Online Supplement: ODD protocol for describing individual-based models (IBM) and supplemental results

Kenneth J. Locey, Jay T. Lennon

Contents

ODD Protocol

Platform description

Here, we describe our individual-based modeling (IBM) platform according to the ODD protocol (Overview, Design concepts, Details), a community standard for describing IBMs (Grimm et al. 2006).

Our IBM platform allows users to study how patterns from multiple ecological theories, disciplines, or paradigms can simultaneously emerge in the same system. These paradigms include metabolic scaling and physiological ecology, community ecology and biogeography, life history, neutral theory, resource limitation, and population ecology. The source code for this platform can be found here: https://github.com/LennonLab/residence-time.

This platform accomplishes three proximate objectives:

1. Assembles and runs IBMs from random combinations of system variables and species traits.
2. Stores the outputs of IBMs in .csv files.
3. Provides code for analyzing simulation data and generating figures.

Purpose

The goal of our platform is to test the predictions of a τ-based theory for biodiversity and to ask whether these predictions are compatible with widely-studied patterns of biodiversity.

Entities & their state variables

Individual organisms — Individuals are distinguished by collections of elements within dictionaries, i.e., data objects that hold key-value pairs. For example, the dictionary holding information on individual organisms (iDict) is structured as follows:

```
iDict = {ind1: sp = sp1, x  = 1.2,  sz = 81.2, q  = 2.8, st = a, age = 10, tis = 8;
         ind2: sp = sp3, x  = 8.5,  sz = 5.2, q  = 10.4, st = a, age = 23, tis = 1,
         ...}
```

where, for example, "ind1" is the key and the variables following the colon are the values for the species ID (sp), position along the dimension of the environment against which flow occurs (x), body size (sz), amount of endogenous resources (i.e., resources within the individual that are available to fuel life processes (q), metabolic state (active or dormant) (st), and the amount of time that the individual has been in its current metabolic state (tis). Individuals undergo changes when randomly sampled from this dictionary.

Species — Each species is characterized by the individuals that share a common set of traits. Species information is stored in a dictionary (sDict) as key-value pairs.

```
sDict = {sp1: gr = 0.8, di  = 0.5, rp  = 0.3, mt = 0.2, mf  = 0.1, ef = [0.1, ..., 0.5];
         sp3: gr = 0.8, di  = 0.5, rp  = 0.3, mt = 0.2, mf  = 0.1, ef = [0.2, ..., 0.0];
         ...}
```

where, for example, species 1 (sp1) has a maximum intrinsic individual-based growth rate (gr) of 0.8, a maximum active dispersal rate (di) of 0.5 units per time step, a 0.3 probability of randomly resuscitating (rp) from a dormant state, a mass specific active metabolic rate (mt) of 0.2, experiences a 10% decrease in mt when entering dormancy, and has 10 resource-specific use efficiencies (ef) that sum to 1. When sampling individuals, the information about their species is gained by accessing sDict.

Resource particles — Our modeling simulates the movement and consumption of individual resource particles. These particles can vary over two orders of magnitude in size and belong to as many as 10 resource types. As

with individual organisms and species, information about individual resource particles is stored in a dictionary (rDict).

```
rDict = {r1: type = 1, v  = 0.5, x  = 0.3;
         r2: type = 10, v  = 1.3, x  = 109.8;
      ...}
```

In the rDict dictionary, r1 is the key for a single resource particle and the variables following the colon are the values for the resource type (type), the size of the particle (v), and the position along the dimension of the environment against which flow occurs (x).

System level state variables

Table S1. Each model was initiated with a set of core parameters, values for which were chosen at random.

Ecosystem properties

Property	Description	Value
Size (V)	Length of the environment	1 - 1000
Flow rate (Q)	Units of V moved per time step	0.001 - 1
Inflowing resource diversity	Number of inflowing resource types	10
Inflowing resource particle size	Range of sizes for inflowing resource particles	1 - 1000

Species traits

Trait	Description	Value
Individual growth rate (g)	Proportional increase in individual size per time step	0.001 - 1
Active dispersal rate (d)	Units of V traveled against direction of flow per time step	0.001 - 1
Resuscitation rate	Probability of resuscitating per time step	0.001 - 1
Basal metabolic rate (B)	Portion of endogenous resources lost to maintenance respiration per time step	0.001 - 1
Reduction of B	Proportional decrease of B upon entering dormancy	0.001 - 1
Resource efficiencies ($e_1, ..., e_{10}$)	For each of 10 resource types, the proportion of a consumed resource particle that is assimilated into an individual's resource quota	0 - 1

Varying properties of individuals

Property	Description	Value
Resource quota (q_i)	Amount of endogenous resources	0 - unconstrained
Body size (M_i)	Individual biomass (does not include q_i)	0 - unconstrained
Spatial Location (x_i)	Position along V	0 - V
Metabolic state	Whether an individual is metabolically active or dormant	active or dormant

Process overview and scheduling

Model assembly — The user runs the main python program (i.e., main.py). The main program chooses random values for system-level state variables. The main program also imports modules (i.e., groups of functions) for diversity analysis and the initiation of output files.

Randomized model parameterization – Each IBM begins with random combinations of traits for each species in the regional pool, randomly drawn values of V and Q (Table 2). V was randomly chosen within a range of three orders of magnitude and represented the length of a system across which flow occurred. Q was also randomly chosen within a range of three orders of magnitude and represented the fraction of a single unit distance covered per time step (Table 2). These randomized starting conditions allow predictions to emerge without explicitly enforced relationships between species traits (see also Locey and Lennon 2017; Locey et al. 2017).

Immigration and resource supply – Each IBM began as an empty system into which resource particles flowed and individual organisms immigrated. Species identities were drawn at random from uniform distributions, as were size and identities of resource particles. Because the equilibrium density of particles in a system should not increase with Q (i.e., the faster particles enter, the faster they leave) the probability of an immigrant or resource particle entering the system was multiplied by Q (0.001 to 1.0).

Overview of life history processes — Our platform simulates individual-based processes of consumption, growth, reproduction, death, passive and active dispersal, and transitions into and out of dormancy. The ordering of life history processes is randomized at each time step to prevent scheduling bias. The rates at which individuals grew, consumed resources, and actively dispersed and whether individuals reproduced, died, or transitioned into or out of dormancy were determined by the products of independent probabilities. These probabilities were influenced by species trait values, along with the metabolic state, body size, and the endogenous resources (q) of individuals. Growth, active dispersal, and transitions to metabolic activity also incurred energetic costs, allowing a biologically meaningful means of simulating death while allowing environmental filtering to act individual variation in traits, size, and q.

Consumption – At each time step, randomly sampled individuals could consume a randomly encountered resource particle from any of 10 possible resource types. Because encounters with any of R resource particles should be influenced by the concentration of resource particles in the environment ($D = R/V$), the probability of consumption (p) was determined by a simple function that allowed D to take any non-negative real number while corresponding to a probability that increased with D and that equaled 0.0 in the absence of resources:

$$p = D/(1 + D)$$

Consumption of a resource particle increased an individual's level of endogenous resources (q). The rate of consumption was determined by the individuals' species-specific rate of consumption for the given resource type (e) and the size of the resource particle (r_0):

$$q_1 = q_0 + min(r_0, e * r_0 * size)$$

The size of the resource particle was then decreased appropriately and led resource particles of zero size to be removed from the system:

$$r_1 = r_0 - min(r_0, e_i * r_0 * size)$$

Growth – Randomly sampled individuals grew proportionally to their q and their species-specific rate of relative individual growth μ, resulting in an equivalent decrease in q:

$$size_1 = size_0 + min(q, \mu * q)$$
$$q_1 = q_0 - min(q_0, \mu * q_0)$$

These simple recurrence relations produce proportional ontogenetic growth limited by q. They also do not allow individuals to grow in the absence of endogenous resources and enforce a conservation of mass between size, q, and r.

Reproduction – At each time step, randomly sampled individuals that were metabolically active could reproduce according to the product of three probabilities, each determined by a simple function that allowed values for the individual's endogenous resources (q) and size to take any non-negative real number while corresponding to probabilities that increased with size and q. In this way, the probability of reproducing equaled 0 if q or size equaled 0.

$$p = (q/(1 + q)) * (size/(1 + size))$$

In this way, greater size and greater endogenous resources contributed to whether individuals reproduced. Reproduction was clonal, as in other models of biodiversity (e.g., Hubbell 2001) and resulted in the halving of q and size. Daughter individuals inherited the spatial coordinates of their mothers, were not subject to mutation, and were given an age of 0.

Transitions between activity and dormancy – Active individuals became dormant based on the product of probabilities determined by their age and q:

$$p = (1/(1 + q)) * (age/(1 + age))$$

As with reproduction, each probability was determined by a saturating function that allowed q, and body size to take any positive real number. In this way, a greater amount of endogenous resource decreased the probability of going dormant, while the age related probability prevented newly produced daughter individuals from going dormant because of their initially small q.

Dormant individuals could transition back to activity according to a probability determined by their species-specific rate of random resuscitation (α) and individual values of q and body size.

$$p = (q/(1 + q)) * \alpha$$

In this way, a low q and small body size reduced the probability of resuscitation. Hence, dormant individuals could eventually become non-viable. Transitions to activity incurred a species-specific energetic cost proportional to α and their size, meaning that larger individuals incurred greater energetic costs when resuscitating.

$$q_1 = q_0 - \alpha * size$$

Metabolic Maintenance and Death – At each time step, randomly sampled individuals incurred a species-specific cost of basal metabolism (BMR) that was likewise, proportional to their size. If the individual was dormant, BMR was reduced by a species-specific value (γ). Hence, dormant individuals incurred metabolic costs that were lower than the cost of active BMR, and the value of γ represented the strength of that reduction. The cost of basal metabolism was paid through the reduction of q:

$$active : q_1 = q_0 - BMR * size$$
$$dormant : q_1 = q_0 - \gamma * BMR * size$$

Individuals were simulated as having died when they were unable to pay the cost of basal metabolism, i.e., q < 0. Hence, both active and dormant individuals could die of starvation.

Dispersal – At each time step, individuals were passively moved according to the rate of flow (Q). However, randomly sampled active individuals could disperse against the direction of flow according to their species-specific dispersal rate (δ).

$$x_1 = x_0 - min(x_0, \delta * size)$$
$$q_1 = q_0 - min(q_0, \delta * size)$$

In this way, individuals could not disperse beyond the upstream edge of the system and also incurred an energetic cost that was proportional to the individuals' size and the distance moved against the direction of Q.

Design concepts

Basic principles

Environmental filtering on species traits: Our IBM platform operates according to a basic principle of ecology, that is, the selection via the abiotic environment on species traits.

Energy-limited life history: Our IBM platform imposes energetic costs to growth and activity. These energetic costs are directly proportional to the values of life history traits. For example, the energetic cost of dispersal is the product of dispersal rate and dispersal distance. This intuitively means that the energetic cost of dispersal is multiplied across distance such that moving a distance of d in a specific direction requires half the energy as moving a distance of $2d$ in the same direction.

Lognormal dynamics: Multiplicative interactions of random variables underpin one of the most successful models of complex systems, i.e., the lognormal (Crow et al. 1988). The lognormal was introduced to ecologists by Preston (1962) as a description of how abundance varies among species. The lognormal has been one of the two most successful models of species abundance for macroscopic plants and animals, and was recently used to form a macroecological theory of microbial diversity (Shoemaker et al. 2017). By "multiplicative interactions" one simply means that two or more variables interact in a multiplicative way, such as the product of body size and the rate of proportional growth or the product of dispersal distance and energetic cost per unit distance. Such interactions are common in ecology and underpin most models of population growth (Putnam 1993). By "random variables" one simply means two independent variables that are either chosen at random or that change stochastically. Our platform explicitly simulates these lognormal dynamics; for example, in the simulation of stochastic proportional growth, active dispersal, and in the simulation of energetic costs.

Simultaneous emergence: A popular advantage of IBMs is the potential to study the emergence of complex patterns from individual-level interactions. Our platform allows emergence in three ways that IBMs are rarely employed. First, patterns emerge from random combinations of traits and state variables. This allows realistic trait syndromes and patterns to emerge without being explicitly enforced. Second, patterns emerge across orders of magnitude in random variation in starting conditions of a large ensemble of IBMs. These degrees of variation can be changed in 'main.py' file and in the 'immigration' function located in the 'bide.py' file. This highly unconstrained approach allows the user to explore a broad spectrum of trait combinations. Third, our platform allows for the study of simultaneous emergence, that is, the emergence of several patterns at once. In this way, our platform initiates with unrealistic ecological communities and unrealistic combinations of species traits, and then allows for ecological realism to develop over time in response to environmental conditions and according to energetic costs. This approach allows the user of our platform to avoid one of the greatest challenges to ecological modeling, i.e., the circularity of documenting outcomes that are otherwise forced to occur by explicitly enforced relationships and arbitrarily chosen parameter values. The code made available for analysis of simulation data is intended to examine the variation and central limiting behavior of ecological patterns among thousands of IBMs.

Other design concepts of the ODD protocol

Hypotheses — Our platform was designed to test predictions of our residence-time theory and to show that these predictions are compatible with well-known patterns of biodiversity.

Learning — There is no aspect of individual-based learning in Emergence. An example of learning would be the adaptation of individuals to local conditions by learning strategies of foraging, competition, etc.

Prediction — Individuals in Emergence do not presently have the ability to anticipate changes in environmental conditions.

Sensing — Individuals in our IBMs cannot currently sense the environment. That is, they do not move towards resource particles or sense the presence of resources or signals that might otherwise induce resuscitation from dormancy (e.g., quorum sensing).

Interaction — Individuals interact through excluding each other from resources, i.e., preemption. There is no explicit communication.

Observation — An unlimited number of IBMs can be run to examine trends and variation in ecological patterns from the recording of 119 metrics and additional aspects of community structure. See 'Output data' below.

Simulation data files

Our IBM platform generates three files of output data. Each IBM quantifies and writes output data for every nth time step after "burn in", where n and the number of time steps given for the model to assemble a community (the "burn in") can be designated by the user. The three files are:

SimData.csv

SimData.csv — Each of the 119 columns of this file corresponds to a piece of data about the system that was modeled (e.g., flow rate, total abundance, species richness, species turnover, resource supply and diversity, rate of disturbance, etc.). Most analyses are conducted on the data in this file. The following is recorded for each Emergence model and stored in SimData.csv:

Table S2. Output of simulation-based data from ensembles of IBMs. These data are found in SimData.csv.

Datum (column header)	Description
sim	Unique identifier for the nth simulation
time	computer clock time
ct	m^{th} time step of the nth simulation
immigration.rate	Probability of an individual immigrating per time step
inflowing.res.dens	Number of resource particles inflowing per time step
N.types	Number of resource types
max.res.val	Greatest size of any inflowing resource particle
max.growth.rate	Maximum intrinsic growth rate for any species
max.met.maint	Maximum fraction of individual endogenous resources lost to metabolic maintanence per time step. A species-specific parameter.
max.active.dispersal	Maximum amount of distance traveled by any individual againt the direction of flow. A species-specific parameter.
starting.seed	Number of individuals that initially occupy the system. Default = 1.
V	Length of the environment
flow.rate	Units of V moved per time step

Continued on next page

Table 1 – *Continued from previous page*

Datum (column header)	Description
total.abundance	Number of individual organisms in the system.
active.total.abundance	Number of active individual organisms in the system.
dormant.total.abundance	Number of dormant individual organisms in the system.
ind.production	Number of individual organisms produced in the mth time step.
resource.particles	Number of resource particles in the system.
resource.concentration	Number of resource particles for every unit of V.
resource.richness	Number of resource types present.
species.richness	Number of species present.
simpson.e	Simpson's measure of species evenness. Equitability of species abundances.
avg.pop.size	Mean population size among species.
pop.var	Variance in population size among species.
N.max	Abundance of the most abundant species.
logmod.skew	Log-modulo skewness of the frequency distribution of species abundances. A measure of species rarity.
whittakers.turnover	Whittaker's measure of species turnover between two time points, quantifying the percent turnover in species membership. Does not consider species abundances.
total.biomass	Mean biomass of individual organisms (i.e., size + endogenous resources).
avg.per.capita.growth	Mean per capita rate of intrinsic individual growth.
avg.per.capita.maint	Mean per capita rate of mass-specific active basal metabolic rate (BMR). Portion of individual endogenous resources lost per unit size, per time step.
avg.per.capita.efficiency	Mean per capita variance in resource efficiency. Resource efficiencies are species-specific.
avg.per.capita.active.dispersal	Mean per capita rate of active dispersal, i.e., portion of a unit of V that individuals can move against the direction of flow. Active dispersal rates are species-specific.
avg.per.capita.rpf	Mean per capita probability of randomly resuscitating from a metabolically dormant state. A species-specific trait.
avg.per.capita.mf	Mean per capita percent by which active BMR is decreased when individuals enter a metabolically dormant state. A species-specific trait.
avg.per.capita.size	Mean per capita body size. Does not include individuals' endogenous resources.
active.species.richness	Number of species in the metabolically active portion of the community.
active.simpson.e	Simpson's measure of species evenness within the metabolically active portion of the community.
active.avg.pop.size	Mean population size among species within the metabolically active portion of the community.
active.pop.var	Variance in population size among species within the metabolically active portion of the community.

Continued on next page

Table 1 – *Continued from previous page*

Datum (column header)	Description
active.N.max	Abundance of the most abundant species within the metabolically active portion of the community.
active.logmod.skew	Log-modulo skewness of the frequency distribution of species abundances within the metabolically active portion of the community. A measure of species rarity.
active.whittakers.turnover	Whittaker's measure of species turnover between two time points, quantifying the percent turnover in species membership within the metabolically active portion of the community. Does not consider species abundances.
active.total.biomass	Mean biomass of individual organisms (i.e., size + endogenous resources) within the metabolically active portion of the community.
active.avg.per.capita.growth	Mean per capita rate of intrinsic individual growth within the metabolically active portion of the community.
active.avg.per.capita.maint	Mean per capita rate of mass-specific active basal metabolic rate (BMR) within the metabolically active portion of the community. Portion of individual endogenous resources lost per unit size, per time step.
active.avg.per.capita.efficiency	Mean per capita variance in resource efficiency within the metabolically active portion of the community. Resource efficiencies are species-specific.
active.avg.per.capita.dispersal	Mean per capita rate of active dispersal within the metabolically active portion of the community, i.e., portion of a unit of V that active individuals can move against the direction of flow.
active.avg.per.capita.rpf	Mean per capita probability of randomly resuscitating from a metabolically dormant state for species within the metabolically active portion of the community. A species-specific trait.
active.avg.per.capita.mf	Mean per capita percent by which active BMR is decreased when individuals enter a metabolically dormant state. With respect to species within the metabolically active portion of the community. A species-specific trait.
active.avg.per.capita.size	Mean per capita body size within the metabolically active portion of the community. Does not include individuals' endogenous resources.
dormant.species.richness	Number of species in the dormant portion of the community.
dormant.simpson.e	Simpson's measure of species evenness within the dormant portion of the community.
dormant.avg.pop.size	Mean population size among species within the dormant portion of the community.
dormant.pop.var	Variance in population size among species within the dormant portion of the community.
dormant.N.max	Abundance of the most abundant species within the dormant portion of the community.
dormant.logmod.skew	Log-modulo skewness of the frequency distribution of species abundances within the dormant portion of the community. A measure of species rarity.
dormant.whittakers.turnover	Whittaker's measure of species turnover between two time points, quantifying the percent turnover in species membership within the dormant portion of the community. Does not consider species abundances.

Table 1 – *Continued from previous page*

Datum (column header)	Description
dormant.total.biomass	Mean biomass of individual organisms (i.e., size + endogenous resources) within the dormant portion of the community.
dormant.avg.per.capita.growth	Mean per capita rate of intrinsic individual growth within the dormant portion of the community.
dormant.avg.per.capita.maint	Mean per capita rate of mass-specific active basal metabolic rate (BMR) within the dormant portion of the community. Portion of individual endogenous resources lost per unit size, per time step.
dormant.avg.per.capita.efficiency	Mean per capita variance in resource efficiency within the dormant portion of the community. Resource efficiencies are species-specific.
dormant.avg.per.capita.dispersal	Mean per capita rate of active dispersal within the dormant portion of the community, i.e., portion of a unit of V that active individuals can move against the direction of flow.
dormant.avg.per.capita.rpf	Mean per capita probability of randomly resuscitating from a metabolically dormant state for species within the dormant portion of the community. A species-specific trait.
dormant.avg.per.capita.mf	Mean per capita percent by which active BMR is decreased when individuals enter a metabolically dormant state. With respect to species within the dormant portion of the community. A species-specific trait.
dormant.avg.per.capita.size	Mean per capita body size within the dormant portion of the community. Does not include individuals' endogenous resources.
percent.dormant	Percent of individuals in the model that are in a metabolically inactive state of zero growth, zero active dispersal, zero reproduction, and decreased metabolic rate.
inflowing.res.rich	Number of resource types that can flow into the nth model.
total.res	Sum of all resource particle sizes in the model.
efficiency	Mean variance in resource efficiency among species. Resource efficiencies are species-specific.
growth	Mean rate of intrinsic individual growth among species.
maint	Mean rate of mass-specific active basal metabolic rate (BMR) among species. The portion of endogenous resources lost per unit mass of an active individual, per unit time.
mf	Mean percent by which BMR is decreased when individuals transition to a dormant state.
rpf	Mean species-specific probability of dormant individuals randomly resuscitating to a metabolically active state.
dispersal	Mean fraction of a unit of V that individuals can move against the direction of flow.
active.efficiency	Mean variance in resource efficiency among species within the metabolically active portion of the community. Resource efficiencies are species-specific.
active.growth	Mean rate of intrinsic individual growth among species within the metabolically active portion of the community.
active.maint	Mean rate of mass-specific active basal metabolic rate (BMR) among species within the metabolically active portion of the community. The portion of endogenous resources lost per unit mass of an active individual, per unit time.

Continued on next page

Table 1 – *Continued from previous page*

Datum (column header)	Description
active.mf	Mean percent by which BMR is decreased when individuals within the metabolically active portion of the community transition to a dormant state.
active.rpf	Mean species-specific probability of dormant individuals randomly resuscitating to a metabolically active state. Here, with respect to individuals within the metabolically active portion of the community.
active.dispersal	Mean fraction of a unit of V that active individuals can move against the direction of flow. Here, with respect to individuals within the metabolically active portion of the community.
dormant.efficiency	Mean variance in resource efficiency among species within the dormant portion of the community. Resource efficiencies are species-specific.
dormant.growth	Mean rate of intrinsic individual growth among species within the dormant portion of the community.
dormant.mf	Mean percent by which BMR is decreased when active individuals transition to a dormant state. Here, with respect to individuals within the dormant portion of the community.
dormant.maint	Mean rate of mass-specific basal metabolic rate (BMR) among species within the dormant portion of the community. For each species, active BMR is decreased by a species-specific factor. The portion of endogenous resources lost per unit mass of a dormant individual, per unit time.
dormant.rpf	Mean species-specific probability of dormant individuals randomly resuscitating to a metabolically active state. Here, with respect to individuals within the dormant portion of the community.
dormant.dispersal	Mean fraction of a unit of V that active individuals can move against the direction of flow. Here, with respect to individuals within the dormant portion of the community.

RAD-Data.csv

A file holding species abundance data for each species in each IBM, at each of 100 recorded time points. This file is useful for analyzing patterns of community structure such as rank-abundance distributions and for analyzing compositional changes across time.

active.RAD-Data.csv

Similar to the RAD-Data.csv file above, but constrained to only the active abundances of species.

dormant.RAD-Data.csv

Similar to the RAD-Data.csv file above, but constrained to only the dormant abundances of species.

Metrics quantified by the IBM platform

Species evenness – Species evenness was calculated using Simpson's evenness metric, an inverse analog of the sample variance.

Species rarity – Rarity was calculated as log-modulo skewness, a transformation of the skewness of the species abundance frequency distribution, as in Locey and Lennon (2016).

Species richness – Richness was quantified as the total number of species, as in Locey and Lennon (2016).

Species dominance – Dominance was quantified as the abundance of the most abundant species, as in Locey and Lennon (2016).

Species turnover – Turnover was quantified using Whittaker's measure, which quantifies the proportional turnover of species between time points.

Resource specialization – Resource specialization was quantified in three ways using resource use efficiencies. These efficiencies can range between 0 and 1. The first of these measures was the variance in efficiencies, to include those that species could not use. The second measure quantified the variance in efficiencies but only included those that species could actually use. The third measure reflected the inverse of the number of resources a species could use (x): 1/(1+x), such that higher values reflect higher specialization.

General ecological patterns generated by the IBM platform

The IBM source code includes python scripts to analyze the following ecological patterns.

Species abundance distribution (SAD) – The SAD is the vector of species abundance in an ecological community and is one of ecology's fundamental patterns of commonness and rarity. SAD's almost universally reflect that few species in ecological communities are highly abundant while most species are relatively rare (McGill et al. 2007). The SAD is predicted by more than 20 ecological models, the two most successful of which are the Poisson-lognormal and the log-series distributions (White et al. 2012, Baldridge et al. 2015, Shoemaker et al. 2017).

Metabolic scaling – The scaling of basal metabolic rate (B) with body size (M) is one of the strongest patterns of ecology (Brown et al. 2004). Taking the form $B \propto B_0 M^z$, the scaling exponent z is most often observed to approach 3/4. However, this scaling exponent has long been argued to be closer to 2/3 and in some cases can approach 1.0 (Glazier 2006). The metabolic theory of ecology (MTE) predicts that $z = 3/4$ and, from this, predicts several other relationships that scale to quarter powers of body size. These include a -1/4 scaling of mass-specific metabolic rate and a -3/4 scaling of population density.

Diversity-abundance scaling relationships – Aspects of species-level diversity have been shown to scale with the total abundance (N) of individuals in a sample at rates that are often similar for microorganisms and macroscopic plants and animals (Locey and Lennon 2016). These aspects include the following: Richness (S), or the number of species in a community. Rarity, or the concentration of species at low abundances. Absolute dominance (N_{max}), or the greatest species abundance. Evenness (E), or the similarity or equitability of abundance among species.

Taylor's Law – The variance (σ^2) in population abundance often scales with the average population abundance (N/S) across space and through time at a rate between 1 and 2, i.e., $\sigma^2 = (N/S)^{1<z<2}$. Named after Taylor (1961), Taylor's Law is observed in many ecological and non-ecological systems.

Supplemental results

Figure S1: Species abundance distribution (SAD)

Kernel density curves show the distribution of r-square values for fits of maximum likelihood forms of two SAD models: the Poisson lognormal distribution (red) and the log-series distribution (blue). The r-square value reveals the percent of variation in the SADs of IBMs that is explained by the SAD models. As commonly observed in ecological communities, both models often explain >75% variation in abundance among species.
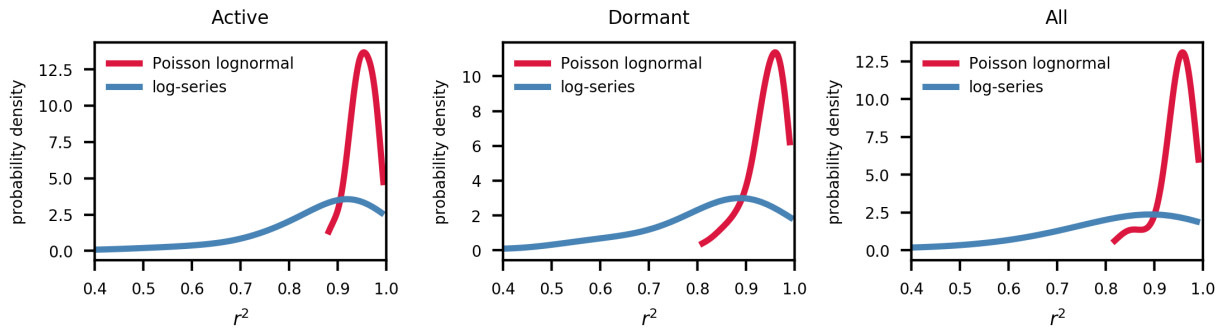
Figure S2: Taylor's Law

Our IBM platform reproduced the scaling of variance in population size with average population size, i.e., Taylor's Law. The scaling exponents for IBMs were well within the observed range of values for Taylor's Law (between 1 and 2). Each data point represents a single species. These results were consistent regardless of whether we focused on species' dormant abundances, active abundances, or total abundances. Rainbow spectrum colors represent species belonging to systems of different τ, with red being shortest and violet being longest.
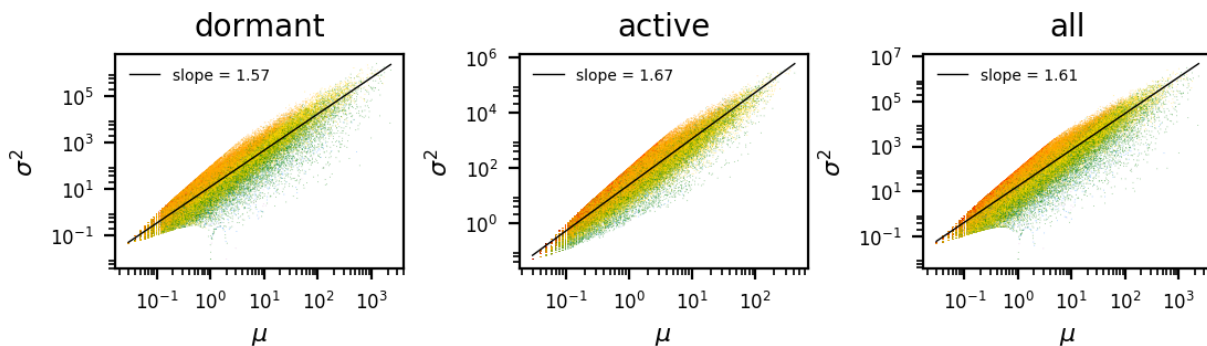
Figure S3: Diversity-abundance relationships

Our IBM platform reproduced the diversity-abundance scaling relationships of Locey and Lennon (2016). Each data point represent the mean for a single model. Rainbow spectrum data points represent systems of different τ, with red being shortest and violet being longest.
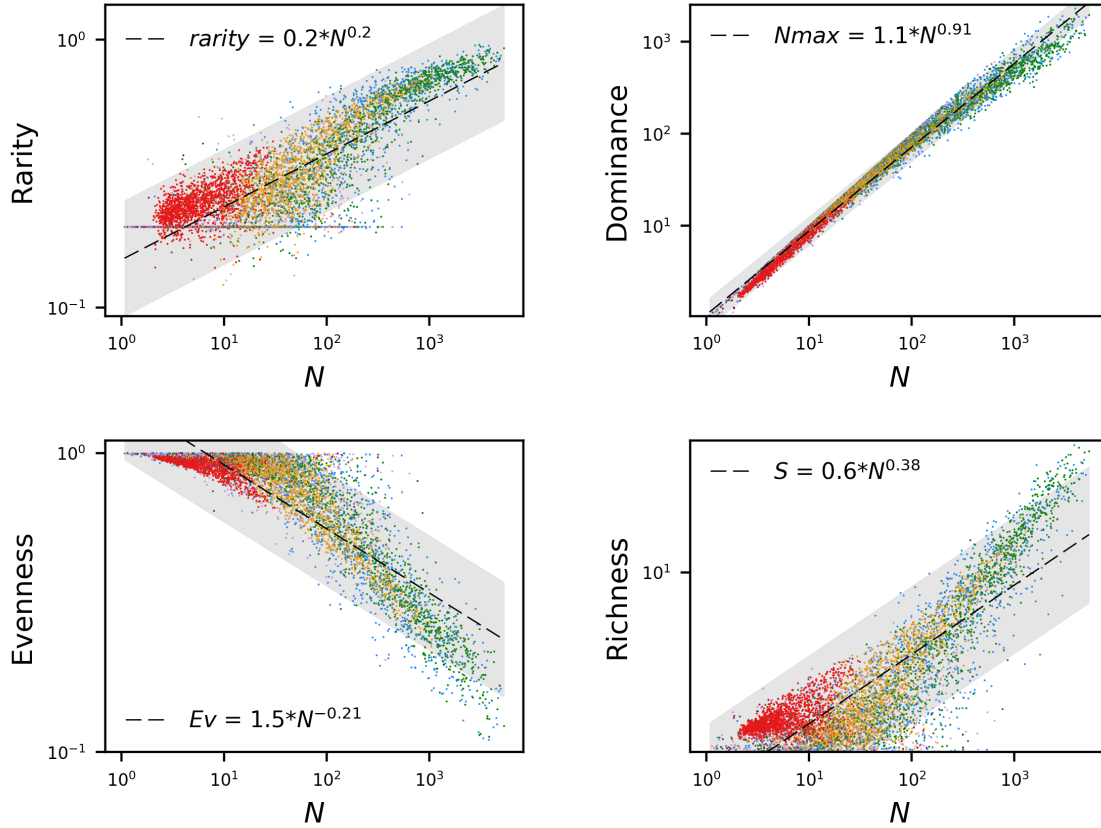
Figure S4: Metabolic scaling

Our IBM platform closely reproduced the 3/4 power scaling of basal metabolic rate to body size for metabolically active individuals. This same scaling does not hold for metabolically dormant individuals, as might be expected. Each data point represent the mean for a single model. Rainbow spectrum data points represent systems of different τ, with red being shortest and violet being longest.
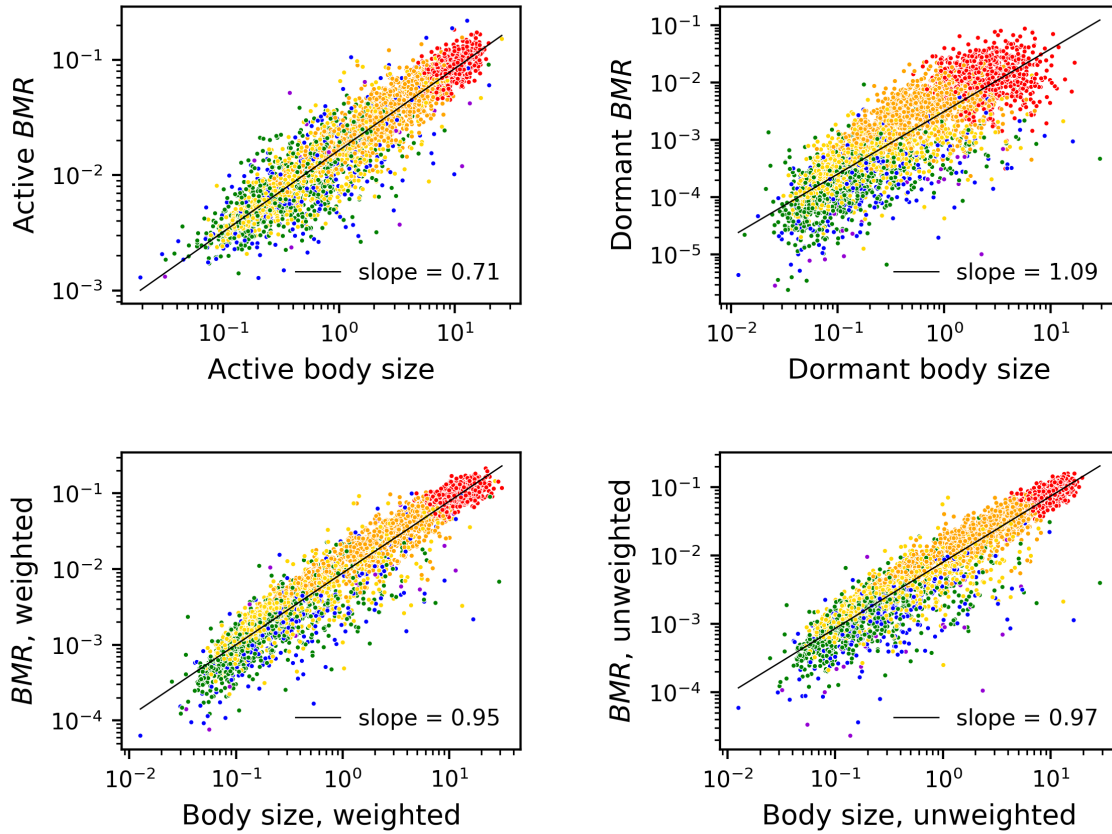
Figure S5: Residence time does not approximate individual age.

In idealized systems such as chemostats, residence time (τ) approximates the average amount of time that passively dispersing individuals remain in a system. This prediction failed in our IBMs, which simulated systems that lack stable-state dynamics and that were characterized by immigration and stochastic dynamics. Each data point represent the mean for a single model.
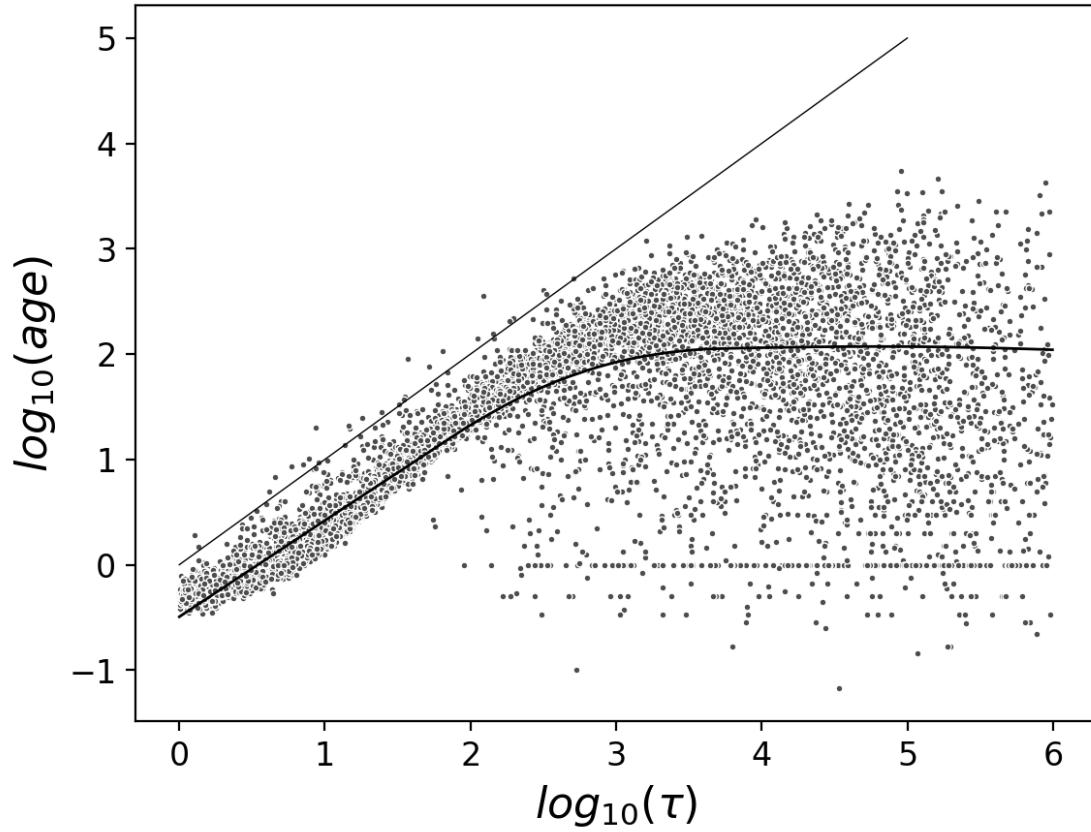
Figure S6: Stochastic population dynamics

Time series of populations from different IBMs. These populations maintained non-zero abundances for the duration of their IBM and are clearly characterized by non-stable state dynamics, stochastic changes, and highly fluctuating population sizes.
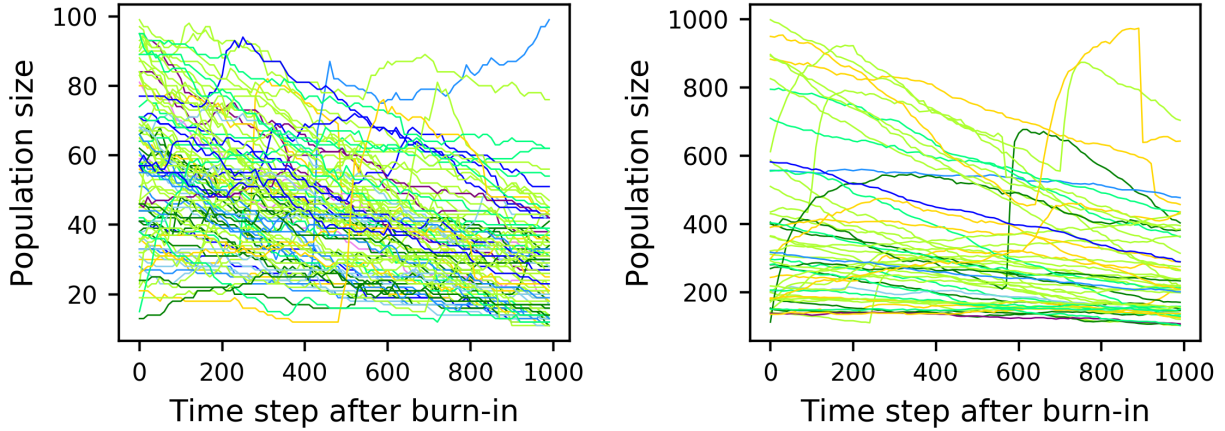

Figure S7: Dilution rate does not approximate growth rate.

In idealized systems such as experimental chemostats, diluation rate $(1/\tau)$ approximates specific growth rate (μ), that is, the proportional rate of population growth. However, this expected relationship could be expected to fail in systems that lack stable-state dynamics and that are characterized by stochastic population and community dynamics, as in our IBMs. Each data point represent the mean for a single model.
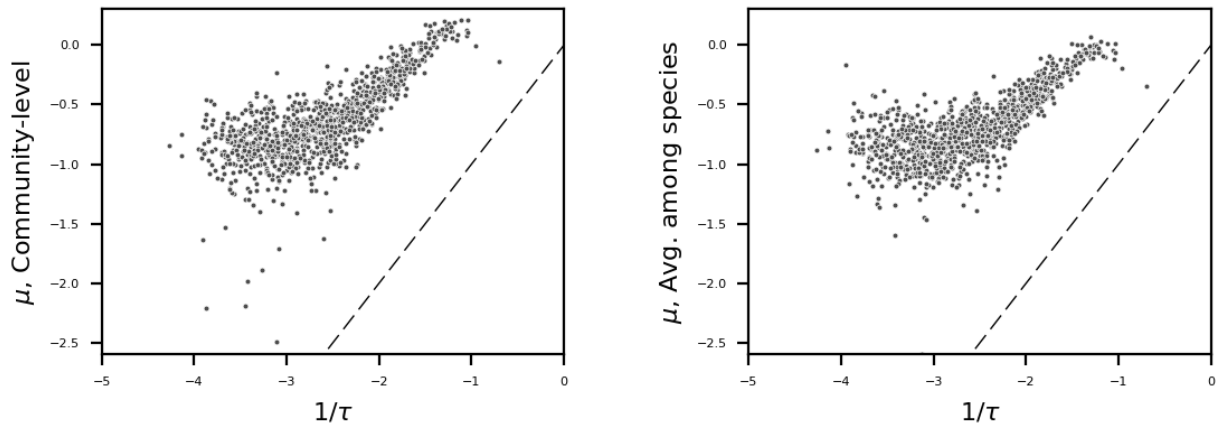
Figure S8: Opposing influence of flow rate and system size

According the definition of residence time as the ratio of a system's size V to its rate of flow Q, V and Q have opposing influences on the value of residence time. Hence, V and Q should have opposing influences on variables that respond to residence time. Each data point represent the mean for a single model. Rainbow spectrum data points represent systems of different V (left column) or different Q (right column), with red representing the smallest and violet representing the largest.
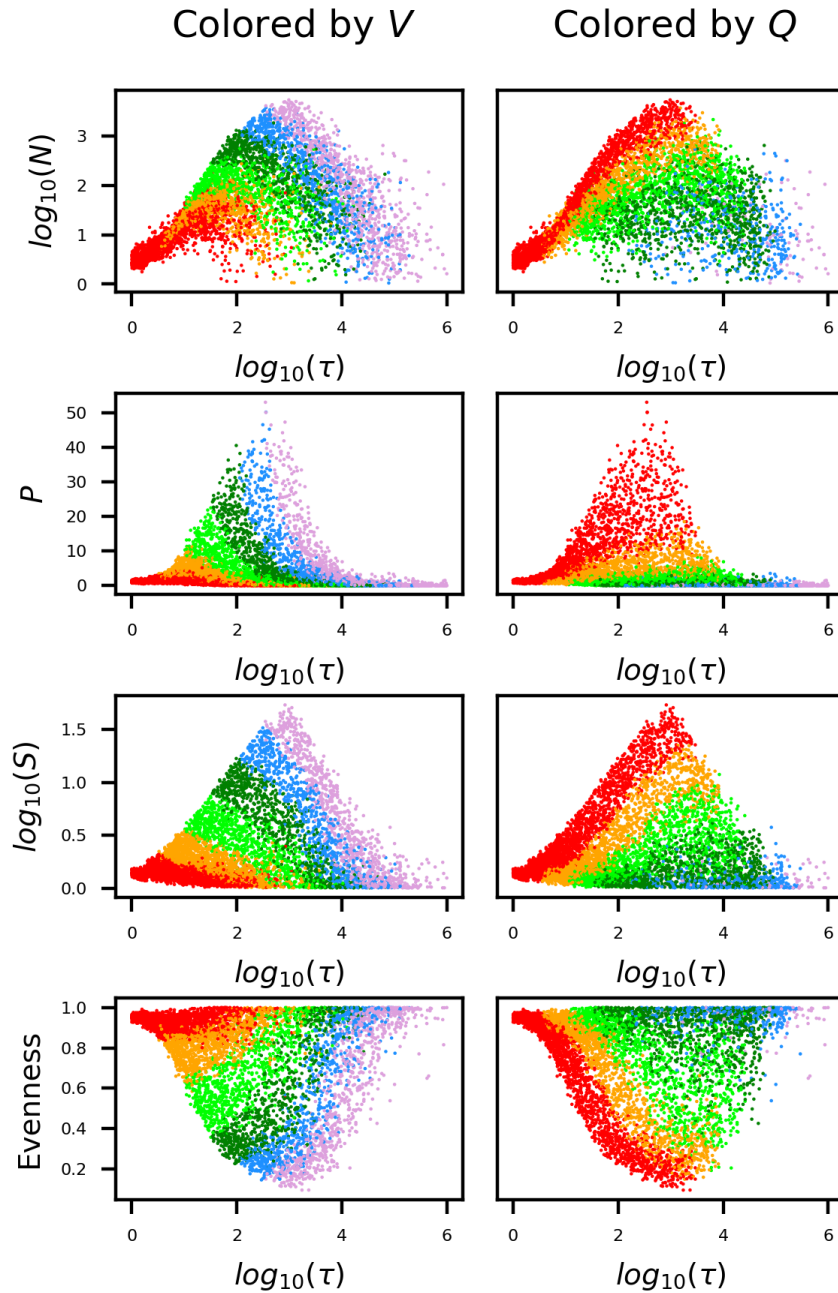
References

1. Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., and G. B. West. 2004. Toward a metabolic theory of ecology. Ecology, 85:1771-1789.

2. Crow, E. L., & K. Shimizu. 1988. Lognormal Distributions: Theory and Applications. New York: Markel Dekker.

3. Glazier, D. S. 2006. The 3/4-power law is not universal: evolution of isometric, ontogenetic metabolic scaling in pelagic animals. AIBS Bulletin, 56:325-332.

4. Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G., and A. Huth. 2006. A standard protocol for describing individual-based and agent-based models. Ecological Modelling, 198:115-126.

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

6. Locey, K. J., and J. T., Lennon. 2016. Scaling laws predict global microbial diversity. Proceedings of the National Academy of Sciences of the United States of America, 113:5970-5975.

7. Locey K.J., and J. T. Lennon. 2017. A modeling platform for the simultaneous emergence of ecological patterns. PeerJ Preprints 5:e1469v3

8. Locey, K. J., Fisk, M. C., and J. T. Lennon. 2017. Microscale Insight into Microbial Seed Banks. Frontiers in Microbiology, 7:2040.

9. McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., ... and A. H. Hurlbert. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology letters, 10: 995-1015.

10. Preston, F. W. 1962. The canonical distribution of commonness and rarity: Part I. Ecology, 43:185-215.

11. Putnam, R. Community Ecology. 1993. Chapman & Hall, London, United Kingdom.

12. Shoemaker, W. R., Locey, K. J., & Lennon, J. T. 2017. A macroecological theory of microbial biodiversity. Nature Ecology & Evolution, 1:0107.

13. Taylor, L. R. 1961. Aggregation, variance and the mean. Nature. 189:4766 732–735.