Suppl. Table 1. Notations and emergent properties associated with each model type.

	Notation	Emergent properties
A	The generalised Lotka-Volterra model is characterised by a set of interacting populations/species, B_i (orange bug) where $i = 1 \dots N$. Each population's intrinsic growth rate, r_i (dark blue arrow), can be impacted by other populations via corresponding interaction coefficients as defined in matrix α_{ij} (maroon arrow). Besides antagonistic interactions, each population is also limited by its intrinsic maintenance/death rate, m_i (bright red arrow).	Community resilience, co-existence
B	The generalised MacArthur Consumer-Resource model is characterised by populations and their direct interactions with resources/cross-fed metabolites, R_α (purple hexagons) where $\alpha = 1 \dots M$. The growth of each population is not only defined by intrinsic growth (or: biomass conversion) rates, but also by a resource utilisation matrix, $C_{i\alpha}$ (rate at which each species uptakes each resource), and a resource quality function, $\Delta w_{i\alpha}$, captures the amount of biomass each species produces per unit of each resource consumed while maintaining energy balance (as described in [88]). To include secretion of metabolic by-products, and therefore the possibility of cross-feeding, the model includes a stoichiometric matrix, $D_{\beta\alpha}^i$ (green dashed arrow), which defines the units of metabolites β (yellow circles) secreted by species i per unit resource α consumed.	Community resilience, coexistence, phenotypic range (i.e., substrate utilisation)
C	Trait-based models are characterised by a set of species interacting with a (set of) environmental factor(s), which influences e.g., nutrient uptake rates. (Initial) nutrient (and species) distributions can be considered uniform, however, there is environmental variability along at least one dimension, producing variable species distributions - mediated by defined trade-offs - along the modelled dimension(s). In the given example, the influence of the environmental factor, T (gradient along the x-axis), on growth rates is represented like an Arrhenius-like equation, where sensitivity to T is captured by a species-specific sensitivity coefficient, S_i (as in Grigoratou et al., [104]). In this example, the trade-off is intrinsic maximum growth rate versus sensitivity. Strategies produced by this trade-off could indirectly depend on modelled resources (e.g., a higher maximum growth rate at the expense of higher sensitivity only 'paying off' if the resource for growth, R_α, is available beyond its half-saturation constant, $K_{i\alpha}$). In the demonstrated case, the resource is fixed, and sensitivity (the trait) to the environmental factor, T, needs to be defined *a priori*.	Community resilience, phenotypic range, spatial structure
D	Individual-based models are defined by individual agents (e.g., cells) interacting with local environment via an *a priori* defined set of rules. In the given example, the model is defined by two spatial dimensions (x, y), where a cell, i, only gets to consume and grow (no longer density-dependent) on resources present within its current grid cell, with each grid cell allowing a maximum of one cell. As it grows, the cell secretes metabolic by-products, R_β (yellow circles), into its immediate environment (i.e., current grid cell). These subcellular agents can disperse across the space via diffusion. Diffusion operates on a different timescale than cellular growth and migration (with $\delta t \ll dt$), and hence, requires separation of timescales when simulated. Cells enter and exit grid cells via a set of rules or probabilities, e.g., a new daughter cell will occupy any of its mother cell's vacant neighbouring grid cells by some probability, P. Cell migration probabilities can be skewed by availability of certain nutrients, allowing for chemotaxis. Cell death events can be made random (e.g., via Moran process), allowing freeing up of grid cells. In the given example, grid cells are only variable by their occupation (which is binary) and concentrations of subcellular agents, but they can also be made environmentally heterogeneous by including variables such as temperature and pH (unless individual protons are modelled explicitly as well).	Community resilience, co-existence, phenotypic range, spatial structure
E	Genome-scale metabolic models are defined by mathematical representations of genome-scale metabolic reconstructions. In the given example, S_i is a stoichiometric matrix for species i, where its elements are stoichiometric coefficients for the corresponding metabolites (rows) and reactions (columns). Flux Balance Analysis [157] then provides the in- and output fluxes (and can be used to estimate growth rates). Unlike previous model types, R_α not only comprises extracellular (cross-fed) resources, but also intra-cellular metabolites, allowing for mechanistic understandings of, e.g., consequences of genetic changes (as opposed to *a priori* definitions). These models can later be used to inform parameters for individual- or trait-based models, or direct community flux simulations [49].	Community co-existence, phenotypic range (i.e., niche expansion through metabolite secretion and utilisation of resources inaccessible to individual members), spatial structure

Suppl. Table 2. Overview over emergent properties, benefits, and limitations for each discussed model type.

	Emergent properties	Model benefits	Model limitations
Lotka-Volterra	• Persistent coexistence & diversity • Resilience & (Multi)stability	• simplicity & generality • adheres to laws of mass action • can model arbitrarily many taxa • high-dimensional Lotka–Volterra systems allow for a large variety of different qualitative behaviours such as limit cycles, chaos or attractors (Stein et al., 2013) • parameters can be inferred from co-occurrence data • intuitive interpretation of parameters • equilibrium conditions and their stability can be attained analytically • explicitly describes population dynamics	• interaction coefficients fixed • assumes homogenous/well mixed environment • *a priori* assumptions interaction types • generating predictions from generalised Lotka-Volterra with many species requires fitting a large number of parameter values from empirical data, i.e., prone to overfitting • missing explicit indirect interactions • assumes only pairwise interactions • linear intrinsic growth rates • lacks biochemical detail • lacks stochasticity • no explicit energetics • lacks evolutionary process
MacArthur consumer-resource	• Resilience • Persistent co-existence & diversity • Community composition self-organisation • Substrate utilisation	• simplicity & generality • adheres to laws of mass action • can model arbitrarily many taxa • can be adapted to model cross-feeding • detailed quantitative specification of way how resource use differs among species • allows for the use of different timescales of resources and consumers • intuitive interpretation of parameters • equilibrium conditions and their stability can be attained analytically • explicitly describes population dynamics • permits arbitrary complexity of resource use • no *a priori* assumptions about species interactions	• assumes no interactions between the resources themselves. • assumes homogenous/well mixed environment • assumes species cannot switch between resources • lacks biochemical detail • lacks stochasticity • lacks evolutionary process • excludes toxins • no explicit energetics
Phenomenological trait-based	• Phenotypic range • Persistent coexistence • Resilience •(Large-scale) patterns	• can include ecological trade-offs • no *a priori* assumptions about species interactions • no *a priori* description of functional groups required • can model arbitrarily many taxa and traits • can model temporal and/or spatial heterogeneity • robustness predictions community dynamics and function (Lajoie & Kembel, 2019) • can include kinetics and diffusion	• needs *a priori* knowledge of: relevant environmental variables, trait assignment, trade-offs between traits • lacks complete evolutionary process • no explicit energetics • lacks inclusion of cross-feeding • assumes no interactions between the resources or abiotic factors themselves. • underrepresentation of species interactions (other than predation) • can be computationally expensive
Individual-based	• Spatial patterns & structures • Resilience to (local) perturbations • Persistent co-existence & Community diversity • Complexity of phenotypes in community	• can model individual cells → increased resolution • cells can respond to changes in local environment • explicit inclusion of physical (thermodynamics), chemical and mechanical laws possible • explicit inclusion of naturally stochastic processes (e.g., cell-cell interactions) is possible • inclusion of metabolites and other compounds possible & these resources can be made to interact • higher-order interactions possible • growth and death rates can be made cell-specific • heterogeneity in (local) environments • can include biochemical detail • can predict concentration gradients of hard-to-detect metabolites/toxins/signals • modular: permitting the addition of further biological and physical processes (Jayathilake et al., 2017)	• high computational requirements • robustness results unclear • parameters not intuitive • large parameter space & prone to overfitting • equilibrium conditions and their stability cannot be attained analytically • *a priori* assumptions cell 'rules' • no universality to methods in Individual-based model