Ecological modelling approaches for predicting emergent properties in microbial communities

Naomi Iris van den Berg, Daniel Machado, Sophia Santos, Isabel Rocha, Jeremy Chacón, William Harcombe, Sara Mitri and Kiran R. Patil

Recent studies have brought forward the critical role of emergent properties in shaping microbial communities and the ecosystems of which they are a part. Emergent properties—patterns or functions that cannot be deduced linearly from the properties of the constituent parts—underlie important ecological characteristics such as resilience, niche expansion and spatial self-organization. While it is clear that emergent properties are a consequence of interactions within the community, their non-linear nature makes mathematical modelling imperative for establishing the quantitative link between community structure and function. As the need for conservation and rational modulation of microbial ecosystems is increasingly apparent, so is the consideration of the benefits and limitations of the approaches to model emergent properties. Here we review ecosystem modelling approaches from the viewpoint of emergent properties. We consider the scope, advantages and limitations of Lotka-Volterra, consumer-resource, trait-based, individual-based and genome-scale metabolic models. Future efforts in this research area would benefit from capitalizing on the complementarity between these approaches towards enabling rational modulation of complex microbial ecosystems.

Microbial communities profoundly contribute to human and planetary health. Yet, the quantitative principles underlying community composition and assembly, and the link between community diversity and function, are largely unknown. Studying the emergent properties of microbial communities requires relating interactions across spatio-temporal scales, as well as their evolutionary dynamics. Unravelling, and ultimately predicting, community dynamics and emergent properties is a topological challenge in microbial ecology.

Higher-order units—from microbial consortia to entire ecosystems—feature macroscopic properties that emerge from microscopic interactions. Such emergent properties refer to any pattern or function that cannot be deduced as the sum of the properties of the constituent parts. Examples in microbial systems include resilience to abiotic and biotic perturbations, stable coexistence and biochemical abilities. Emergent properties typically arise when community members reach some threshold level of community size and connectivity. Conditions such as high flow rates or high medium viscosity can obstruct connectivity and hence limit the onset of emergent properties. Finding the threshold connectivity is often difficult, even in a defined community, due to the non-linear nature of emergence. The threshold is property-specific and may also (non-linearly) depend on environmental factors. Hence, descriptions of emergent properties have been largely qualitative, although synthetic biology approaches are starting to characterize emergent properties quantitatively.

Natural communities are difficult to study in situ due to their inherent dynamic nature and heterogeneity of habitats. Furthermore, abiotic variables influencing communities are difficult to measure or control in situ. Laboratory experiments, often considering the interactions between two or three community members, are amenable to quantitative analysis but limited in capturing emergent properties of natural, more diverse microbial communities. Defined communities with larger membership can be constituted in vitro yet are limited to culturable microbes. Therefore, mathematical models are indispensable for linking community composition and connectivity to emergent functions. Models can bridge principles learned from simple laboratory systems to complex natural systems; this is a continuity that is very difficult to achieve in an experimental setting.

Complex natural systems such as microbial communities are never fully closed, and more than one model will be plausible for any data available from the system of interest. The choice of the model depends on the research question and the data available to estimate the parameters. When experimental data are limited, models can be set up using first principles. Such models can be used to probe system responses to perturbations that are inaccessible to experimentation or observations. In comparison to statistical analyses, models can help in generating mechanistic hypotheses and establishing causal relations. Examples include the identification of bacteria conferring colonization resistance against Clostridium difficile and emergence of stability through competition.

Nevertheless, models are often most useful in complementation with statistical patterns; for example, a recent study using metabolic modelling uncovered polarization between cooperation and competition in microbial communities.

Two general approaches have been commonly used to model community dynamics and emergent properties: ecological models with species or cells as basic units and interactions among them as the focus, and genome-scale metabolic models that have intracellular reactions as main units and nutrient generation/consumption as the focus. As the use of genome-scale metabolic models has previously been reviewed, we focused on four commonly used ecological models. We provide an overview of their advantages in...
Review Article

A wide variety of mathematical models have been developed to address various ecological properties (Fig. 2): compositional stability; response to abiotic perturbations or invasions; environment dependency of interspecies interactions; spatio-temporal patterns of colonization and succession; and the conflict between collective versus individual interest (that is, the public goods problem, Black Queen hypothesis). Choosing the right model depends on the question(s), the assumptions made and the amount of information known or feasibly measurable. We review the following models: Lotka–Volterra; MacArthur consumer–resource; trait-based; individual-based; and genome-scale metabolic models (which can be regarded as metabolism-centred, trait-based models). We note that boundaries between these model types are not always clear-cut and adaptations to one model type can take it into the realm of another. For the sake of clarity, we structured this review in accordance with the common formulations of these modelling types.

Lotka–Volterra models. The Lotka–Volterra model is the most basic ecological model, originally used to describe macrofauna/habitat predation–prey dynamics, such as those between the Canadian lynx and snowshoe hare. This was among the first mathematical approaches to describe complex ecological dynamics. The interspecies interaction coefficients in the model (Fig. 2a) can be adjusted to reflect different types of interactions, such as competition between the moose and snowshoe hare. The classic Lotka–Volterra model has been instrumental in studying two-species interactions and can be solved both numerically and analytically. However, the classic model is limited in that it generates oscillations that are neutrally stable (that is, do not have a characteristic amplitude or period) instead of true limit cycles. Small modifications in the model can thus lead to qualitatively distinct outcomes. Furthermore, the classic model does not capture interactions involving more than two species. Hence, of relevance to microbial communities is the generalized Lotka–Volterra model, which can accommodate any number of species.

The Lotka–Volterra model consists of a set of non-linear, coupled, first-order differential equations. The growth rate of a species population is described as a function of its intrinsic growth rate and the linear effects exerted by the other populations. The Lotka–Volterra model requires relatively few parameters, intrinsic growth rates for each species (as a vector) and interaction coefficients for all species pairs (as a matrix). The advantage of having few parameters is that they can be inferred from (co-)culture data or from correlations observed in metagenomic datasets. The interaction parameters are easy to interpret.

Ecological models

Metagenomics has enabled culture-free mapping of microbial diversity across a broad range of habitats. Statistical approaches have helped to reconstruct co-occurrence networks from which interspecies interactions are often inferred. However, co-occurrence does not capture asymmetric relations, nor implies direct interactions. Therefore, classical ecological models from macrofaunal ecosystems have been adapted to microbial communities to unravel ecological mechanisms underlying emergent properties. These adaptations consider the following differences between macrofaunal/habitat systems and microbial communities: (1) large community membership. With the number of individual cells in microbial communities often in the range of 10^6, the use of concepts from statistical physics, such as central limit theorem, is justified; (2) short generation times. Microbial generation times are often in the range of minutes to hours, creating possibilities for easier generation of empirical data for model parameterization and validation. Short generation times combined with rapid horizontal gene transfer mean fast dynamics and evolution of interactions, which can alter the balance between processes such as migration and competitive exclusion. Conversely, short generation times also confound the separation of ecological from evolutionary dynamics and thus simulation intervals need to be considered accordingly; (3) trophic layer separation. Microbial communities not only exhibit biomass transfers between members via classic, predator–prey-like trophic interactions but also, and primarily, via secreted molecules through cross-feeding. As many of the secreted common goods create one-to-many and reciprocal interactions, microbial communities cannot be structured according to the classic ecological notion of layered food webs.

Fig. 1 | Examples of emergent properties arising from community complexity. a, Stability and resilience. Interactions among community members can buffer against biotic and/or abiotic perturbations leading back to a stable compositional state. The blue arrow indicates a perturbation event. b, Phenotypes. Emergent functions, such as substrate use, biomass production and cross-protection, result from the cooperation between community members. c, Persistence. Strains that would otherwise face competitive exclusion can coexist via, for instance, the intermediary of a cross-fed metabolite secreted by another community member. d, Self-organization. Balance between competitive and cooperative interactions can lead to spatial patterning, such as the formation of clonal patches.
Fig. 2 | Model types commonly used for studying emergent properties of microbial communities. a–d. Ecological models. Four commonly used ecological models are illustrated with conceptual diagrams showing the captured ecological interactions/processes, and the associated mathematical model formulations. The equations shown are examples and could be varied to be more realistic by, for instance, making growth rates non-linearly dependent on resource availability, for example, via Monod-like kinetics (as in c) or by introducing species-specific carrying capacities to make growth logistic. Parts of the equation that correspond to elements in the conceptual diagram are matched by colour, whereas the grey elements are those not explicitly present in the equations shown. See Supplementary Table 1 for further details including the notations used. A uniform background colour represents a homogenous (and closed) environment. e, Genome-scale metabolic models. Akin to trait-based models, these aim at capturing the near-complete metabolic and biosynthetic capabilities of all community members as encoded in their genomes. a,b. Populations and/or resources are shown as groups for the purpose of illustration but the model assumes a spatio-temporally homogeneous distribution (effectively making the system like a well-stirred reactor). We further note that the boundaries between different model types depicted here (a–e) are not rigid and each model type can be extended beyond its typical application.
(The sign indicates either positive or negative effects of species on one another and the magnitude indicates effect strength.) This simplicity and phenomenological nature makes Lotka–Volterra models easy to generalize. In addition, equilibrium conditions and their stability can be identified analytically. Therefore, the Lotka–Volterra model has been widely used for connecting species interactions to community dynamics in various complex communities including the gut microbiome and cheese fermentation communities.

The model's simplicity also has its drawbacks. First, interactions are static in time and space, which contrasts with natural systems wherein changes in the abiotic environment or species densities can shift interactions from, for example, mutualism to competition. Positive feedback loops in the system can lead to unstrained population growth. Consequently, fitting data to a Lotka–Volterra model may underestimate the prevalence of mutualism. A third limitation is the assumption that community dynamics can be described only using pairwise interactions, whereas the interaction between a pair can be altered by the presence of a third species (that is, higher-order interactions). However, there is a debate on the added value of including higher-order interactions in modelling. Furthermore, in its usual form, the Lotka–Volterra model does not account for stochastic processes such as genetic drift or environmental fluctuations. Lastly, small molecule-mediated interactions such as metabolite exchange or quorum sensing are not explicitly modelled, limiting the study of emergence of community-level biochemical properties. This omission can also lead to qualitatively wrong conclusions.

Some of the limitations of Lotka–Volterra models can be addressed through more parameters. Unstrained growth in a mutualistic scenario can be circumvented by replacing linear species interaction terms with non-linear saturating functions. The net sign of species interactions can be allowed to vary through time by, for example, adding a joint carrying capacity. Higher-order interactions can also be included. Each of these additions comes with the difficulty in meaningful interpretations of the parameters. Also, the number of interaction parameters scales quadratically with community membership. To account for higher-order interactions without considerably increasing the parameter space, pairwise interaction coefficients can be made to represent the net apparent effects of both higher-order and pairwise interactions.

The requirement for a priori assignment of interaction types has been addressed through combining linear regression with bootstrap aggregation to estimate the topology of the interaction network in the presence of measurement noise. Similarly, the confidence intervals for inferred interaction coefficients can be estimated using stochastic fitting. Environmental heterogeneity has been modelled as fluctuations in mortality rates and by using time-varying interactions. Another means to incorporate stochasticity is through adding a noise term to the equations or by randomly drawing interaction coefficients and/or intrinsic growth rates from predefined distributions. Indirect, metabolite-mediated interactions can also be explicitly modelled as a quadratic species-metabolite interaction model. This allows for the emergence of stable coexistence and diversity via cross-feeding, a phenomenon that is generally missing from Lotka–Volterra type models.

In sum, the generalized Lotka–Volterra model and its extensions are useful when a coarse-grained understanding of an ecosystem is needed or when measurements of species traits are not available. As for emergent properties, the generalized Lotka–Volterra model and its adaptations are often used to model community diversity, stable coexistence and multistability.

MacArthur consumer–resource models. An alternative approach to the Lotka–Volterra model is a resource-explicit model, where a population's growth rate depends on chemical mediators such as nutrients, toxins or signals, whose concentrations are coupled to population dynamics. Explicit modelling of metabolic by-products allows for the microbes as active modulators of their environment. In contrast to the Lotka–Volterra model, where the net species interactions are a priori assumptions and researchers must grapple with what they mean mechanistically, resource-explicit models are defined ‘mechanistically’ by how species consume and deplete resources, excrete by-products and are hindered or helped by signals. Thus, species are not denoted as ‘competitors’ or ‘mutualists’ a priori, but these labels are outcomes of the model and may be time- and environment-dependent.

The most basic resource-explicit model is the MacArthur consumer–resource model. Like the Lotka–Volterra model, these models were initially used in macrofaunal ecology. MacArthur consumer–resource dynamics mathematically captures the classical theory of competitive exclusion or the R* resource-ratio hypothesis, that is, a species needs to be the best at consuming a certain resource to persist. The model allows one to quantify how resource use differs between species: the bigger the resource overlap, the smaller the realm of coexistence. The growth rate of each species is proportional to the resources consumed, weighted by an energy factor minus a maintenance requirement that includes death. Species interactions are not parameterized directly but are mediated by the depletion of (shared) resources. The resources themselves are assumed not to interact with each other but can be introduced and consumed by arbitrarily complex dynamics. Additionally, MacArthur consumer–resource equations allow for the use of different timescales for resources and consumers. In the limit where the dynamics of resources are fast, the system reduces to Lotka–Volterra dynamics.

To make the consumer–resource model applicable to microbial communities, it is necessary to extend its scope much beyond two species and two resources. For this, parameters such as uptake rates can be randomly drawn from statistical distributions. Analytical results can be derived using generalized versions of the model also, such as the absence of multistability for linear functional responses or species having identical stoichiometries of essential nutrients. To include cross-feeding, the model can be extended with stoichiometric coefficients describing how the consumption of one compound relates to the secretion of another. This method captures the emergent property of stable coexistence and diversity, even for a single externally supplied resource. In line with empirical observations, cross-feeding critically alters the model's qualitative behaviour, increasing the potential for coexistence over competitive exclusion.

One limitation of the model is the assumption of fixed metabolism of community members. This can be partly overcome by including dynamic metabolic adaptations, allowing each species to temporally change its preferred nutrient. Thus, emergent properties of community self-organization and persistent coexistence and maintenance of diversity could be captured. Another limitation is the absence of taxonomic and hierarchical metabolic structures. This has been addressed through imposing correlation structures of the consumer preference and metabolic matrix. To tackle missing thermodynamic and biochemical detail, which is a common limitation of many ecological models, proteomic and energetic constraints have been incorporated. This allowed the elucidation of emergent energy flow topologies in the community and distinct regimes of community structure.

The MacArthur consumer–resource model is deterministic in nature, which contrasts with the stochastic biotic and abiotic influences on community dynamics. Such probabilistic events can be simulated within the framework of consumer–resource models. Similarly, demographic noise can be incorporated, capturing the emergence of apparent neutrality, that is, giving all community members the appearance of being ecologically equivalent. Adaptation of the model to include environmental fluctuations...
revealed a U-shaped relationship between community diversity and disturbance intensity, which a Lotka–Volterra model failed to capture\textsuperscript{96}. To go beyond resources that positively contribute to growth\textsuperscript{108}, the model was adapted to include toxins, describing how resources and toxins jointly influence interspecies interactions. In line with results based on game theory\textsuperscript{117}, this adaptation demonstrated how coexistence and diversity are shaped by the hostility of the environment.

Explicitly coupling the dynamics of chemicals to that of populations enables capturing interactions mediated by more than one chemical, including those with opposing effects. For example, a species could be obtaining an essential resource from several community members. Thus, the effects of environmental change can be studied without the need to re-parameterize the system. The downside is that the system needs to be characterized in more detail: parameters for production and uptake rates for each chemical or chemical reaction rates need to be estimated, which can be difficult in practice\textsuperscript{110}. Additionally, chemical interactions must be posited a priori, a considerable challenge in complex or poorly characterized communities. When choosing statistical physics to overcome this problem, it is important to explicitly include thermodynamics in parameterization to ensure compliance with the second law of thermodynamics\textsuperscript{106–108}.

Yet, classic ecological models lack biochemical detail and do not have the resolution to explain experimental results beyond those explicitly modelled, for example, pH change\textsuperscript{98}. The MacArthur consumer–resource model and its adaptations are generally suitable to model emergent properties arising via cross-feeding such as diversity and coexistence.

### Trait-based models

The family of phenomenological models, besides the classic Lotka–Volterra and consumer–resource models, includes trait-based models wherein the focal variables are phenotypic traits rather than phylogenetic groups. Community members are defined by their traits and the models describe how trait combinations respond to and in turn influence environmental variables\textsuperscript{96,102,109}. The interactions between members and environmental variables can be directly defined and mediated by ecological trade-offs\textsuperscript{110,111–113} such as stress-tolerance versus combative dominance\textsuperscript{96}. Whereas the classic ecological models typically assume a constant environment, trait-based models focus on community dynamics along major environmental gradients. The system dynamics created by the interaction between various ‘guilds’ and their environment may lead to a functional trait distribution that optimizes a certain community-level functional property such as nitrification in soil systems\textsuperscript{98,112}, organic matter decomposition\textsuperscript{96,111} or resiliency to invasion\textsuperscript{112}. Trait-based models can elucidate the emergence of selection pressures and diversity patterns along modelled environmental gradients; these are important aspects to understand evolutionary processes\textsuperscript{113,114}. Trait-based models are not confined by a priori definitions of community structure and functional groups but instead allow for the emergence of species and groups with new trait combinations as simulated environmental conditions change\textsuperscript{117}.

The trait combinations, often referred to as ‘genes’, can be informed by (meta)genomic data\textsuperscript{114}, by theoretical considerations\textsuperscript{115} or randomly drawn from empirical distributions\textsuperscript{116}. Random sampling of trait combinations can even be used to simulate metagenomes and metatranscriptomes through the replacement of unfit phenotypes This strategy demonstrated the emergent property of community structure/assembly and ocean biogeochemistry\textsuperscript{117}. Trait-based models allow community members to not only respond to, but also influence, modelled environmental gradients\textsuperscript{113,114}. Furthermore, trait-based models can be applied to larger microbial communities than those in ocean systems\textsuperscript{113,116,117}, which are too big to be modelled via individual-based methods and too spatio-temporally heterogenous to be captured via classic ecological models. As trait-based models simulate how system-scale—reaching to global-scale—environmental changes influence community dynamics, applications have been largely restricted to where local variations in environmental variables are considered to be of negligible importance to overall emergent functionalities\textsuperscript{96}. Thus, trait-based models have been applied to microbial communities in oceanic\textsuperscript{104,105,111–113,115} or soil systems\textsuperscript{96,110–111,113,115}. For systems with high local variability such as the human gut, other approaches with higher resolution (for example, individual-based models) may be more appropriate.

Trait-based models provide a bridge between community structure and function through the environment\textsuperscript{122}. Empirical evidence has shown that microbial communities subjected to environmental change are more likely to exhibit shifts in functional trait distributions rather than in taxonomic distributions\textsuperscript{122,124}. A trait-based approach may hence be more useful than a strict taxon-based approach in light of environmental heterogeneity. This is in agreement with an in silico analysis based on a consumer–resource-based model\textsuperscript{122}. However, unlike the classic ecological models, a trait-based model’s main purpose is not to elucidate interaction dynamics among community members (defined in the Lotka–Volterra model by direct influence on fitness or in the MacArthur consumer–resource model through the relationship with a shared resource) nor how these relate to coexistence. Rather, trait-based models primarily consider, besides predation in a few cases\textsuperscript{125}, interactions between the community and environmental heterogeneity. Therefore, the influence of inter-member interactions, both direct, for example, antibiotic-mediated antagonism, and indirect, for example, cross-feeding, on fitness and community dynamics may be underestimated\textsuperscript{112}.

Trait-based models do not require prior imposition of trait combinations. However, they require defining which traits are included. This can pose a problem in systems such as the gut microbiota wherein both environmental gradients and associated fitness determinants are poorly known\textsuperscript{96}. This limitation has been addressed by, for example, changing the modelling unit from species to enzymes of interest\textsuperscript{117}. In addition, not all relevant environmental variables or interactions between environmental variables may be known, as is the case in poorly studied communities. This problem was tackled by modelling trait distributions along interaction effects between different factors\textsuperscript{110}. A trait-based model might be further limited in understanding community ecology under new environmental conditions without the introduction of new traits. Trait diffusion through subsequent generations could be used to overcome this limitation\textsuperscript{114}. However, neither emergent trait distribution nor randomized initialization fully capture evolution via natural selection, which requires de novo mutations and associated traits arising during model simulation.

In sum, if local environmental variations can be neglected, trait-based models, through defining community members as trait combinations, are well suited to simulate adaptation, emergent phenotypic structures\textsuperscript{122} and collective function\textsuperscript{123} along environmental gradients.

### Individual-based models

Increase in computational power has enabled the use of individual-based, also called agent-based, models. Like classic ecological models, individual-based models were first used to study macrofaunal communities such as bird flocks but increasingly also microbial communities\textsuperscript{124,125}. The individual-based model is inherently stochastic\textsuperscript{130}. It takes a bottom-up approach by modelling every community member as an individual interacting agent with each having any number of ‘rules’ of any complexity\textsuperscript{131}. This contrasts with the top-down models that simulate adaptations at the population level and hence cannot capture (micro)evolutionary dynamics starting from individual mutations.
While top-down models can simulate the fate of mutants, both mutants and their characteristics need to be predefined. Individual-based models can simulate the complete evolutionary process via inheritable mutations. This allows capturing the emergence of community resilience and phenotypic complexity in a changing and/or heterogeneous environment. The rules given to each individual cell can either be empirically informed, with recent advances in single-cell microbiology and flow cytometry allowing for realistic parameterization of single-cell behaviours, or based on genome-scale metabolic models. The inherent discrete nature of individual-based models allows capturing discrete events that cause a critical shift in cell activity, for example, cell attachment/disruption and community size via grid occupancy rules. By contrast, top-down models typically resort to a priori defined carrying capacities, community size via grid occupancy rules. By contrast, top-down models are those that consider communities in terms of population averages, whereas bottom-up models simulate communities at the individual level. The model types shown and their links to parameters and emergent properties are based on their common forms and applications. The model types can be adapted to be informed by additional parameters and be made suitable to model other emergent properties than indicated. For instance, genome-scale metabolic models can be used to study spatial patterns.

Another benefit to individual-based models is the possibility to include thermodynamic constraints on community interactions and diversity. With limited exceptions, adaptations of classic ecological models tend not to explicitly include energetic/thermodynamic constraints. Furthermore, complete accounting of thermodynamic constraints requires incorporation of pH and chemical speciation, which can be done with an individual-based model. Other important forces constraining individual cell activity/dispersal and community size are of physical/mechanical nature, such as fluid drag and shear forces. These can also be explicitly included. Moreover, inclusion of biophysical processes, such as cell shoving, can not only increase model accuracy but also reduce computation time through confining the solution space.

In addition to capturing individual-level heterogeneity, individual-based models are well suited for simulating stochastic processes, such as genetic drift, horizontal gene transfer and cell–cell interactions like type VI secretion stabbing. While changes in and dispersal of biomass are modelled via discrete individual interactions, soluble substrates/chemical species typically need to be modelled as reaction–diffusion systems.

The main limitation to individual-based modelling is its complexity, making it not only computationally expensive but also difficult to assess for robustness. Reduced dimensional descriptions are essential for robustness of conclusions, for instance, through setting hydrodynamic limits, moment closure and other approaches to aggregation. Another strategy is to limit the number of modelled agents, or group cell types to reduce the community-wide variation in agent-specific rules. Similarly, dividing the total cell population into cheaters and non-cheaters is a common approach in individual-based models investigating evolutionary game theory. This strategy demonstrated the emergence of a cooperative network. Simplifications of individual-based models still allow studying the behaviour of individuals after perturbation departing from the binary question of whether the system is stable.

Another limitation is a priori assumptions about direct interactions; if incorrect, owing to emergent (and unconstrained) feedbacks between individual and population behaviour, they can escalate to

---

**Fig. 3 | Choosing a suitable modelling approach.** The flow diagram indicates the model choice trajectory, starting with the available input parameters until the emergent properties that can be potentially modelled. The dotted line marks the potential for inferring interaction coefficients from omics data (for example, co-occurrence networks predicted by genome-scale metabolic models), albeit without being indicative of direct and density-dependent interactions. Top-down models are those that consider communities in terms of population averages, whereas bottom-up models simulate communities at the individual level. The model types shown and their links to parameters and emergent properties are based on their common forms and applications. The model types can be adapted to be informed by additional parameters and be made suitable to model other emergent properties than indicated. For instance, genome-scale metabolic models can be used to study spatial patterns.
outcomes incongruent with population-level data. Comparison of model types in predicting emergent properties is challenging due to variations in the mechanistic and biochemical details included as well as the number of modelled dimensions. Thus, the ability of individual-based models to identify potential local processes that lead to emergence of system-wide properties also varies.

In sum, individual-based models are generally suitable to identify spatially explicit emergent properties, such as self-organization, resilience to local perturbations and feedbacks between individual and population behaviour influencing the spread and fixation of a mutation. These models are also suitable for identifying system-wide emergent properties, such as community cooperation and coexistence\textsuperscript{22}; however, robustness and stability are hard to determine and can typically not be derived analytically.

**Genome-scale metabolic models**

Genome-scale metabolic models are mathematical representations of the near-complete set of biochemical reactions encoded in the genome of an organism. These models can be built in a fully or semi-automated manner from a genome assembly and ideally also by using physiological data, such as nutrient requirements\textsuperscript{114, 124}. They can be used to simulate the metabolic phenotype of an organism—including its growth rate, intracellular fluxes and secretion/uptake rates—and its response to environmental and genetic perturbations. The most common simulation approach, flux balance analysis\textsuperscript{123}, identifies an optimal metabolic state within the bounds of mass balance and reaction directionality constraints. The commonly used criterion for optimality (objective function) in case of (single) microbes is maximization of growth rate\textsuperscript{123}. Other relevant optimality criteria, such as maximum ATP production or minimal adjustment against a reference state, can also be used.

Genome-scale metabolic models have been applied to microbial communities in several ways. Flux balance analysis can be used to identify the range of feasible metabolic exchanges and community-scale metabolic properties. When simulating a community, choosing a biologically relevant objective function is often difficult and poses a major limitation to the approach. Yet, the foundation in first principles means that these metabolic models can identify the phenotypic range within which a community must operate. This advantage can be used, for example, to identify candidate cross-fed metabolites\textsuperscript{157} and assess the extent of metabolic cooperation and competition. Additionally, flux balance analysis can be used to predict compositional and metabolite dynamics by solving a series of steady-state problems in discrete time steps, while the uptake/secretion of metabolites is modelled as differential equations\textsuperscript{123}. Multiple species or genotypes can be modelled simultaneously allowing for quantitative prediction of community dynamics resulting from intracelluar metabolic events\textsuperscript{114, 124}. By further combining dynamic flux balance analysis with metabolite diffusion equations, spatial effects can also be captured as in biofilms\textsuperscript{156, 159}. Modelling frameworks employing this strategy, such as COMETS\textsuperscript{58} and CROMICS\textsuperscript{160}, can thus allow investigating both spatial and temporal dynamics.

In the context of ecological models, genome-scale metabolic models are akin to trait-based models. Yet, there are several distinctions: (1) a direct link to the (meta-)genome and hence the ability to simulate specific genetic changes; (2) scale (typically) hundreds to thousands of reactions and metabolites per species; and (3) the possibility to include additional molecular mechanistic details, such as proteome allocation and thermodynamic constraints. These distinguishing characteristics, together with advances in automated model reconstruction\textsuperscript{114, 161, 162}, have accelerated their use in modelling microbial communities. These models have provided testable hypotheses on (nutritional) habitat range, ecosystem productivity and cross-fed metabolites\textsuperscript{123, 157, 163}. They have also been used to simulate community dynamics in response to altered nutrient environment or addition/removal of species\textsuperscript{15}.

One of the main limitations of genome-scale metabolic models is the absence of negative interactions beyond resource competition, such as those mediated by toxins and other antimicrobials that are prevalent in ecosystems\textsuperscript{30, 31, 164}. Furthermore, these models often require extensive curation to achieve high accuracy, especially for microbes that are phylogenetically distant from model organisms like *Escherichia coli*. This problem can get further amplified when modelling at the single-cell level, as generalization can become difficult due to model complexity\textsuperscript{160, 161, 166}. Nevertheless, owing to their amenability for automated reconstruction and integrating (meta-)omics data\textsuperscript{14}, there is exciting potential for merging genome-scale metabolic models with ecological model approaches to capture emergent properties.

**Parameter inference and data integration**

As the quality of a model depends on the accuracy and uncertainty in parameter estimates, choice of model type and complexity is best guided by data availability (Fig. 3). The parameters required can be categorized into member-level parameters, such as carrying capacity, and interaction parameters, such as cross-feeding rates. The parameters should ideally be obtained in the community context. This is challenging owing to experimental difficulties in probing the combinatorial parameter space. Most models thereby resort to parameters obtained from monocultures or small-scale community studies. Developments in omics methods and high-throughput culturing and interaction mapping\textsuperscript{155, 156} are helping to fill the gap. Metagenomic analyses can suggest which members of the community may interact, for example, as statistical covariation\textsuperscript{115, 116}. Metagenomics can also help in estimating in situ growth rates using read coverage over the origin of replication\textsuperscript{115}. Meta-transcriptomics and metabolomics can be used to constrain the model structure, for example, by selecting interactions consistent with the data\textsuperscript{152, 170}.

**Discussion**

As microbial community dynamics and properties can be attributed to (the interaction of) various complex networks, such as gene regulatory, metabolic, cell-environment and cell–cell, identifying which of these should be included to accurately predict the emergent property of interest is key in choosing an appropriate modelling approach (Fig. 3 and Supplementary Table 2). Depending on the temporal and spatial characteristics of the property of interest, a combination of approaches is advantageous. Each approach has advantages and limitations in terms of integrating omics data, handling missing data, generating analytical solutions, predictive power, scope of applications and amenability for generating hypotheses that can be tested experimentally. The choice of model(s) also depends on whether a general phenomenological understanding is sufficient or if the goal is to fit data to generate predictions in a specific system. While top-down, population-based approaches (that is, Lotka–Volterra, MacArthur consumer–resource and phenomenological trait-based models) are generally more suited for modelling resilience and coexistence, bottom-up approaches (that is, individual-based and genome-scale metabolic models) are typically better suited for modelling phenotypic trait ranges and spatial patterns. Each of these approaches can be adjusted in the number of dimensions included and units modelled.

The distinction between the modelling approaches detailed in this review is not necessarily categorical. For example, adaptations of a consumer–resource model can make it fall within the category of individual-based models, whereas an individual-based model with few dimensions and low resolution can be less accurate in predicting the emergent phenotypic property of substrate use than a multi-population consumer–resource model. As ecosystem stability and resilience are a pressing issue on the health and environmental fronts\textsuperscript{5, 171, 172}, the need to develop predictive models cannot be overstated. With computational power becoming less limiting, and
molecular and single-cell data becoming increasingly available, we envisage that fusion between different ecological and cellular-level models will, in the coming years, enable predictive modelling of emergent properties in complex ecosystems from the molecular to the ecosystem scale.

Received: 23 September 2021; Accepted: 23 March 2022; Published online: 16 May 2022

References

1. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. Nat. Rev. Genet. 13, 260–270 (2012).
2. Shi, Z. Gut microbiota: an important link between Western diet and chronic diseases. Nutrients 11, 2287 (2019).
3. Lynch, S. V. & Pedersen, O. The human intestinal microbiome in health and disease. N. Engl. J. Med. 375, 2369–2379 (2016).
4. Glowacki, R. W. P. & Martens, E. C. In sickness and health: effects of gut microbial metabolites on human physiology. PLoS Pathog. 16, e1008370 (2020).
5. Nazaries, L. et al. Evidence of microbial regulation of biogeochemical cycles from a study on methane flux and land use change. Appl. Environ. Microbiol. 79, 4031–4040 (2013).
6. Komorpa, A. What is microbial community ecology? ISME J. 3, 1223–1230 (2009).
7. Flemming, H.-C. et al. Biofilms: an emergent form of bacterial life. Nat. Rev. Microbiol. 14, 563–575 (2016).
8. Bauer, E., Zimmermann, J., Baldini, F., Thielle, I. & Kaleta, C. BacArena: individual-based metabolic modeling of heterogeneous microbes in complex communities. PLoS Comput. Biol. 13, e1005544 (2017).
9. van Hoek, M. J. A. & Merks, R. M. H. Emergence of microbial diversity due to cross-feeding interactions in a spatial model of gut microbial metabolism. BMC Syst. Biol. 11, 56 (2017).
10. Gotter, F. A., Manhart, M. & Ackermann, M. Understanding the evolution of interspecies interactions in microbial communities. Philos. Trans. R. Soc. Lond. B Biol. Sci. 375, 20190256 (2020).
11. Wintermute, E. H. & Silver, P. A. Emergent cooperation in microbial metabolism. Mol. Syst. Biol. 6, 407 (2010).
12. Chen, J., Yoshinaga, M. & Rosen, B. P. The antibiotic action of methylarsenate is an emergent property of microbial communities. Mol. Microbiol. 111, 487–494 (2019).
13. Konstantinidis, D. et al. Adaptive laboratory evolution of microbial co-cultures for improved metabolite secretion. Mol. Syst. Biol. 17, e10189 (2021).
14. Park, H. et al. Artificial consortium demonstrates emergent properties of enhanced cellulose-sugar degradation and biofuel synthesis. NPl Biofilms Microbiomes 59 (2020).
15. Schwartzenberg, J. A. et al. Bacterial growth in multicellular aggregates leads to the emergence of complex lifecycles. Preprint at bioRxiv https://doi.org/10.1101/2021.11.01.466752 (2021).
16. Levin, R. & Lewontin, R. The Dialectical Biologist (Harvard Univ. Press, 1985).
17. Diaz, P. I. & Valm, A. Microbial interactions in oral communities mediate emergent biofilm properties. J. Dent. Res. 99, 18–25 (2020).
18. Buenger, A. N. et al. Gastrointestinal dysbiosis following diethylhexyl phthalate exposure in zebrafish (Danio rerio): altered microbial diversity, functionality, and network connectivity. Environ. Pollut. 265, 114496 (2020).
19. Kim, M. K., Ingreameu, E., Zhao, A., Basoler, B. L. & Stone, H. A. Local and global consequences of flow on bacterial quorum sensing. Nat. Microbiol. 1, 15005 (2016).
20. Ebrahimi, A. & Or, D. Hydration and diffusion processes shape microbial community organization and function in model soil aggregates. Water Resour. Res. 51, 9804–9827 (2015).
21. Falconer, R. E. et al. Microscale heterogeneity explains experimental variability and non-linearity in soil organic matter mineralisation. PLoS ONE 10, e0123774 (2015).
22. Fredrickson, J. K. Ecological communities by design. Science 348, 1425–1427 (2015).
23. Singer, E. et al. Next generation sequencing data of a defined microbial mock community. Sci. Data 3, 160081 (2016).
24. Marino, S., Baxter, N. T., Huffnagle, G. B., Petrosino, J. F. & Schloss, P. D. Mathematical modeling of primary succession of murine intestinal microbiota. Proc. Natl Acad. Sci. USA 111, 439–444 (2014).
25. Carboni, J., Gatelli, D., Liska, R. & Saltelli, A. The role of sensitivity analysis in ecological modelling. Ecol. Model. 203, 167–182 (2007).
26. Oreskes, N., Shrader-Frechette, K. & Belitz, K. Verification, validation, and confirmation of numerical models in the Earth sciences. Science 263, 641–646 (1994).
27. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 46, 7542–7553 (2018).
28. Ruffie, C. G. et al. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature 517, 205–208 (2015).
29. Hammarlund, S. P., Chacón, J. M. & Harcombe, W. R. A shared limiting resource leads to competitive exclusion in a cross-feeding system. Environ. Microbiol. 21, 759–771 (2019).
30. Coyle, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, functions, and applications. Science 350, 663–666 (2011).
31. Cariboni, J., Gatelli, D., Liska, R. & Saltelli, A. The role of sensitivity models will, in the coming years, enable predictive modelling of emergent properties in complex ecosystems from the molecular to the ecosystem scale.
118. Ratcliffe, C., Barrere, J. & Gore, J. Strength of species interactions determines biodiversity and stability in microbial communities. Nat. Ecol. Evol. 4, 376–383 (2020).

119. Bradford, M. A. et al. Quantifying microbial control of soil organic matter dynamics at macrosystem scales. Biogeochemistry 156, 19–40 (2021).

120. Ward, B. A., Dutkiewicz, S., Moore, C. M. & Follows, M. J. Iron, phosphorus, and nitrogen supply ratios define the biogeography of nitrogen fixation. Limnol. Oceanogr. 58, 2059–2075 (2013).

121. Severin, I., Östman, Ö. & Lindström, E. S. Variable effects of dispersal on productivity of bacterial communities due to changes in functional trait composition. Proc. Natl Acad. Sci. U.S.A. 112, 8025 (2015).

122. Staley, C. et al. Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. Front. Microbiol. 5, 414 (2014).

123. Worden, L. Conservation of community functional structure across changes in composition in consumer-resource models. J. Theor. Biol. 493, 110239 (2020).

124. van der Plas, F. et al. Plant traits alone are poor predictors of ecosystem properties and long-term ecosystem functioning. Nat. Ecol. Evol. 4, 1602–1611 (2020).

125. Song, H.-S. et al. Regulation-structured dynamic metabolic model provides an accelerated gro simulator. In Front. Microbiol. 10, 76–89 (2019).

126. van der Plas, F. et al. Plant traits alone are poor predictors of ecosystem properties and long-term ecosystem functioning. Nat. Ecol. Evol. 4, 1602–1611 (2020).

127. Song, H.-S. et al. Regulation-structured dynamic metabolic model provides an accelerated gro simulator. In Front. Microbiol. 10, 76–89 (2019).

128. Gutiérrez, M. et al. A new improved and extended version of the multicell individual-based modelling of biofilms. Methods Mol. Biol. 2349, 367–380 (2022).

129. Jayathilake, P. G. et al. A mechanistic individual-based model of microbial communities. PLoS ONE 12, e0181965 (2017).

130. Clark, J. R., Daines, S. J., Lenton, T. M., Watson, A. J. & Williams, H. T. P. Individual-based modelling of adaptation in marine microbial populations using genetically defined physiological parameters. Ecol. Modell. 222, 3823–3837 (2011).

131. Nadell, C. D. et al. Cutting through the complexity of cell collectives. Proc. Biol. Sci. 280, 20122770 (2013).

132. Allen, B., Gore, J. & Nowak, M. A. Spatial dilemmas of diffusible public goods. Interface Focus 2, 726–737 (2012).

133. Griesemer, M. & Sindi, S. S. Rules of engagement: a guide to developing agent-based models. Methods Mol. Biol. 2349, 367–380 (2022).

134. A. J. Shou, W. Spatial self-organization favors properties and long-term ecosystem functioning. Nat. Ecol. Evol. 4, 1602–1611 (2020).

135. Zwart, J. A., Solomon, C. T. & Jones, S. E. Phytoplankton traits predict fixation. Biogeochim. Acta 69, 504–515 (2000).

136. Allen, B., Gore, J. & Nowak, M. A. Spatial dilemmas of diffusible public goods. Interface Focus 2, 726–737 (2012).

137. Song, H.-S. et al. Regulation-structured dynamic metabolic model provides an accelerated gro simulator. In Front. Microbiol. 10, 76–89 (2019).

138. Allison, S. D. & Goulden, M. L. Consequences of drought tolerance traits for microbial decomposition in the DEMENT model. Soil Biol. Biochem. 107, 104–113 (2017).

139. Allison, S. D. A trait-based approach for modelling microbial litter decomposition. Ecol. Lett. 15, 1058–1070 (2012).

140. Doleman, A., Varghese, H., Miller, C. D. & Flann, N. S. Modeling de novo granulation of anaerobic sludge. BMC Syst. Biol. 13, 344–349 (2010).

141. Parise, F., Lygeros, J. & Russ, E. Bayesian inference for stochastic individual-based models of ecological systems: a pest control simulation study. Front. Environ. Sci. 3, https://doi.org/10.3389/fes.2015.00042 (2015).

142. Allison, S. D. & Goulden, M. L. Consequences of drought tolerance traits for microbial decomposition in the DEMENT model. Soil Biol. Biochem. 107, 104–113 (2017).

143. Allison, S. D. A trait-based approach for modelling microbial litter decomposition. Ecol. Lett. 15, 1058–1070 (2012).

144. Chacón, J. M., Möbius, W. & Harcombe, W. R. The spatial and metabolic basis of colony size variation. ISME J. 12, 669–680 (2018).

145. Worden, L. Conservation of community functional structure across changes in composition in consumer-resource models. J. Theor. Biol. 493, 110239 (2020).

146. Chacón, J. M., Möbius, W. & Harcombe, W. R. The spatial and metabolic basis of colony size variation. ISME J. 12, 669–680 (2018).

147. Worden, L. Conservation of community functional structure across changes in composition in consumer-resource models. J. Theor. Biol. 493, 110239 (2020).
NATURE ECOLOGY & EVOLUTION

Review Article

(grant no. 866028) and from the UK Medical Research Council (project no. MC_UU_00025/11). S.M. thanks the Swiss National Science Foundation for funding the Swiss National Centre of Competence in Research Microbiomes and an Eccellenza project and the ERC for starting grant no. 715097. W.H. received funding from the National Institutes of Health through grant no. R01-GM121498; S.S. received funding from the Portuguese Foundation for Science and Technology under the scope of a PhD grant (no. SFRH/BD/121695/2016) and the strategic funding of the UIDB/04469/2020 unit.

Author contributions
N.I.V.D.B. conceived the review, carried out the literature survey and wrote the manuscript. D.M., S.S. and I.R. contributed to the section on the genome-scale metabolic models. J.C., W.H. and S.M. helped with the literature survey and contributed to the manuscript structuring and writing. K.R.P. conceived the review, helped with the literature survey and wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01746-7.

Correspondence should be addressed to Kiran R. Patil.

Peer review information Nature Ecology & Evolution thanks Alvaro Sanchez, Stefanie Widder and Manoshi Datta for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2022