91.012820

Akjouj, I., Barbier, M., Clenet, M., Hachem, W., Maïda, M., Massol, F., Najim, J., & Tran, V. C. (2024). Complex systems in ecology: A guided tour with large Lotka–Volterra models and random matrices. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *480*(2285), 20230284. https://doi.org/10.1098/rspa.2023.0284

Amarasekare, P. (2003). Competitive coexistence in spatially structured environments: A synthesis. *Ecology Letters*, *6*(12), 1109–1122. https://doi.org/10.1046/j.1461-0248.2003.00530.x

Amarasekare, P., & Nisbet, R. M. (2001). Spatial Heterogeneity, Source-Sink Dynamics, and the Local Coexistence of Competing Species. *The American Naturalist*, *158*(6), 572–584. https://doi.org/10.1086/323586

Barbier, M., Arnoldi, J.-F., Bunin, G., & Loreau, M. (2018). Generic assembly patterns in complex ecological communities. *Proceedings of the National Academy of Sciences*, *115*(9), 2156–2161.

Barbier, M., Bunin, G., & Leibold, M. A. (2023). *Getting More by Asking for Less: Linking Species Interactions to Species Co-Distributions in Metacommunities* [Preprint]. Ecology. https://doi.org/10.1101/2023.06.04.543606

Barbier, M., De Mazancourt, C., Loreau, M., & Bunin, G. (2021). Fingerprints of High-Dimensional Coexistence in Complex Ecosystems. *Physical Review X*, *11*(1), 011009. https://doi.org/10.1103/PhysRevX.11.011009

Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, *23*(7), 1050–1063. https://doi.org/10.1111/ele.13525

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the Spatial Component of Ecological Variation. *Ecology*, *73*(3), 1045–1055. https://doi.org/10.2307/1940179

Borregaard, M. K., Rahbek, C., Fjeldså, J., Parra, J. L., Whittaker, R. J., & Graham, C. H. (2014). Node-based analysis of species distributions. *Methods in Ecology and Evolution*, *5*(11), 1225–1235. https://doi.org/10.1111/2041-210X.12283

Carugno, G., Neri, I., & Vivo, P. (2022). Instabilities of complex fluids with partially structured and partially random interactions. *Physical Biology*, *19*(5), 056001. https://doi.org/10.1088/1478-3975/ac55f9

Chang, C.-Y., Bajić, D., Vila, J. C. C., Estrela, S., & Sanchez, A. (2023). Emergent coexistence in multispecies microbial communities. *Science*, *381*(6655), 343–348. https://doi.org/10.1126/science.adg0727

Chave, J., Muller-Landau, H.C. and Levin, S. (2002). Comparing classical community models: theoretical consequences for patterns of diversity. *The American Naturalist* 159, 1-23.

Choler, P., Michalet, R., & Callaway, R. M. (2001). FACILITATION AND COMPETITION ON GRADIENTS IN ALPINE PLANT COMMUNITIES. *Ecology*, *82*(12), 3295–3308. https://doi.org/10.1890/0012-9658(2001)082[3295:FACOGI]2.0.CO;2

Connell, J. H. (1961). The Influence of Interspecific Competition and Other Factors on the Distribution of the Barnacle Chthamalus Stellatus. *Ecology*, *42*(4), 710–723. https://doi.org/10.2307/1933500

Connell, J. H. (1983). On the Prevalence and Relative Importance of Interspecific Competition: Evidence from Field Experiments. *The American Naturalist*, *122*(5), 661–696. https://doi.org/10.1086/284165

Connor, E. F., & Simberloff, D. (1979). The Assembly of Species Communities: Chance or Competition? *Ecology*, *60*(6), 1132–1140. https://doi.org/10.2307/1936961

Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters*, *8*(11), 1175–1182. https://doi.org/10.1111/j.1461-0248.2005.00820.x

Crombie, A. C. (1945). On competition between different species of graminivorous insects. *Proceedings of the Royal Society B: Biological Sciences*.

Crombie, A.C. (1946). Further experiments on insect competition. *Proceedings of the Royal Society of London. Series B - Biological Sciences*, *133*(870), 76–109. https://doi.org/10.1098/rspb.1946.0004

Cui, W., III, R. M., & Mehta, P. (2024). *Les Houches Lectures on Community Ecology: From Niche Theory to Statistical Mechanics* (arXiv:2403.05497). arXiv. https://doi.org/10.48550/arXiv.2403.05497

Darwin, C. (1859). *On the Origin of Species*. John Murray.

Friedman, J., Higgins, L. M., & Gore, J. (2017). Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution*, *1*(5), 0109. https://doi.org/10.1038/s41559-017-0109

Gause, G. F. (1932). *Experimental Studies on the Struggle for Existence In Mixed Population of Two Species of Yeast*. *9*, 389–402.

Gause, G. F. (1934). *The Struggle for Existence*. Williams and Wilkins.

Germain, R. M., Hart, S. P., Turcotte, M. M., Otto, S. P., Sakarchi, J., Rolland, J., Usui, T., Angert, A. L., Schluter, D., Bassar, R. D., Waters, M. T., Henao-Diaz, F., & Siepielski, A. M. (2021). On the Origin of Coexisting Species. *Trends in Ecology & Evolution*, *36*(4), 284–293. https://doi.org/10.1016/j.tree.2020.11.006

Gleason, H. A. (1926). The Individualistic Concept of the Plant Association. *Bulletin of the Torrey Botanical Club*, *53*(1), 7. https://doi.org/10.2307/2479933

Goh, B. S. (1979). Stability in models of mutualism. *The American Naturalist*, *113*, 261–275.

Gotelli, N. J., & McCabe, D. J. (2002). SPECIES CO-OCCURRENCE: A META-ANALYSIS OF J. M. DIAMOND'S ASSEMBLY RULES MODEL. *Ecology*, *83*(8), 2091–2096. https://doi.org/10.1890/0012-9658(2002)083[2091:SCOAMA]2.0.CO;2

Gould, E., Fraser, H., Parker, T., Nakagawa, S., Griffith, S., Vesk, P., Fidler, F., Abbey-Lee, R., Abbott, J., Aguirre, L., Alcaraz, C., Altschul, D., Arekar, K., Atkins, J., Atkinson, J., Barrett, M., Bell, K., Bello, S., Berauer, B., … Tompkins, E. (2023). *Same data, different analysts: Variation in effect sizes due to analytical decisions in ecology and evolutionary biology*. https://doi.org/10.32942/X2GG62

Grace, J. B. (2024). An integrative paradigm for building causal knowledge. *Ecological Monographs*, *94*(4), e1628. https://doi.org/10.1002/ecm.1628

Grace, J. B., Harrison, S., & Damschen, E. I. (2011). Local richness along gradients in the Siskiyou herb flora: R. H. Whittaker revisited. *Ecology*, *92*(1), 108–120. https://doi.org/10.1890/09-2137.1

Gravel, D., Massol, F., & Leibold, M. A. (2016). Stability and complexity in model meta-ecosystems. *Nature Communications*, *7*(1), 12457. https://doi.org/10.1038/ncomms12457

Gurevitch, J., Morrow, L. L., Wallace, A., & Walsh, J. S. (1992). A Meta-Analysis of Competition in Field Experiments. *The American Naturalist*, *140*(4), 539–572. https://doi.org/10.1086/285428

Guzman, L. M., Thompson, P. L., Viana, D. S., Vanschoenwinkel, B., Horváth, Z., Ptacnik, R., Jeliazkov, A., Gascón, S., Lemmens, P., Anton-Pardo, M., Langenheder, S., De Meester, L., & Chase, J. M. (2022). Accounting for temporal change in multiple biodiversity patterns improves the inference of metacommunity processes. *Ecology*, *103*(6), e3683. https://doi.org/10.1002/ecy.3683

Hart, S. P., Freckleton, R. P., & Levine, J. M. (2018). How to quantify competitive ability. *Journal of Ecology*, *106*(5), 1902–1909. https://doi.org/10.1111/1365-2745.12954

Harte, J., & Newman, E. A. (2014). Maximum information entropy: A foundation for ecological theory. *Trends in Ecology & Evolution*, *29*(7), 384–389. https://doi.org/10.1016/j.tree.2014.04.009

Holling, C. S., & Allen, C. R. (2002). Adaptive Inference for Distinguishing Credible from Incredible Patterns in Nature. *Ecosystems*, *5*(4), 319–328. https://doi.org/10.1007/s10021-001-0076-2

Holt, R. D., & Hochberg, M. E. (2001). Indirect interactions, community modules and biological control: A theoretical perspective. In E. Wajnberg, J. K. Scott, & P. C. Quimby (Eds.), *Evaluating indirect ecological effects of biological control. Key papers from the symposium "Indirect ecological effects in biological control", Montpellier, France, 17-20 October 1999* (1st ed., pp. 13–37). CABI Publishing. https://doi.org/10.1079/9780851994536.0013

Holt, Robert D. (1997). Community modules. In *Multitrophic interactions in terrestrial ecosystems* (pp. 333–350). Blackwell.

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.

Kéfi, S., Miele, V., Wieters, E. A., Navarrete, S. A., & Berlow, E. L. (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. https://doi.org/10.1371/journal.pbio.1002527

Khattar, G., & Peres-Neto, P. R. (2024). The Geography of Metacommunities: Landscape Characteristics Drive Geographic Variation in the Assembly Process through Selecting Species Pool Attributes. *The American Naturalist*, *203*(5), E142–E156. https://doi.org/10.1086/729423

Kneitel, J. M., & Chase, J. M. (2004). Trade-offs in community ecology: Linking spatial scales and species coexistence. *Ecology Letters*, *7*(1), 69–80. https://doi.org/10.1046/j.1461-0248.2003.00551.x

Leibold, M. A. (1996). A Graphical Model of Keystone Predators in Food Webs: Trophic Regulation of Abundance, Incidence, and Diversity Patterns in Communities. *The American Naturalist*, *147*(5), 784–812. https://doi.org/10.1086/285879

Leibold, M. A., & Chase, J. M. (2018). *Metacommunity Ecology, Volume 59*. Princeton University Press. https://doi.org/10.1515/9781400889068

Leibold, M. A., Govaert, L., Loeuille, N., De Meester, L., & Urban, M. C. (2022). Evolution and Community Assembly Across Spatial Scales. *Annual Review of Ecology, Evolution, and Systematics*, *53*(1), 299–326. https://doi.org/10.1146/annurev-ecolsys-102220-024934

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, *7*(7), 601–613. https://doi.org/10.1111/j.1461-0248.2004.00608.x

Leibold, M. A., & Mikkelson, G. M. (2002). Coherence, species turnover, and boundary clumping: Elements of meta-community structure. *Oikos*, *97*(2), 237–250. https://doi.org/10.1034/j.1600-0706.2002.970210.x

Leibold, M. A., Rudolph, F. J., Blanchet, F. G., De Meester, L., Gravel, D., Hartig, F., Peres-Neto, P., Shoemaker, L., & Chase, J. M. (2022a). The internal structure of metacommunities. *Oikos*, *2022*(1), oik.08618. https://doi.org/10.1111/oik.08618

Leibold, M. A., Rudolph, F. J., Blanchet, F. G., De Meester, L., Gravel, D., Hartig, F., Peres-Neto, P., Shoemaker, L., & Chase, J. M. (2022b). The internal structure of metacommunities. *Oikos*, *2022*(1), oik.08618. https://doi.org/10.1111/oik.08618

Lerch, B. A., Rudrapatna, A., Rabi, N., Wickman, J., Koffel, T., & Klausmeier, C. A. (2023). Connecting local and regional scales with stochastic metacommunity models:

Competition, ecological drift, and dispersal. *Ecological Monographs*, *93*(4), e1591. https://doi.org/10.1002/ecm.1591

Lotka, A. (1925). *Elements of Physical Biology*. Williams and Wilkins.

MacArthur, R. H. (1974). *Geographical Ecology: Patterns in the Distribution of Species*.

MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. Princeton University Press.

May, F., Huth, A., & Wiegand, T. (2015). Moving beyond abundance distributions: Neutral theory and spatial patterns in a tropical forest. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1802), 20141657. https://doi.org/10.1098/rspb.2014.1657

May, F., Wiegand, T., Lehmann, S., & Huth, A. (2016). Do abundance distributions and species aggregation correctly predict macroecological biodiversity patterns in tropical forests? *Global Ecology and Biogeography*, *25*(5), 575–585. https://doi.org/10.1111/geb.12438

May, R. M. (1972). Will a large complex system be stable? *Nature*, *238*, 413–414.

May, R. M. (1973). *Stability and Complexity in Model Ecosystems. .* (Vol. 6). Princeton University Press.

McGill, B. J. (2010). Towards a unification of unified theories of biodiversity. *Ecology Letters*, *13*(5), 627–642. https://doi.org/10.1111/j.1461-0248.2010.01449.x

McPeek, M. A. (2022). *Coexistence in ecology: A mechanistic perspective*. Princeton University Press.

Miller, Z. R., Clenet, M., Della Libera, K., Massol, F., & Allesina, S. (2024). Coexistence of many species under a random competition–colonization trade-off. *Proceedings of the National Academy of Sciences*, *121*(5), e2314215121. https://doi.org/10.1073/pnas.2314215121

Mühlbauer, L. K., Schulze, M., Harpole, W. S., & Clark, A. T. (2020). gauseR: Simple methods for fitting Lotka-Volterra models describing Gause’s “Struggle for Existence.” *Ecology and Evolution*, *10*(23), 13275–13283. https://doi.org/10.1002/ece3.6926

Ovaskainen, O., & Abrego, N. (2020). *Joint species distribution modelling: With applications in R*. Cambridge university press.

Ovaskainen, O., Rybicki, J., & Abrego, N. (2019). What can observational data reveal about metacommunity processes? *Ecography*, *42*(11), 1877–1886. https://doi.org/10.1111/ecog.04444

Pascual, M. A., & Kareiva, P. (1996). Predicting the outcome of competition using experimental data: Maximum likelihood and bayesian approaches. *Ecology*, *77*, 337–249.

Peake, J., Leibold, M., Schrandt, M., & Stallings, C. (2024). *Assembly processes in estuarine metacommunities are dependent on spatial scale*. https://doi.org/10.22541/au.172506225.52119085/v1

Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation Partitioning of Species Data Matrices: Estimation and Comparison of Fractions. *Ecology*, *87*(10), 2614–2625. https://doi.org/10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2

Peres-Neto, P. R., Olden, J. D., & Jackson, D. A. (2001). Environmentally constrained null models: Site suitability as occupancy criterion. *Oikos*, *93*(1), 110–120. https://doi.org/10.1034/j.1600-0706.2001.930112.x

Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J. S., & Thuiller, W. (2021). On the Interpretations of Joint Modeling in Community Ecology. *Trends in Ecology & Evolution*, *36*(5), 391–401. https://doi.org/10.1016/j.tree.2021.01.002

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A., & McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, *5*(5), 397–406. https://doi.org/10.1111/2041-210X.12180

Qi, M., Sun, T., Xue, S., Yang, W., Shao, D., & Martínez-López, J. (2018). Competitive ability, stress tolerance and plant interactions along stress gradients. *Ecology*, *99*(4), 848–857. https://doi.org/10.1002/ecy.2147

Ranjan, R., Koffel, T., & Klausmeier, C. A. (2024). The three-species problem: Incorporating competitive asymmetry and intransitivity in modern coexistence theory. *Ecology Letters*, *27*(4), e14426. https://doi.org/10.1111/ele.14426

Saavedra, S., Rohr, R. P., Bascompte, J., Godoy, O., Kraft, N. J. B., & Levine, J. M. (2017). A structural approach for understanding multispecies coexistence. *Ecological Monographs*, *87*(3), 470–486. https://doi.org/10.1002/ecm.1263

Sanderson, J. G., & Pimm, S. L. (2015). *Patterns in Nature: The Analysis of Species Co-occurences* (s). University of Chicago Press.

Schaffer, W. M. (1981). Ecological Abstraction: The Consequences of Reduced Dimensionality in Ecological Models. *Ecological Monographs*, *51*(4), 383–401. https://doi.org/10.2307/2937321

Schoener, T. W. (1983). Field Experiments on Interspecific Competition. *The American Naturalist*, *122*(2), 240–285. https://doi.org/10.1086/284133

Serván, C. A., Capitán, J. A., Miller, Z. R., & Allesina, S. (2024). Effects of Phylogeny on Coexistence in Model Communities. *The American Naturalist*, E000–E000. https://doi.org/10.1086/733415

Shoemaker, L. G., Sullivan, L. L., Donohue, I., Cabral, J. S., Williams, R. J., Mayfield, M. M., Chase, J. M., Chu, C., Harpole, W. S., Huth, A., HilleRisLambers, J., James, A. R. M., Kraft, N. J. B., May, F., Muthukrishnan, R., Satterlee, S., Taubert, F., Wang, X., Wiegand, T., … Abbott, K. C. (2020). Integrating the underlying structure of stochasticity into community ecology. *Ecology*, *101*(2), e02922. https://doi.org/10.1002/ecy.2922

Shoemaker, W. R., Sánchez, Á., & Grilli, J. (2023). *Macroecological patterns in experimental microbial communities*. https://doi.org/10.1101/2023.07.24.550281

Sih, A., Crowley, P., McPeek, M., Petranka, J., & Strohmeier, K. (1985). Predation, Competition, and Prey Communities: A Review of Field Experiments. *Annual Review of Ecology and Systematics*, *16*(1), 269–311. https://doi.org/10.1146/annurev.es.16.110185.001413

Ter Braak, C. J. F., & Prentice, I. C. (1988). A Theory of Gradient Analysis. In *Advances in Ecological Research* (Vol. 18, pp. 271–317). Elsevier. https://doi.org/10.1016/S0065-2504(08)60183-X

Thom, R., Noël, E., Tsatsanis, S. P., & Lisker, R. (2016). *To predict is not to explain: Conversations on mathematics, science, catastrophe theory, semiophysics, natural philosophy and morphogenesis*. Thombooks Press.

Thompson, P. L., Guzman, L. M., De Meester, L., Horváth, Z., Ptacnik, R., Vanschoenwinkel, B., Viana, D. S., & Chase, J. M. (2020). A process-based metacommunity framework linking local and regional scale community ecology. *Ecology Letters*, *23*(9), 1314–1329. https://doi.org/10.1111/ele.13568

Tilman, D. (1980). Resources: A Graphical-Mechanistic Approach to Competition and Predation. *The American Naturalist*, *116*(3), 362–393. https://doi.org/10.1086/283633

Tilman, D. (1982). *Resource Competition and Community Structure*.

Vandermeer, J. H. (1969). The Competitive Structure of Communities: An Experimental Approach with Protozoa. *Ecology*, *50*(3), 362–371. https://doi.org/10.2307/1933884

Vellend, M. (2010). Conceptual synthesis in community ecology. *Quarterly Review of Biology*, *85*, 183–206.

Vellend, M. (2016). *The Theory of Ecological Communities*. Princeton University Press.

Volterra, V. (1926). *Variazioni e Fluttuazioni Del Numero d'individui in Specie Animali Conviventi*. *2*, 31–113.

Werner, E. E. (2001). Ecological experiments and a research program in community ecology. In *Experimental Ecology: Issues and Perspectives* (pp. 3–26). Oxford University Press.

Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, *30*(3), 279–338. https://doi.org/10.2307/1943563

Yanco, S. W., McDevitt, A., Trueman, C. N., Hartley, L., & Wunder, M. B. (2020). A modern method of multiple working hypotheses to improve inference in ecology. *Royal Society Open Science*, *7*(6), 200231. https://doi.org/10.1098/rsos.200231

Yodzis, P. (1981). The stability of real ecosystems. *Nature*, *289*(5799), 674–676. https://doi.org/10.1038/289674a0

Appendix 1: Glossary of terms:

Table S1. An annotated glossary of terms used in this essay.Their intended definition is to refine how we link patterns to processes in metacommunities (here, our system of study).

Term	Definition
Data	Direct observations of the focal elements in a system. Here, usually represented by a site-by-species-by-time matrix of presence/absence, abundance/biomass, or relative abundance/biomass.
Pattern	Informative or interesting regularities within a system often described as simplified representations of data.
Mechanism	Specific configurations of processes and constraints driving processes that produce particular or relevant outcomes. Mechanisms enable the stages of a process to occur by providing the elements necessary for each step in a process .
Processes	Fundamental elements of the ecological dynamics of a system. A series of steps necessary to achieve a particular pattern or behaviour in a system. These are reasonably well summarized by Vellend's (2010) suggestion to link to well-developed evolutionary theory.
Parameters	Variables or constants used to define and/or control the behavior or characteristics of a system or process.
Module	Small set of (possibly interacting) species/taxa/entities.
Model	A simplified representation of an ecological system, constructed to understand, analyze, and

	predict outcomes within an ecological system such as a population, a community, or a metacommunity. Models are usually encoded in mathematical equations which track variables whose dynamics are modulated by parameters.

Appendix 2: How are communities and metacommunities represented by matrices?

When there are more than a few (say 3) species, using individual equations to solve their mathematical effects becomes difficult. Fortunately, mathematicians have developed tools that greatly facilitate this using matrix algebra. This applies to linked sets of equations, those that depend on the parallel structure of linear equations that are linked to each other because they have the same parameters. In community ecology this means taking a matrix that does this to the Lotka-Volterra models, then solving for a possible equilibrium point and linearizing the equations near that point to get something called “the Jacobian” of the matrix. It is then possible to solve for the dynamic behavior of these equations near the equilibrium. With this, the most important point is resolving whether this point is realistic and then stable/unstable. For an ecologist using Lotka-Volterra models, the first thing is to find out if the equilibrium point has any negative values; if so, this would mean negative densities and be meaningless. The second is that if it checks out (all positive values), then solving to find the conditions (if any) when the point is stable. This is done by calculating the ‘dominant eigenvalue’ that describes the tendency of the whole system to come back to equilibrium. Stability here is inferred if the ‘real part’ of this eigenvalue is less than 0.

May (1972) did this by developing the idea of the “community matrix”, which can be expressed as the Jacobian of the matrix describing the interactions among a group of species in a single local community (see Novak et al. for more details on this). This matrix is often denoted by the letter “A” and is the Jacobian of the interactions in the community.

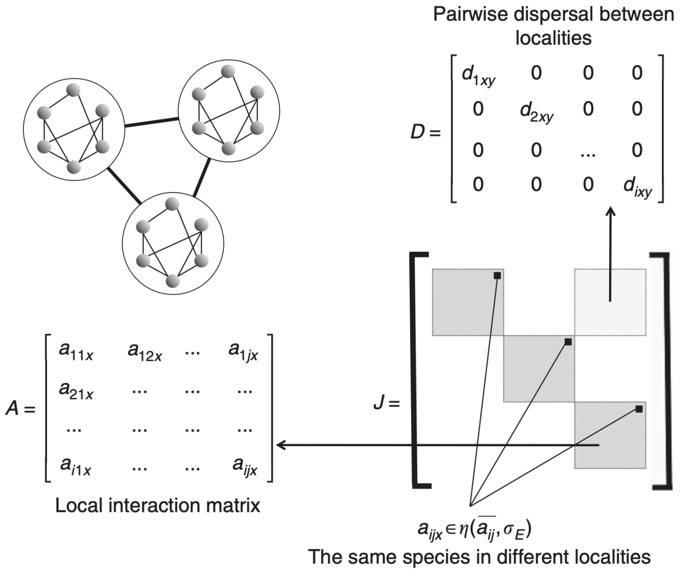

Figure B1: Going from local interaction matrices that describe how all species interact with each other (here this is the A matrix) by adding dispersal parameters to obtain the equivalent metacommunity matrix. This is done by lining up all the local matrices (keeping in mind that their separate values can be arbitrarily different) along the diagonal of the metacommunity matrix (darkly shaded part of the J matrix). Dispersal can then be described by other entries that are off diagonal light shading (matrices that have the structure shown in the D matrix. These matrices describe the dispersal rate (going from the column to the equivalent row elements of this matrix. They will only have entries along the main diagonal since species cannot change their identities while they disperse.

When May (1972) used this method on the A matrix where all off-diagonal entries were sampled randomly, he found that the system tended to have a very distinct stability only when:

$$\sigma \sqrt{c\,(S-1)} < m$$

Where S is the number of species present, c is the proportion of interactions that are not zero, δ is the variance in interspecific interaction strength, and *m* is the average intraspecific interaction strength.

Gravel et al. (2016) expanded this very basic model by asking how this would manifest in a metacommunity. To do so, they linked multiple local communities via dispersal as described in Figure B1. The stability criterion was then:

$$\sigma \sqrt{c\,(S-1)/N} < m$$

Where N was the ‘effective number of ecologically distinct’ local matrices. That is to say, May’s inequality was modified because ecologically distinct communities acted multiplicatively to enhance diversity in the metacommunity. This was also true in each local community since dispersal at these levels allowed them to do so in each locality as sink populations as well.

Of course, there are many other ways that multispecies ecology has used matrices such as the Jacobian of the community (and now the metacommunity), but the comparison of May (1972) and Gravel et al. (2016) illustrates some of the major features that allow the modeling of purely local interactions (May 1972), here modeled as a completely disordered system, and metacommunity (Gravel et al 2016) dynamics, here modeled as a partially structured extension of May's (1972) work.

Appendix 3: Contrasting challenges for modeling approaches at the interface between patterns and processes

Figures in the main text present a simplified contrast between approaches to modeling communities, focusing mainly on the “bottom-up” and “top-down” perspectives. In fact, there are other important axes of variation between models, implying different methods and difficulties when trying to bridge the process-to-pattern gap.

To understand why different modeling approaches may encounter distinct challenges, we categorize them as follows:

1. **How the model is selected/tested against patterns**
    a. **Qualitative prediction:** The model predicts the presence or absence of some qualitative feature in the data, e.g. environment filtering suggests a positive correlation between phylogenetic relatedness and co-occurrence. This is often the case of simple theoretical scenarios such as the main metacommunity paradigms (Leibold et al. 2004), and one common challenge is finding ways to compare how strongly the evidence supports one paradigm over another.
    b. **Fitting:** Alternatively, a model can also generate a range of hypothetical outcomes, among which we identify the best match(es) to our observations by quantitative model fitting. One important challenge is the risk of mis-specifying the model, so that the best fit within its range is not the most appropriate picture of the world.
    c. **Cross-pattern consistency:** If the model is successful at predicting one pattern, does it correctly predict other patterns? As the number of congruent patterns increases, they each act to more strongly validate the model.
2. **How it represents processes**

   a. **As qualitative assumptions:** The model, explicitly or sometimes implicitly, makes a statement about the presence or importance of one or more processes (e.g. absence of dispersal, dominance of abiotic filtering). This is especially relevant when we want to consider scenarios that differ in many ways (e.g. aquatic vs terrestrial biomes, wind dispersal vs chemotaxis) so that they are more easily represented as distinct archetypes rather than on a very multidimensional continuum. The risk is not knowing how different assumptions would have changed the model's predictions.
   b. **As parameters:** Quantitative knobs relating to processes (e.g. rate of each process).This is especially tractable when we want to consider scenarios that differ continuously in a few significant ways. One challenge is interpreting parameter values and how they relate to predictions, for instance interaction rates and dispersal rates are in different units and cannot easily be compared to determine which of the two processes is most important, and theoretical analysis is needed to know how strong dispersal must be to homogenize a metacommunity.

3. **Level of description:**
   a. **Disaggregated**: We retain species identities by characterizing them by their species-specific parameters, and we seek to derive their individual contributions to overall community-level patterns. The usual problem in metacommunities is the lack of sufficient statistical power to resolve these properties for up to hundreds of species.
   b. **Aggregated:** We lose species-specific labels, and instead look at overall distributions or summary parameters and metrics. This method characterizes "disordered systems models" described in text. It can, however, produce meaningless results when we aggregate variables that cannot be (for instance, it is not obvious that we can sum biomasses or richness of species at different trophic levels).

   We note that both may be used together. For example, Guisan and Rahbek (2011) combine macroecological constraints and SDMs to obtain more informative predictions. Or we can fit a JSDM at the species level (as in Leibold et al. 2022) and then aggregate some components over all the species. We also note that, in principle, there can be some sort of intermediate level aggregation such as described for "partially structured models".

4. **Process interplay:** How does our model assume that different processes (e.g. environments, dispersal, interactions, including all the processes that make up the “context” e.g. biome type) come together?
    a. **Totally separable:** we can ask of each process type what is its contribution to patterns, irrespective of what others are doing (possibly after some transformation, e.g. GLMs and GAMs). For instance, an SDM attempts to quantitatively partition variance in species abundance into effects of spatial fluxes and environment.
    b. **Totally entangled:** Various specific combinations/rankings of processes can give various outcomes, without any way to separate them into contributions of each process. For instance, the “patch dynamics” metacommunity paradigm captures the effects of having dispersal and stochasticity and local interactions acting together, without being able to meaningfully separate these processes.
    c. **Dynamical models:** A special case where we assume that processes are totally separable in the short-term rate of change (often not measured), but entangled in outcomes such as observed abundances.
    d. **Multiple regimes** can be seen within the same model (often seen for dynamical models), several “regions” within which processes have consistent (and at least conceptually, if not practically/quantitatively, separable) effects on patterns/outcomes, but these effects differ from region to region – e.g. the notion of “limiting factor”, with a region of limitation where increasing the factor steadily increases its importance and consequences, then a region without limitation where changing it has no impact.
5. **How a model is interpreted:** Do we try to use the model to infer underlying mechanisms and resolve causal relations that link process and pattern? (Our ability to do so is conditioned by the choices made along the previous axes.)
    a. **Prediction only:** Model assumptions and parameter values are not interpreted in any way, but only used to predict patterns in other similar situations. Processes thus only intervene in building the model (e.g. deciding which parameters and relations to include), but not in interpreting it. This includes the usual practice involving SDMs and JSDM to forecast future changes in species distributions under changing environmental or landscape conditions.
    b. **Qualitative agreement:** If the model correctly predicts patterns, we interpret this as suggesting that its qualitative assumptions regarding

processes are correct (for instance, seeing a spatial checkerboard pattern is interpreted as indicative of competition).

c. **Partitioning:** The values of model parameters are not necessarily meaningful in themselves, but they are used to deduce the relative importance of various processes e.g. by variance partitioning.

d. **Theory:** Model assumptions and parameter values must be analyzed theoretically to clarify their ecological implications (for instance, we have a theory of how strong interactions should be to induce nonequilibrium dynamics within a patch).

Some Examples:

We find that models linking processes and patterns span much of (if not all of) the range of attributes described above. To illustrate this, Table S1 classifies model attributes used across a wide array of studies, with many of these coming from our own work as individual researchers.

Example patterns, processes, and interpretations for various models typically used in metacommunity research.

Model	**Selection by data/ patterns**	**Representation of processes**	**Level of description**	**Process interplay**	**Interpretation**
Competitive checkerboard vs null model	Presence or absence of specific pattern	Categorical: competition vs stochasticity	Species	Single dominant process	Qualitative
Functional trait over/underdispersion	Presence or absence of specific pattern	Categorical: competition vs env. selection	Species	Single dominant process	Qualitative
Biodiversity-Function relations	Fitting a specific pattern	Parameters: selection/complementarity	Community	Separable	Partitioning
Environmental niche model	Fitting to full compositio	Parameters	Species	Separable	Partitioning

	nal data				
Joint Species Distribution Model	Fitting to full compositional data	Parameters	Species	Separable	Partitioning
Fitted gLV	Fitting to full composition or time series data	Parameters	Species	Dynamical, multi-regime	Requires theory
Probabilistic macroecology (Grilli)	Fitting several distributions	Categorical: species fitness + stochasticity	Community	Separable	Qualitative
Disordered systems model	Global means and variances (covariances) of parameters	Parameters	Community	Separable	Requires theory
Partially structured model	Subsetted means and variances (covariances) of parameters	Parameters	Community	Separable	Requires theory