Imperial College London

Imperial College London

Department of Life Sciences

Is everything everywhere? Deterministic and stochastic processes in microbial biogeography

Amy Solman

External supervisor:
Prof. Ryan Chisholm
University of Singapore

Internal supervisors:
Dr. James Rosindell
Imperial College London
Prof. Thomas Bell
Imperial College London

August 2020

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Research at Imperial College London

Formatted in the journal style of Methods in Ecology and Evolution
Submitted for the MRes in Computational Methods in Ecology and Evolution

Declaration

Concept: The concept for this work came from Prof. Thomas Bell and Prof. Ryan Chisholm.

Data: All data for this analysis was collected by myself from the literature as listed in the Supplementary Materials section.

Simulation: All simulation code was written by myself, excluding Function 2: coalescence_test provided by Dr. James Rosindell.

Model: The Classic, Depth and Perimeter models, as well as critical area formulas, was supplied by Prof. Ryan Chisholm.

Analysis: I declare that all analysis was carried out by myself.

Report: I declare that the report was written by myself.

COVID-19

The hypotheses presented in this report were originally investigated using a laboratory-based experiment in Professor Thomas Bell's Microbial Ecology Laboratory, Imperial College London, Silwood Park. Two months into the investigation, laboratory work was ceased due to the COVID-19 pandemic. The following report seeks to test the same hypotheses, using data from the literature.

Abstract

1. Microbial communities play an essential role in biogeochemical processes and ecosystem functioning. Despite their importance, relatively little is known about the mechanisms driving spatial scaling within these communities.

2. The Theory of Island Biogeography predicts that species richness increases with island area through stochastic colonisation-extinction processes. It has been widely use to describe macro-organism spatial patterns, but there has been little application to microbial communities.

3. Despite the popularity of this theory, small islands often show no clear relationship between species richness and area. Chisholm *et al.,* have addressed this by developing a unified theory of species-area relationships that transitions from a niche-structured regime at smaller spatial scales to stochastic mechanisms at larger spatial scales.

4. This work modifies the Chisholm model to address microbial communities in a range of habitats. Through model fitting I assess to what extent microorganisms are subject to deterministic or stochastic processes. I also explore how the transition between regimes differs among habitat types and taxonomic groups.

5. The model gave good fits to the data. Multiple regression analysis showed significant evidence that taxonomic group explained variance in the critical area of transition between deterministic and stochastic regimes. Less motile taxonomic groups exhibited higher critical areas of transition, supporting my hypothesis. Habitat type was non-significant in predicting critical area.

6. A proportion of the datasets exhibited a biphasic species-area relationship. Broadly the critical area hypotheses of lower transition areas for more motile taxa was supported by the data.These results can assist in predicting the spatial scaling of microbial diversity, with application to climate change modelling.

Keywords: species-area relationships, taxa-area relationships, microbial biogeography, small-island effect, niche structure, colonisation-extinction balance, island biogeography

Contents

List of Figures

List of Tables

Introduction

The species–area relationship (SAR) is one of the oldest fundamental ecological laws (Gooriah and Chase, 2019). SARs describe the relationship between community diversity and habitat area. The observation that species richness increases with sampling area, a positive SAR, has been observed for a broad range of faunal (Ricklefs and Lovette, 1999) (Lomolino, 1982) (Eadie et al., 1986) and floral groups (Zacharias and Brandes, 1990) (Price and Wagner, 2011). The ubiquitous nature of positive SARs has been used to inform conservation practises in natural (Haila, 2002) (Samson, 1980) and urban environments (Davis and Glick, 1978). Whilst a large number of studies have examined macroorganism SARs, relatively little is known about the spatial scaling of microbial biodiversity.

Understanding the factors that regulate microbial community structure is important as they play a vital role in biogeochemical cycling and ecosystem functioning (Griffiths et al., 2011). Despite bacteria and fungi representing major contributors to soil biodiversity and processes, little is known about below-ground regulators of biodiversity (Griffiths et al., 2011) (Li et al., 2020). This is especially challenging as few terrestrial environments present insular habitats for microbial community dynamics to be easily studied. Microorganisms also play a metabolically active role in polar regions previously believed to be abiotic (Stibal et al., 2020). Rapid climate change is leading to the exposure of soils dominant in high-latitude carbon (Bradley et al., 2017). As these soils are colonised by microbial communities, biogeochemical transformations release CO_2, CH_4 and N_2O. Understanding the mechanisms that drive microbial colonisation of polar environments can help produce accurate models of greenhouse gas release (Malard and Pearce, 2018).

Debate around the applicability of SARs to microbial systems stems from the assumption that they are limited only by niche-filtering, as articulated by Bass-Becking: '*Everything is everywhere*, but, *the environment selects*' (Baas-Becking, 1934). This classic tenet of mi-

crobiology assumes that the abundance, short generation times and small size of microorganisms gives them an almost cosmopolitan distribution (Green and Bohannan, 2006). High abundances increase the probability of transport between environments via an accidental vector. Small size also increases the likelihood of passive transport via air or water, leading to high dispersal rates (Green and Bohannan, 2006). Uninhibited dispersal may also be facilitated by dormancy as a biogeographical response (Locey, 2010).

One of the most commonly used tools in biogeography is the power-law model:

$$S = cA^z \tag{1.1}$$

Where S is species richness as a function of area (A), c is a constant specific to that taxa/habitat and the z exponent is the slope of the line associating area and species richness (Darcy et al., 2018). z typically falls in the range of 0.1 to 0.3 for continuous habitats and 0.25 to 0.35 for insular habitats (Green and Bohannan, 2006). Microbial z values are typically well below those seen in macro-organisms ($z < 0.1$), supporting the idea of cosmopolitan distribution (Green and Bohannan, 2006).

One of the limitations for microbial biogeography has been in quantifying taxa, given that many cannot be accurately identified using morphological techniques (Green and Bohannan, 2006). Commonly, microbial biogeography is concerned with taxa-area relationships (TARs), rather than SARs as microbial diversity is quantified in operational taxonomic groups (OTUs). Deciduous leaves as 'island' habitats for aquatic fungi found morphospecies-based diversity increased with leaf area, but next generation sequencing did not. (Duarte et al., 2017). With recent advances in molecular approaches such as single-celled sequencing, the genomes of previously uncultivated bacterial taxa are filling in the phylogenetic tree providing a higher resolution picture of microbial community structure (Lasken and McLean, 2014). Limited data on temporal and spatial microbial distributions has led to a lack of detailed distribution maps. Distribution maps allow us to estimate the true number of taxa in a given environment when total counts are not available, without which estimated z values may be artificially low (Green and Bohannan, 2006).

The mechanisms driving island SARs have been of particular interest to ecologists since the

1800s (MacDonald et al., 2018). Islands are considered important paradigms for fragmented habitats as well as larger geographic regions (Simberloff, 1974). Their insular nature allows for ecological processes and patterns to be investigated in a simplified and relatively closed system.

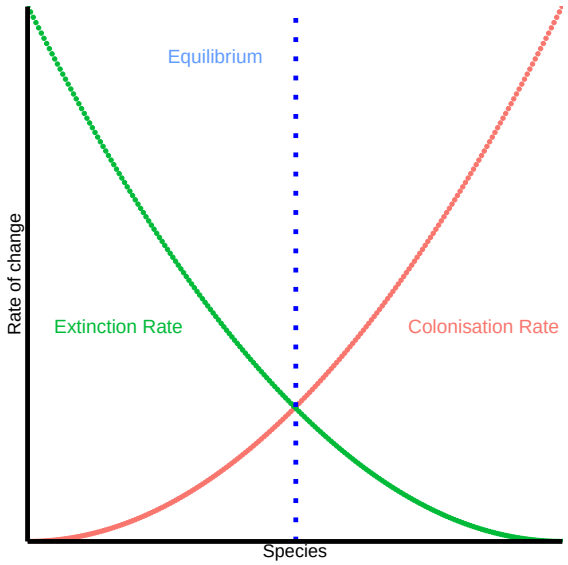


Figure 1.1: Colonisation-Extinction Dynamic Equilibrium

MacArthur and Wilson's Theory of Island Biogeography (MacArthur, 1967) is one of the most widely accepted island SAR theories. Explains the maintenance of biodiversity on islands through the stochastic processes of colonisation and extinction. The rates of these processes are determined by island area and isolation from the mainland. Islands that are nearer to source populations will experience a higher rate of immigration. This in turn can produce a rescue effect leading to decreased extinction rates (Brown and Kodric-Brown, 1977). Larger islands will receive more immigrants as species actively target larger habitats with more resources, or will be more likely to immigrate randomly due to island size. A larger population is also less susceptible to inbreeding depressions and random extinction (MacDonald et al., 2018). This results in higher species richness at the point of balance between immigration and extinction rates (i.e. the colonisation-extinction dynamic equilibrium, Figure 1.1) for larger, less isolated islands.

It has been suggested that the stochastic significance of area in predicting species richness has been overplayed, to the exclusion of deterministic mechanisms such as interspecific relationships, biotic and abiotic factors (Abbott, 1974). This is due to empirical evidence suggesting that smaller islands do not always follow the positive SAR pattern (Triantis

et al., 2006) (Sfenthourakis and Triantis, 2009). MacArthur and Wilson noted that archipelagos showed unusual SARs, with smaller island species-richness varying independently of size (MacArthur, 1967). It appears when smaller habitats within a broad range of spatial scales are assessed, both deterministic and stochastic patterns can emerge (Lomolino and Weiser, 2001). This exception to MacArthur and Wilson's putative ecological law has been dubbed the small-island effect (SIE).

Several hypotheses have been offered to explain the SIE. The 'subsidized island biogeography' hypothesis suggests that smaller islands have a greater edge to interior ratio, thus receive a greater amount of nutrients per unit area (Barrett et al., 2003) (Anderson and Wait, 2001). Secondly, extinction rates on islands may operate independently of area due to their environmental instability and high temporal turnover, where major episodic disturbances periodically wipe out colonising species (MacArthur, 1967). Thirdly, the 'habitat hypothesis' suggests that diversity is limited on smaller islands, compared to larger islands (Triantis et al., 2008). However, the environmental instability and habitat hypotheses contradict empirical data that indicate small islands have unusually high numbers of species. The Habitat–Diversity Hypothesis addresses the SIE phenomena by stating that as observation area increases we encounter a greater range of habitats (Connor and McCoy, 1979). Therefore, the theory predicts that species richness should increase with habitat diversity, which varies independently of area (MacDonald et al., 2018).

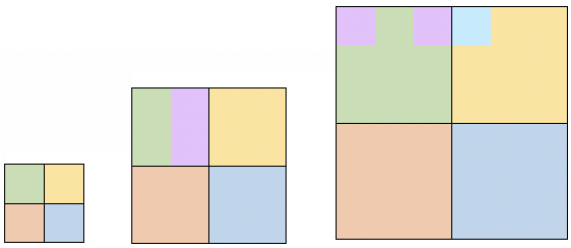


Figure 1.2: A graphical representation of a simulation (using the Classic Model, see Methods) of three islands of varying size, with the same number of niches (K=4) and **low immigration rate** (m_0 = 0.03). Each of the three main squares represents an island. Each smaller square represents an individual niche. Each unique colour patch within a niche represents a unique species. The smallest island has one species her niche, the medium size island has four individuals per niche and the largest island has nine individuals per niche. Species richness on the smallest island is **4**, on the medium island is **5** and the large island is **6**

Chisholm *et al.,* (2016) explain both deterministic and stochastic SARs in a unified theory. They posit that this pattern of species-richness is due to a transition from a niche-structured

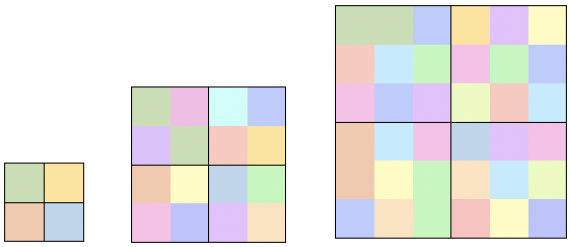


Figure 1.3: A graphical representation of a simulation (using the Classic Model, see Methods) of three islands of varying size, with the same number of niches (K=4) and **high immigration rate** ($m_0 = 0.9$). Each of the three main squares represents an island. Each smaller square represents an individual niche. Each unique colour patch within a niche represents a unique species. The smallest island has one species her niche, the medium size island has four individuals per niche and the largest island has nine individuals per niche. Species richness on the smallest island is **4**, on the medium island is **15** and the large island is **33**

regime on smaller islands, to a colonisation-extinction regime on larger islands. The niche-structured regime is characteristic of deterministic theories like the Habitat–Diversity Hypothesis, where habitat structure and intra- and interspecific interactions determine species richness (Chase and Myers, 2011). The colonisation-extinction regime is characteristic of stochastic mechanisms such as the Theory of Island Biogeography and ecological neutral theory, where richness is dictated by random colonisation and extinction events, as well as ecological drift (Hubbell, 2001). Chisholm *et al.,* hypothesise that species richness on all islands is maintained by these two mechanisms. They suggest that niche diversity increases slowly with area, whilst immigration rate increases quickly. Thus smaller islands are constrained by niche-structured regimes, until a critical area threshold where species richness is constrained by immigration. Figures 1.2 and 1.3 show the effect of immigration rate and area on species richness, where each 'island' is made up of four niches and the different coloured patches inside each niche represent a species unique to that niche. For islands with low immigration (Figure 1.2) or high immigration (Figure 1.3), small islands harbour the same number of species. Small islands where each niche can support only a small number of individuals will be constrained by those niches. Larger islands have less species at lower immigration rates and more species at higher immigration rates. They are less constrained by the number or size of their niches, and their species richness is dictated by random immigration and extinction events.

Chisholm *et al.,* developed a parsimonious mechanistic model to test their hypotheses, and applied it to 100 archipelago datasets. Their results supported the prediction that critical area will be lower for species with higher motility and less isolated habitats.

Previous research indicates that microbial TARs may be controlled by either deterministic, environmental mechanisms, or stochastic, neutral processes (Stegen et al., 2012). Phylogenetic analysis of subsurface microbial communities showed related taxa utilised similar habitats, illustrating that environmental filtering determined community composition (Stegen et al., 2012). Niche filtering also had a greater influence in the most spatially and temporally varied environments (Stegen et al., 2012). The relative strength of these mechanisms has also been shown to vary with community functionality (Caruso et al., 2011).

Whilst both stochastic and deterministic processes are demonstrated for microbial communities, few studies discuss the transition of mechanisms across a spatial scale. An investigation of phytoplankton TARs in water bodies indicated that for the smallest spatial scales, niche relations determine OTU richness, before transitioning to an immigration dominated regime (Várbíró et al., 2017). The SIE has also been seen in benthic diatoms where it is suggested stochastic variation in OTU richness is a function of the decreased stability of smaller habitats (Bolgovics et al., 2016).

Aquatic habitats are some of the most studied in microbial biogeography due to the availability of insular water bodies and their range of spatial scales. An investigation into bacterial diversity in aquatic tree holes found a z value comparable to macro-organisms ($z = 0.26$) (Bell et al., 2005). Antarctic cryoconite holes have also exhibited positive TARs on glaciers where biomass influx was limited, illustrating the significance of immigration rate (Darcy et al., 2018). Positive associations between habitat area and microbial OTU richness have also been reported for habitats as diverse as lakes (Battes et al., 2019), membrane bioreactors (Van Der Gast et al., 2006) (Van Der Gast et al., 2005) and vertebrate bodies (Godon et al., 2016).

Previous investigations into ectomycorrhizal fungi communities within 'tree island' root systems showed that total OTU richness increased significantly with size, although distance effects vary (Glassman et al., 2017) (Peay et al., 2007). An investigation into bacterial and fungal diversity in a group of land-bridge islands showed OTU richness for both groups was positively correlated with area, but these same patterns were driven by different mechanisms (Li et al., 2020). The bacterial TAR was a produce by differences in habitat quality

with island area, and the fungal TAR was driven by within–island dispersal limitation. Laboratory based experiments have supported the presence of soil microbial TARs as well as the influence of resource availability on OTU richness (Delgado-Baquerizo et al., 2018). Country and continent-scale patterns of pathogen diversity have also been shown to be a function of area and isolation (Jean et al., 2016) (Cashdan, 2014). In both terrestrial and aquatic systems microbial communities exhibit significant TARs. The varying mechanisms underlying these TARs warrant further investigation.

In this project I apply three modified versions of the model presented by Chisholm *et al.,*:

· **Classic Model**: Where per capita immigration rate is proportional to habitat area (e.g. in the case of aerial and directed dispersal species immigrating into a two-dimensional habitat)
· **Perimeter Model**: Where per capita immigration is proportional to habitat perimeter (e.g. in the case of water dispersed species immigrating into a two-dimensional habitat)
· **Depth Model**: Where per capita immigration rate is proportional to depth (e.g. in the case of species dispersing into a volume via its surface into a three-dimensional habitat)

These models are applied to bacterial, archaeal and micro-eukaryote insular spatial data with the aim of testing whether there is a biphasic microbial TAR, as well as investigating the impact of habitat type and taxonomic group on critical area of transition between the deterministic and stochastic regimes.

Methods

I developed a simulation model with three variants, based on the model presented by Chisholm *et al.,* (2016). The variants consist of the Classic Model, and two modifications (Depth Model and Perimeter Model) for use in microbial TARs. I verified the simulation data by comparison to analytic results from the simplified equation presented by Chisholm *et al.,* (2016) (2.2). The model fitting procedure was validated by fitting the model to simulation data with known parameters. The model fitting procedure was then applied to empirical data from the literature.

2.1 The Model

The model describes both a metacommunity and an island community. The metacommunity represents a mainland or source population from which propagules can immigrate to island communities. Both the metacommunity and island communities are made up of K non-overlapping niches. We assume that area is measured in number of individuals present ($\rho = 1$). Neutral theory assumes that an individual's probability of birth and death do not depend on its species or density. In my model, each niche follows two suppositions of ecological neutral theory: niches operate under a zero-sum assumption (where niche community size is constant) and each species within the niche is considered ecologically equivalent (with the same probability of producing a propagule or dying). The metacommunity is generated by a modified coalescence algorithm (Rosindell et al., 2008). A separate metacommunity is generated for each of K niches, where K is the number of niches to be simulated and each niche consists of 10 000 individuals (J). Therefore the total number of individuals in the metacommunity is J^*K. The metacommunity is assumed to be constant over timescales relevant to the island communities. For details of the modified metacommunity algorithm see Supplementary Materials.

Each island community consists of K niches, and each niche is initiated with one unique species. At each timestep we process one randomly selected niche of every island. When we process a niche of an island, one individual in that niche dies and leaves a gap for another individual of a species suitable to occupy that niche. With probability m (the per capita immigration rate), the dead individual is replaced with a randomly chosen propagule from the metacommunity. Species may only immigrate from the same niche in the metacommunity, to the corresponding niche in the island community. With probability 1 - m, the dead individual is replaced with a local propagule from the same niche. The species richness for each niche is calculated, totalled across all niches for each island and stored at every 5000 timesteps.

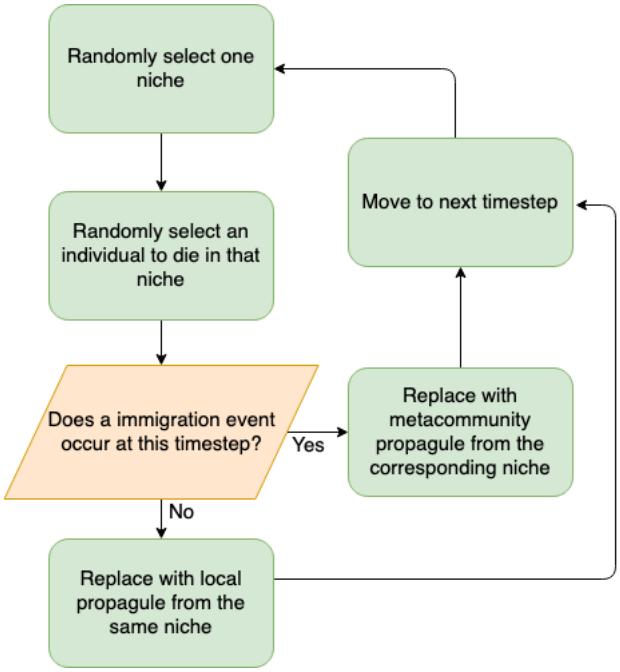


Figure 2.1: Flowchart of simulation design

The probability of an immigration event occurring at each birth/death event (m) is calculated with three variations, the Classic Model, the Depth Model and Perimeter Model (Figure 2.2). m_0 is the immigration constant parameter, given to the simulation, from which the per capita immigration rate is calculated. The Classic Model is the original method presented by Chisholm *et al* and is most appropriate for species utilising aerial or directed dispersal, where immigration rate is directly proportional to area and the habitat is considered two dimensional. The Depth Model simulates immigration into a three-dimensional space where species inhabit depth as well as area. For the Depth Model, each island simulation is given a depth of 1 unit. The Perimeter Model has immigration propor-

tional to perimeter, for species that immigrate across land or water and whose likelihood of encountering a habitat is based on its perimeter. The Perimeter Model also considered habitat to be two-dimensional.

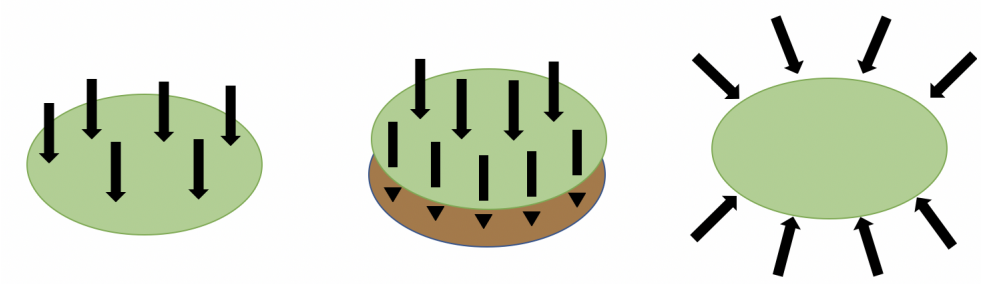


Figure 2.2: **Classic** (m=m_0), **Depth** (m=$m_0/depth$), **Perimeter** (m=$m_0/\sqrt{area}$)

I ran 200 simulations 100 times for each of the three models (Classic, Depth and Perimeter) using the High Performance Computing (HPC) service at Imperial College London. The 200 simulations consisted of 20 simulation sizes differing in powers of 2 multiplied by 10 unique parameter combinations (Table 2.1). The m_0 values for the Perimeter Model simulation were higher than those for the Classic and Depth Models, to allow the simulation to reach equilibrium within 24 hours. The parameter values of my simulations are not indicative of real-world speciation or immigration rates and are arbitrary values that facilitated building a metacommunity and running the simulation to equilibrium within 24 hours. The simulations were run in groups on the cluster for 24 hours to reach equilibrium. I plotted a timeseries for each simulation to ensure equilibrium has been reached, before calculating the mean species richness for set of simulation parameters.

Table 2.1: Simulation Parameters where the same range of parameters is applied to each of the Classic, Depth and Perimeter Model simulations. Higher m_0 values were given to the Perimeter Model simulation to allow it to reach equilibrium within the 24 hours. Simulations with larger areas and low immigration rates (m) take longer to reach equilibrium and for the Perimeter Model simulation m=$m_0/\sqrt{area}$ thus immigration rates were considerably lower than for the Classic and Depth Models.

Parameter	**Values**
speciation rate (nu)	0.00001 - 0.0001
immigration constant (m_0)	0.01 - 0.1 (0.09-0.1 Perimeter)
number of niches (K)	5 - 50
simulation areas (i.e. number of individuals)	5 - 20,000

2.1.1 Validation and Model Fitting Procedure

The results of the model were found analytically by applying the simplified mathematical model presented by Chisholm *et al.,* (2.2-2.5). Analytic and simulation results were compared to ensure the model had been simulated successfully. In these analytical solutions the species richness (S) is given by:

$$S = \theta\{\psi(\frac{\theta}{K} + \gamma(\psi(\gamma + J) - \psi(\gamma))) - \psi(\frac{\theta}{K})\} \tag{2.1}$$

Where ψ is the digamma function, J is the number of individuals per niche, K is the number of niches and θ is the fundamental biodiversity number calculated as *nu*(J*K-1)/(1-nu)*

$$\gamma = (J-1)m/(1-m) \tag{2.2}$$

and

$$m = m_0 \quad \text{or} \quad m = m_0/D \quad \text{or} \quad m = m_0/\sqrt{A} \tag{2.3}$$

Where m_0 is the immigration constant used to calculate per capita immigration rate m, D = depth, A = area and

$$J = A\ \rho/K \quad \text{or} \quad J = A\ \rho\ D/K \tag{2.4}$$

Where $J = A\ \rho/K$ is used for the two-dimensional models (Classic and Perimeter) and $J = A\ \rho\ D/K$ is used for the three-dimensional model (Depth). For all simulations ρ is given as 1 as area (A) is measured in units corresponding to number of individuals.

The mathematical model was fit to the simulation data using non-linear least squares (NLLS) fitting in R (version 3.6.1 (2019-07-05)). The model fitting procedure uses the minpack.lm package. This package provides a Levenberg-Marquardt NLLS fitting function (nlsLM) that uses a more robust algorithm than it's base-R counterpart.

The three parameters of the model are K, m_0 and θ and to aid the fitting procedure I gave it starting parameters $\hat{\theta}$ and $\hat{m_0}$ corresponding to estimated values of what those three parameters will be. Values of K were looped through from 1 to maximum species richness

and from this $\hat{m}_0$ and $\hat{\theta}$ starting values were calculated, where A_{med} is median surface area of habitats, $S_{A_{max}}$ is species richness in the largest habitat and $W_{-1}(x)$ is the lower branch of the Lambert W function. The best-fit values were stored for the K and corresponding $\hat{m}_0$ and $\hat{\theta}$ that gave the highest R^2 score. They were calculated as follows:

$$\hat{m_0} = \frac{-K}{\rho\ A_{med}\ W_{-1}\ (-K\ \rho\ A_{med})} \tag{2.5}$$

$$\hat{\theta} = \frac{S_{A_{max}}\ \hat{\gamma}\ log\ \hat{m_0}}{S_{A_{max}} - \hat{\gamma}\ log\ \hat{m_0}\ W_{-1}(exp(S_{A_{max}}/\hat{\gamma}\ log\hat{m_0}\ S_{A_{max}})/(\hat{\gamma}\ log\ \hat{m_0}))} \tag{2.6}$$

Where $\hat{\gamma}$ is calculated as:

$$\hat{\gamma} = \frac{(\rho\ A_{max}) - 1)\hat{m_0}}{1 - \hat{m_0}} \tag{2.7}$$

Using NLLS to fit the three models to their corresponding simulations I was able to validate the fitting procedure by recapturing the known parameters (K, m_0 and θ) using area and species data alone.

2.1.2 Critical Area

The critical area of transition from a niche-structured regime to an extinction-colonisation equilibrium regime where the TAR starts to increase to a steeper gradient, can be calculated according to Chisholm *et al.*, by these formulas:

Classic Model

$$A_{\text{crit}} = \frac{\theta(1 - m_0)(exp(K/\theta) - 1)}{m_0\ \rho\ log(1/m)} \tag{2.8}$$

Depth Model

$$A_{\text{crit}} = \frac{1}{\rho\ D}\left[\frac{\theta(exp^{\frac{K}{\theta}}-1)(D-m_0)}{m_0\ log(\frac{m_0}{D})}+1\right] \tag{2.9}$$

Perimeter Model

$$x = \frac{\theta(exp^{\frac{K}{\theta}}-1)}{m_0\rho} \quad \text{and} \quad A_{\text{crit}} = \left\{\frac{x}{W_0(x/m_0)}\right\}^2 \tag{2.10}$$

For each model fitting, critical area is estimated from the best-fit parameters of K, θ and m_0. For higher values of K and lower values of θ, m_0 and ρ, the critical area will be larger. The critical area formulas were validated by visual inspection of fits to the simulation data.

With parameter estimations obtained from the NLLS fitting of the three models to the empirical datasets, A_{crit} was estimated. For habitats with varying area and depth values A_{crit} was calculated for each depth value and the mean taken. The mean A_{crit} was then multiplied by the mean depth of the dataset to give a mean critical volume (V_{crit}) and plotted with habitat volume (depth x area) and species richness. The mean A_{crit} from these fittings is used in the statistical analysis for comparison with homogenous depth habitats. By finding A_{crit} for empirical datasets, I was able to test the theory that the regime shift would occur at lower areas for less isolated habitats and for more motile taxa.

2.2 Data Collection and Analysis

I compiled 57 datasets from 29 studies on microbial TARs (Table 2.2). A full description of each dataset can be found in Supplementary Materials. They include a range of taxonomic groups as well as habitat types. Before performing the model fitting process, the datasets were tested for Spearmans Rank correlation to assess if there was a positive correlation between area and OTU richness. This allowed me to exclude datasets that could not be fit with the model. The datasets with a positive correlation were imported into each of the three NLLS fitting scripts (Classic, Depth, Perimeter). The parameter ρ was estimated separately from the other three parameters. Estimations were taken from the original study or proxy papers (see Supplementary Material). In addition to plotting the three

Table 2.2: Summary of datasets collected from the literature

Taxa	*Habitat*	*Count*
algae	lacustrine	10
algae	riverine	3
archaea	lacustrine	1
bacteria	lacustrine	9
bacteria	machine	4
bacteria	terrestrial	4
fungi	lacustrine	4
fungi	plant	5
fungi	terrestrial	4
pathogen	terrestrial	4
protozoa	lacustrine	7
protozoa	riverine	1
protozoa	terrestrial	1

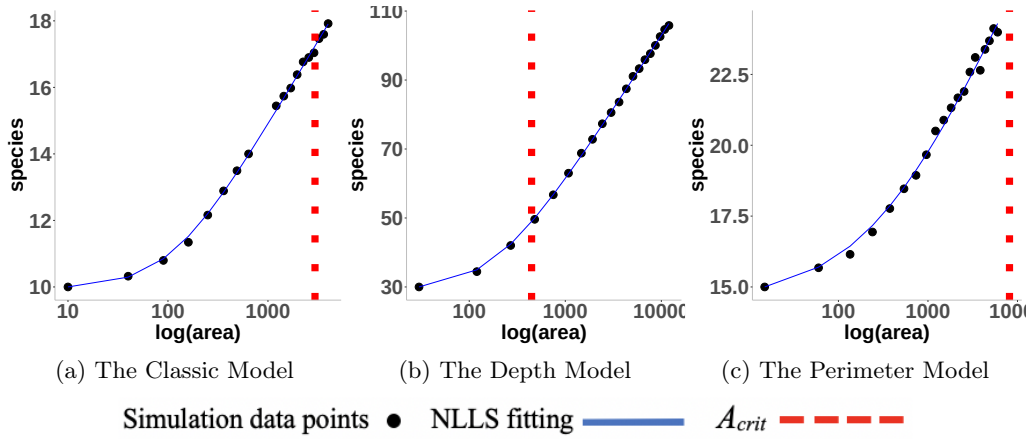
model variations, I also fitted each dataset with the simple power-law model (Introduction, Equation 1.1). The power-law model fit is compared to each of the Classic, Depth and Perimeter model fits using Akaike Information Criterion (AIC) score to determine if, by incorporating θ, K, m_0 and ρ parameters, our models are better fits to the data.

2.3 Statistical Analysis

For empirical datasets where adjusted R^2 was between 0 and 1 and at least one of the three models was a reasonable fit, I found the mean R^2 and adjusted R^2 as well as the standard deviation and range. The median, standard deviation and range of the fitted parameters was also found. Spearmans rank tests were used to assess correlations between parameters. Differences between mean critical area across habitat types and taxonomic groups was found. Multiple regression analysis of log critical area (A_{crit}) was carried out with habitat type and taxonomic group as categorical explanatory variables.

Results

3.1 Simulation



(a) The Classic Model (b) The Depth Model (c) The Perimeter Model

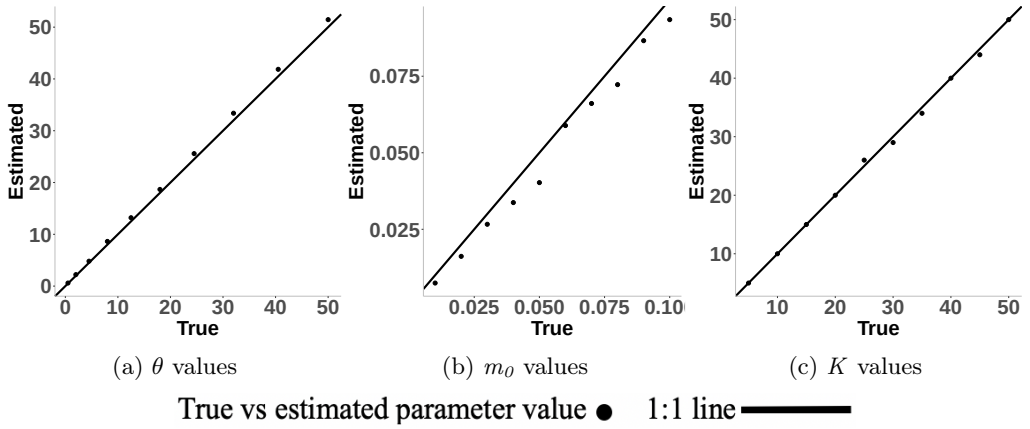
Figure 3.1: NLLS fitting of the the simulation data. A) The Classic Model with true parameters θ=20, m_0=0.2, K=10 and estimated parameters θ=20, m_0=0.2, K=10. B) The Depth Model with true parameters θ=18, m_0=0.06, K=30 and estimated parameters θ=19, m_0=0.05, K=30. C) The Perimeters Model with true parameters θ=30, m_0=0.5, K=15 and estimated parameters θ=57, m_0=0.46, K=15.

My results verified that the simulation and analytic formula are in agreement as expected. The three critical area formulas (see Methods 2.8-2.10) for each of the three model variations (Classic, Depth, Perimeter) give reasonable estimations of critical area (log A_{crit}).

3.1.1 Classic Model

The Classic Model fitting procedure fitted the simulated data well. Estimated parameters were slightly higher than the true parameters for θ and slightly lower for m_0 and K. There was a significant difference between the true and estimated values for θ (p=0.0008, 9 df) and m_0 (p=0.0003, 9 df). There was no significant differences in true and estimated parameters for K (p=0.34, 9 df). As K true and estimated parameters showed no significant

difference, and there was only a small difference between m_0 and θ true and estimated parameters, the model fitting process is considered validated and fit for applying to the empirical datasets.



(a) θ values (b) m_0 values (c) K values

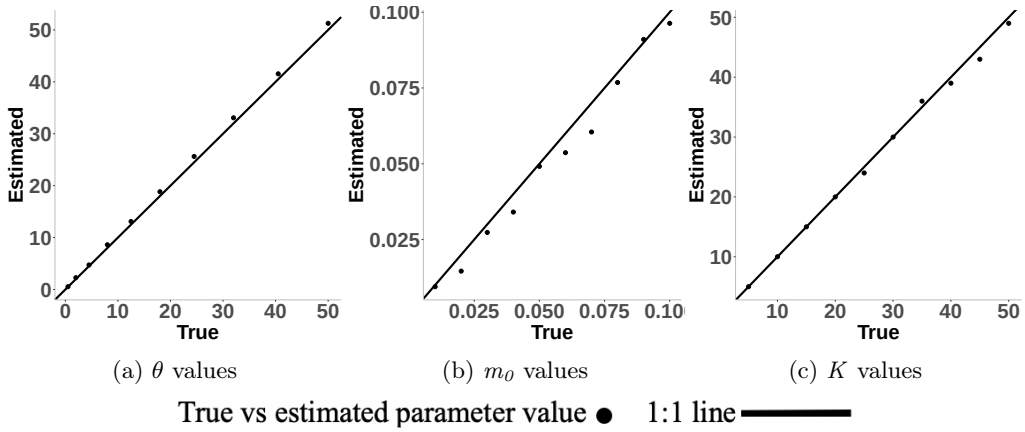
Figure 3.2: The true parameters values of θ (A), m_0 (B) and K (C) were simulated and the results fitted using the Classic Model analytical NLLS fitting procedure to get the parameters back. True and estimated values for the three fitted parameters are plotted above. Fittings of the Classic Model to simulated data returned mean $R^2 = 0.99$, adjusted $R^2 = 0.99$.

Table 3.1: Comparison between true and estimated mean parameters across 200 Area Model simulations clustered into 10 groups where parameter values (θ, m_0, K) were the same for each simulation group with varying areas.

Parameter	*True*	*Estimated*	*Difference*
theta	19.251	20.040	−0.789
m0	0.055	0.050	0.005
K	27.500	27.300	0.200

3.1.2 Depth Model

The Depth Model fitting procedure fitted the simulated data well. Estimated values of θ were higher, K and m_0 values were lower than the true parameters (Table 3.2). There was a significant difference between the true and estimated values for θ (p=0.0005, 9 df) and m_0 (p=0.005, 9 df), but there was no significant difference for K (p=0.168, 9 df). As K showed no significant difference, and the difference in estimated θ and m_0 values was so low, the model fitting process is considered validated and fit for applying to the empirical datasets.



(a) θ values (b) m_0 values (c) K values

Figure 3.3: The true parameters values of θ (A), m_0 (B) and K (C) were simulated and the results fitted using the Depth Model analytical NLLS fitting procedure to get the parameters back. True and estimated values for the three fitted parameters are plotted above. Fittings of the Depth Model to simulated data returned mean $R^2 = 0.99$, adjusted $R^2 = 0.99$.

Table 3.2: Comparison between true and estimated mean parameters across 200 Depth Model simulations clustered into 10 groups where parameter values (θ, m_0, K) were the same for each simulation group with varying areas.

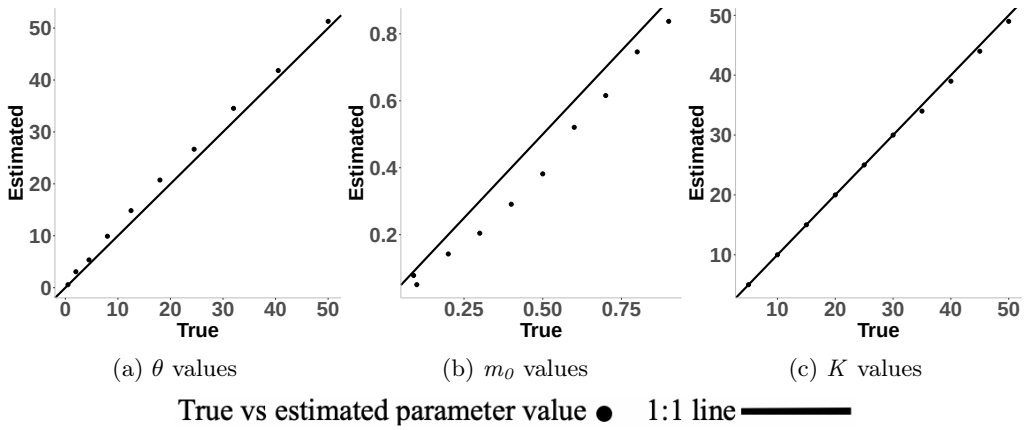
Parameter	*True*	*Estimated*	*Difference*
theta	19.251	19.965	−0.713
m0	0.055	0.051	0.004
K	27.500	27.100	0.400

3.1.3 Perimeter Model

Table 3.3: Comparison between true and estimated mean parameters across 200 Perimeter Model simulations clustered into 10 groups where parameter values (θ, m_0, K) were the same for each simulation group with varying areas.

Parameter	*True*	*Estimated*	*Difference*
theta	19.251	20.885	−1.634
m0	0.459	0.387	0.072
K	27.500	27.100	0.400

The Perimeter Model fitting procedure fitted the simulated data well. Estimated parameters for θ were slightly higher than true parameters, whilst m_0 and K were slightly lower than the true parameters (Table 3.3). There was significant difference between the true and estimated values for θ (p=0.0002, 9 df), m_0 (p=5x10^5, 9 df) and K (p=0.04, 9df). Despite the significant difference between the estimated and true values for my simulation parameters, the differences are small enough that the model fitting process is considered



(a) θ values (b) m_0 values (c) K values

Figure 3.4: The true parameters values of θ (A), m_0 (B) and K (C) were simulated and the results fitted using the Perimeter Model analytical NLLS fitting procedure to get the parameters back. True and estimated values for the three fitted parameters are plotted above. Fittings of the Perimeter Model to simulated data returned mean $R^2 = 0.99$, adjusted $R^2 = 0.99$.

validated and fit for applying to the empirical datasets.

3.2 Model Fitting

3.2.1 Non-Linear Least Squares Fitting

50 of the 57 datasets exhibited a positive TAR and were used for the NLLS fitting. Of the 50 datasets fitted with the three models, 26 failed to achieve adjusted R^2 scores of between 0 and 1. These datasets were excluded from further analysis (see Supplementary Materials, Figure 7.2).

Table 3.4: The mean R^2 and adjusted R^2 results for each model (Classic, Depth, Perimeter) after being successfully fitted to 26 empirical datasets.

$Model$	R^2	$AdjR^2$
Classic	0.492	0.409
Depth	0.493	0.410
Perimeter	0.486	0.401

All three models had similar mean R^2 and adjusted R^2 scores and total results show the models fit the data moderately well (Table 3.4). The Classic Model was best-fit for 1 dataset, Depth and Perimeter were best for 2 each and the rest of the datasets were either best described by both Classic and Depth or all of the models (Table 3.5). The differences in R^2 scores for each of the fittings was small (see Supplementary Materials, Table 7.5). For datasets that were equally well fit to two or more of the models I found the mean

A_{crit} and parameter estimations (θ, m_0, K) for each of those fittings and used these in the proceeding analysis.

Table 3.5: The best-fit models (Classic, Depth, Perimeter) by highest adjusted R^2 value for each empirical dataset (note some datasets had equal adjusted R^2 values for two or more models).

Models	*BestFit*
Classic	1
Depth	2
Perimeter	2
Classic and Depth	10
Classic and Perimeter	0
Depth and Perimeter	0
All	9

The best model fits had mean $R^2 = 0.49$ and mean adjusted $R^2 = 0.41$ with standard deviation 0.28 and range 0.01 - 0.96. The median value of θ was 8, with a range of 0.28 – 159709. The median value of m_0 was 2.17 x10^{-9} with a range of 4.97 x 10^{-16} – 0.56. The median value for K was 7, with range 1 – 424. There was no correlation detected between the best fitted-values of the four parameters (Table 3.6).

Table 3.6: p-values of correlations between the four model parameters (θ, m_0, K, ρ) that show no correlation

Parameter	*K*	*Theta*	m_0	*rho*
K	NA	0.472	0.369	0.518
Theta	0.472	NA	0.110	0.200
m0	0.369	0.110	NA	0.140
rho	0.518	0.200	0.140	NA

Mean z value for the 50 datasets (prior to removing failed fits) was 0.16. The power-law model had the same number of successful fittings as the Classic, Depth and Perimeter models. After removing failed fits the mean z value remained the same. The power-law model performed slightly more poorly than the other three models (R_2=0.47, adjusted R_2=0.38) (see Supplementary Materials, Table 7.3). AIC scores indicated that the power-law model was not a more parsimonious model than the Classic, Depth or Perimeter models relative to model fit for any of the 24 successfully fitted datasets (see Supplementary Materials, Table 7.4). The Classic and Depth models were significantly better than the

power-law model for 9 datasets each. The Perimeter model was a better fit than the power-law model for 5 datasets.

3.3 Critical Area

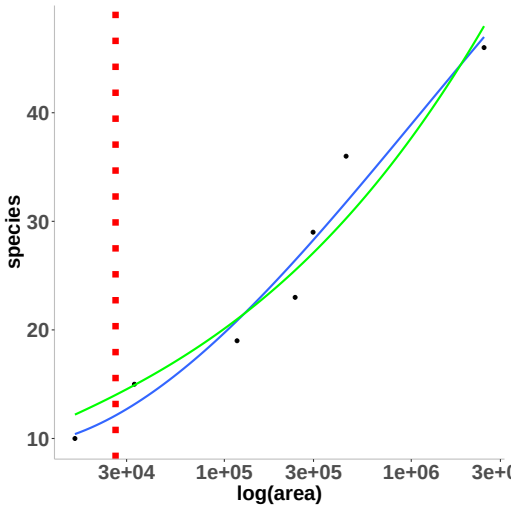
Of the five habitat types (terrestrial, riverine, lacustrine, plant and machine) and six taxonomic groups (algae, archaea, bacteria, fungi, pathogens and protozoa), the riverine habitat and archaea group did not have any successful fittings and are excluded from the following analysis.

The log A_{crit} data were not normally distributed with non-homogenous variances. Despite the violation of normality I have proceeded with the multiple regression analysis, although interpretation of results will take this into consideration.

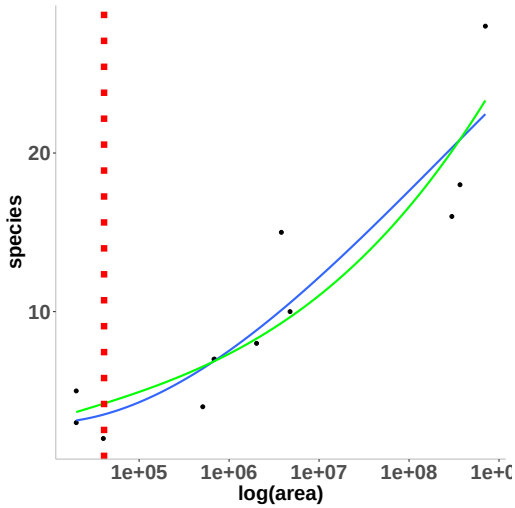
Initial multiple regression revealed that the model was a poor fit to the data (R^2=0.39, adjusted R^2=0.05, p=0.374) and neither categorical variable was significant in predicting log A_{crit} (habitat type p=0.1759, taxonomic group p=0.6402). A plot of the model indicated that there was an outlying data point. After removing the outlying data point the model was significant in describing the data (R^2=0.62, adjusted R^2=0.44, p=0.02). Taxonomic group became weakly significant in predicting log A_{crit} (p=0.0187) but habitat type did not (p=0.097). After removing habitat type as a categorical dependent variable the model was a similar fit to the data but more significant (R^2=0.55, adjusted R^2=0.45, p=0.004).

Multiple regression including taxonomic group only upheld the prediction that A_{crit} would occur at lower areas for more motile OTUs as bacteria show the lowest logA_{crit} estimate and host-dependent pathogens show the highest (Table 3.7).

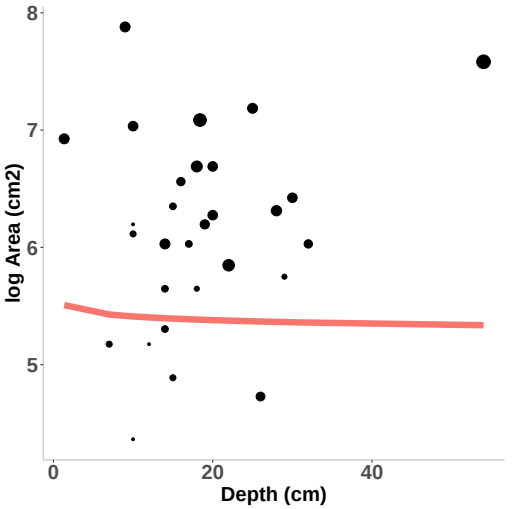
There was a large variation in mean log A_{crit} between habitats and taxonomic groups. Terrestrial habitats showed the highest mean log A_{crit} (27.33), whilst machine habitats showed the lowest (4.66) (Figure 3.6). Pathogens exhibited the largest mean A_{crit} for taxonomic groups (55.15), with bacteria having the lowest (4.06) (Figure 3.6).



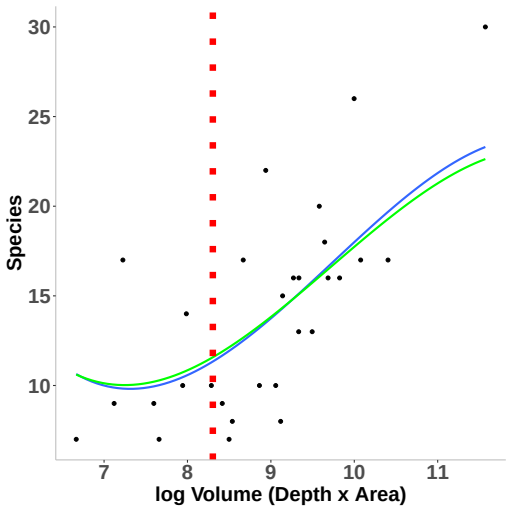
(a) Dataset 45, bacteria in biomembrane reactors



(b) Dataset 44, fungi in plant root soil



(c) Dataset 46, bacteria in tree holes (log area plotted with depth)



(d) Dataset 46, bacteria in tree holes (log volume plotted with OTU richness)

Figure 3.5: A) Best-fit Classic Model for dataset 45, bacteria in biomembrane reactors. Red line indicates A_{crit}, blue line indicates NLLS fit and green line indicates power-law fit (R^2=0.96, adjusted R^2=0.88, θ=9, m_0=4.97 x 10^{-16}, K=7). B) Best-fit Perimeter Model for dataset 44, fungi in plant soil (R^2=0.85, adjusted R^2=0.77, θ=5, m_0=6.15 x 10^{-11}, K=2). C) Best-fit Depth Model for dataset 46, bacteria in freshwater treeholes. The size of the black circles represents increasing OTU richness at that corresponding depth (x-axis) and log area (y axis) (R^2=49, adjusted R^2=0.40, θ=8, m_0=3.75 x 10^{-9}, K=6). Where the red line passes through depth and area space is where A_{crit} occurs. D) Dataset 46 plotted as log Volume by OTU richness to illustrate the model fit and log critical volume (A_{vol})

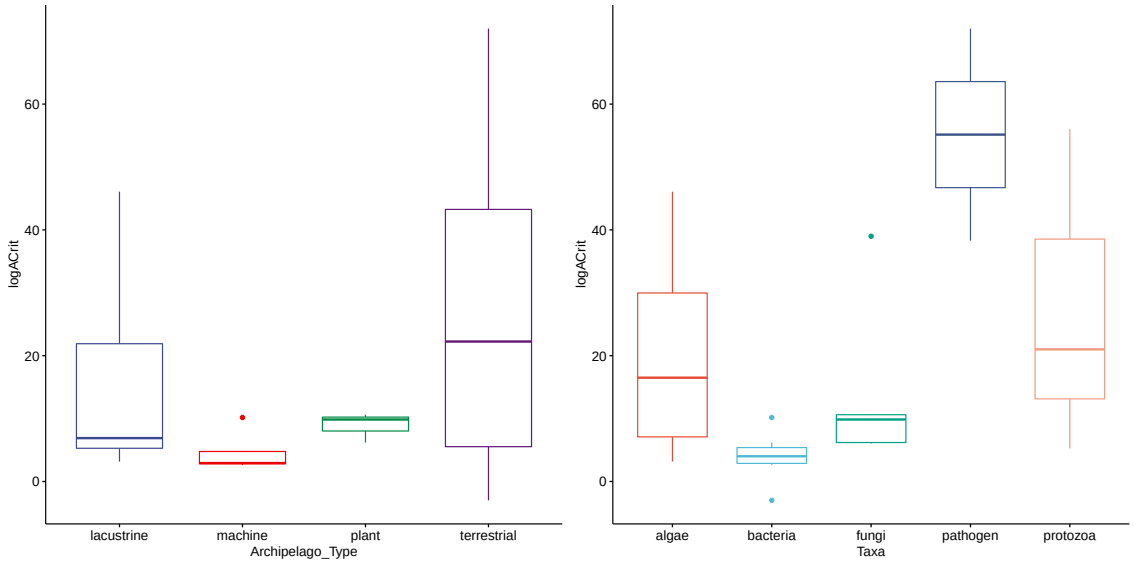


Figure 3.6: log A_{crit} by habitat type and taxonomic group after removing anomalous result

Table 3.7: Table showing the results of multiple regression analysis of estimated effect of taxonomic group only on log A_{crit}

Variable	**Estimate**	**95% CI**	**p-value**
intercept (algae)	20.56	[5.07, 36.07]	0.0121
taxonomic group			0.004
bacteria	-16.500	[-35.12, 2.12]	0.0791
fungi	-6.222	[-27.01, 14.56]	0.5373
pathogens	34.587	[7.75, 61.42]	0.0144
protozoa	6.879	[-16.79, 30.55]	0.5491

Discussion

I have presented three variations of the Chisholm model (Chisholm et al., 2016) that take into account varying habitats and immigration routes and have successfully fit all three to microbial TAR data. The relatively equal success of the three model variations (Classic, Depth, Perimeter) suggests that immigration route is not a significant factor in defining microbial TARs (see Supplementary Materials, Table 7.5). Microorgansims can cross oceanic barriers via airborne dust particles (Rosselli et al., 2015), enter glacial habitats via stream deposition (Darcy et al., 2018) and reach pole-to-pole via airborne, animal vector and anthropogenic mediated dispersal (Kleinteich et al., 2017). Microbial OTUs likely utilise a variety of immigration routes when entering a new environment. No significant accuracy was lost in assessing three-dimensional habitats using the two-dimensional models (Classic, Perimeter), suggesting habitat depth did not affect OTU richness as strongly as area, where area acts as immigration portal into the three-dimensional habitat. Algae and bacteria have shown negative correlations between OTU richness and depth (Battes et al., 2019) (Turner et al., 2017). These patterns emerge as increasing habitat depth often accompanies nutrient-poor, low-energy environments. The immigration portal may be characterised by a nutrient-rich, high-energy stratification that is a more potent predictor of OTU richness than depth.

The phenomenological power-law model did not perform significantly better than the mechanistic Classic, Depth and Perimeter models, according to the AIC measure of parsimony, relative to model fit (see Supplementary Materials, Table 7.4). This indicates that the model parameters (θ, m_0, K, ρ) were useful in fully describing the shape of the observed TARs, thus supporting the hypothesis that OTU richness is influenced by the parameters, rather than being simply a constant power of the area. The mean slope of positive TARs across these datasets was comparable to macroorganisms (z=0.16) and higher than those previously reported for microbial taxa (Rosenzweig et al., 1995) (Green et al., 2004)

(see Supplementary Materials, Table 7.3). These observations show habitat area has a relatively strong influence on OTU richness. Isolated habitats may provide the stability needed for microbial taxa to reach equilibrium and for stronger TARs to arise, in contrast to turbulent continuous habitats (Bell et al., 2005).

The successfully fitted datasets exhibit both the classic MacArthur and Wilson (MacArthur, 1967) biogeographic pattern of increasing OTU richness with area and the small island effect of OTU richness varying independently of area. The results demonstrate that some microbial communities are constrained by niche-structured regimes at smaller areas where immigration is low, before transitioning to colonisation-extinction balance regimes at larger areas where immigration is high. This lends support to the theory that microbial species are not ubiquitous and unlimited in dispersal, that they can be limited by habitat heterogeneity, resource availability and dispersal barriers, but this is not a ubiquitous pattern with over 50% of the datasets failing to be fit by the model.

Many datasets with positive TARs (as indicated by Spearman's rank correlation coefficients) could not be successfully fit with the power-law, Classic, Depth or Perimeter models (see Supplementary Materials, Figure 7.2). Despite positive z values for these datasets, confidence intervals included zero and therefore were not statistically significant. Stochastic variation between data points inhibited the models from discerning significant TARs (see Supplementary Materials, Figure 7.2 a & b). The majority of failed fits were aquatic habitats and may be due to uncertainty in the spatial sample regime of a heterogenous habitat. In order to elucidate the the spatial patterns within these habitats, it might be useful to take a stratified approach. Some of the failed fits had too few data points in comparison to the parameters of the model, producing low adjusted R^2 values (see Supplementary Materials, Figure 7.2 b). Microbial TARs may also be undetected due to the disparity between sample and community sizes meaning rare taxa are missed (Woodcock et al., 2006).

This project is, to the best of my knowledge, the first attempt to apply a biphasic mechanistic TAR model to microbial data. The model demonstrates that when niche diversity increases slowly or remains constant and immigration increases quickly with area, a biphasic TAR is produced. At an A_{crit} specific to that habitat and taxonomic group, the TAR will transition from deterministic to a stochastic mechanisms. I hypothesised that A_{crit} would

be lower where immigration is higher (i.e. for more motile OTUs and less isolated habitats). My analysis indicated that taxonomic group was significant in predicting A_{crit}, while habitat type was not. It is likely that taxonomic group is significant in predicting A_{crit} as taxa are constrained (or liberated) by their own range of life cycles (activity and dormancy) and dispersal methods (sporulation, meteorological, biotic, anthropogenic, passive). According to this data isolated habitats present no significant dispersal barriers to microorganisms as a whole, although their relative accessibility varies between taxonomic groups.

My analysis indicated that pathogenic OTUs had overwhelmingly higher A_{crit} values (mean 9.43 x 10^{30} cm^2), thus they are more constrained by resource availability and dispersal barriers. I suggest this is due to their dependence on host species, although this will be directly related to the motility and sociability of their hosts. The two datasets used in this study quantify human pathogen richness on 'true' islands (Jean et al., 2016). Human pathogen OTU richness is negatively correlated with disease control efforts (Dunn et al., 2010). I suggest that global mitigation strategies such as behavioural change, medicine and vaccination (Nicolaides et al., 2020) mean pathogens face considerable dispersal barriers that limit immigration and constrain them to niche-structured spatial regimes over larger areas.

Bacterial OTUs exhibited the lowest mean A_{crit} value (3.02 x 10^3 cm^2). The small size of bacteria allows them to dispersal more freely than size-limited macroorganisms (Martiny et al., 2006). They may also overcome dispersal limitation through dormancy as a biogeographical response and as a consequence of enormous population sizes (Locey, 2010) (Fenchel and Finlay, 2004). Bacteria have a variety of ecological traits that allow them to move freely and access isolated habitats, thus they transition to stochastic TAR mechanisms at lower areas.

Fungi also showed low A_{crit} values (mean 1.72 x 10^{16} cm^2). Mycorrhizal fungi, where there are beneficial associations with plant roots, have large spores that immigration slowly through soil (Bueno and Moora, 2019), however, the close proximity of potential host plants might mitigate low fungal motility. For other fungal groups, long distance spore dispersal is facilitated by meteorological, biotic and anthropogenic vectors (Golan and Pringle, 2017). Fungal sporulation allows taxa to overcome local and regional barriers, thus contributing to the low A_{crit} values seen in these datasets.

Algae (mean 2.56 x 10^{19} cm^2) and protozoa (mean 7.43 x 10^{23} cm^2) exhibited similar midrange A_{crit} values. The broad range of A_{crit} values for these taxonomic groups may be due to the issue of spatial sampling regime in spatially heterogenous aquatic environments. Issues of taxonomic classification, particularly for protists may contribute to varying estimations of diversity (Foissner, 2006). Whilst seems that algae and protists transition from deterministic to stochastic mechanisms of spatial scaling in the midrange of areas, further investigation is needed to discern a true pattern within the wide range of A_{crit} values estimated. The multiple regression model coefficients (Table 3.7) broadly confirm the overall taxonomic results.

As the multiple regression analysis showed that habitat type was non-significant in predicting A_{crit} I cannot assess the relative isolation of habitats or how they may affect A_{crit}. It is interesting however to look at the mean A_{crit} values for each habitat type, as a sign post towards what may be found with a more comprehensive dataset. Terrestrial habitats show the highest mean A_{crit} values (2.36 x 10^{30} cm^2). This may be due to immigration via an accidental vector being limited to aerial species that can reach the land island. Passive immigration by water or air to land islands relies on stochastic success which may limit dispersal, although fungal and bacteria OTU richness has been shown be unaffected by isolation (Li et al., 2020).

Lacustrine habitats exhibited low mean A_{crit} values (1.28 x 10^{19} cm^2) suggesting immigration to these habitats is high. Aquatic taxa such as algae and protozoa utilise a variety of dispersal mechanisms between habitats, including dispersal via insects and waterfowl (Stewart and Schlichting Jr, 1966). It may be easier for microbial OTUs to colonise inland lacustrine environments where animal activity increases the probability of transport via an accidental vector. Passive transport to lacustrine environments may have a greater success rate than terrestrial islands due to the interconnectivity of rivers and streams that empty into watershed areas, filling lakes and ponds.

Plant habitats also have low mean A_{crit} values (2.01 x 10^4 cm^2). For many symbiotic plant-microbe species relationships, plant seeds are already inoculated with associated microbial taxa on dispersal (Ho et al., 2017). Thus dispersal barriers between plant and microbes are removed, contributing to low A_{crit} values. Many plant communities are comprised of the same species in close proximity, providing ready access to source populations and increasing immigration.

Four of the six best-fit datasets were for bacteria in machine habitats (membrane bioreactors and metal cutting machine sump tanks) (Van Der Gast et al., 2006) (Van Der Gast et al., 2005). It may be that the strong TAR found in these environments is a function of their isolation, relative to natural habitats. Despite the large numbers, rapid asexual reproduction and resilience to extinction of bacteria, when constrained by immigration, more prominent and easily quantifiable TAR patterns arise. The model fitting process supports this by estimating extremely low immigration rates for machine habitats. Despite this, machine habitats had the lowest mean A_{crit} (6.56 x 10^3 cm^2). A_{crit} is not only affected by immigration as in my primary hypothesis, but can also be affected by number of niches (K), density (ρ) and θ. In the fitted model the low A_{crit} for machine habitats in spite of their isolation is caused by the low K values of homogenous, man-made environments, a characteristic of these unusual habitats that warrants further investigation.
The non-significance of habitat type, despite marked differences in the mean A_{crit} is due to the large, overlapping estimate ranges. Overall, after removing the outlying datapoint and removing habitat type as an explanatory variable, the model accounts for nearly half of the variation in log A_{crit} using the broad taxonomic groups.

The reason for the lack of successful fittings for riverine habitats is due to the low number of data points in each of these studies (see Table 7.1, Supplementary Material). Whilst the majority of these datasets had high R^2 values, once adjusted R^2 values were calculated the fittings were unsuccessful. Only one dataset included archaeal TARs and no significant relationship between area and OTU richness was found. This is likely due to the importance of environmental filtering for extremophile OTUs in soda lakes (Lanzén et al., 2013). It is interesting to note other datasets removed from the fitting process due to a lack of positive TARs. These included, fungi in the Antarctic cryoconite holes of two glaciers where extreme biomass influx negated observable TARs (datasets 10 & 11) (Darcy et al., 2018). Inappropriate diversity metrics and spatial scaling may have led to undetectable toot-symbiotic fungi TARs (dataset 18) (Davison et al., 2018). Fungi OTU richness did not increase with area on submerged leaves due to a lack of energy increase with corresponding area as expressed by the species-energy theory (dataset 39) (Feinstein and Blackwood, 2012) (Wright, 1983). TARs may not have arisen in protozoan communities in submerged substrate due to a failure to reach equilibrium (dataset 55) (Henebry and Cairns Jr, 1980). It is clear that the factors affecting microbial TARs are diverse and each habitat/taxa pairing may require unique assessment.

The anomalous result removed from analysis concerned pathogenic bacterial OTU richness on 'true' geographic islands (Jean et al., 2016). The model was a poor fit to the data (R^2=0.23, adjusted R^2=0.18) and it's likely the error associated with estimating density for pathogenic bacteria over such large geographic scales lead to poor estimations of the remaining parameters and an excessive critical area estimate.

Whilst the data here indicate that habitat type is non-significant in predicting log A_{crit}, the large variation in mean log A_{crit} suggests there many be too few data points to discern a significant pattern. I also encountered challenges when trying to compare studies that used a variety of methods and quantification techniques. Microbial OTUs inhabit three-dimensional habitats and whilst steps have been taken here to account for this there is more work to be done to incorporating this fully. In a future extension of this mode I would consider each stratification of a habitat separately, to account for spatial heterogeneity. Volume has been shown to be more accurate in quantifying microbial TARs (Van Der Gast et al., 2006). It would be useful to further modify the model to explicitly incorporate volume and V_{crit} across datasets, as nearly all of them concern habitats within a volume even though often only surface area data is provided. Here I have used area with a depth metric (Depth Model) which suggests the habitat maintains the same area for the full depth, whereas natural habitats rarely take this shape and this reduces the accuracy of my results.

Another issue I encountered was density estimations. The model required an estimation of individual density per unit area, however, direct counts are rarely given for microbial OTUs. Estimations were made in various ways, using gene sequence numbers or proxy papers, although these methods introduce error into the model fitting process. It would be beneficial to this project to develop more robust methods for estimating density as data taken from proxy papers introduces error into the fitting process. A broad scale experiment to quantify microbial TARs in a laboratory, where data specific to the needs of these model could be collected (i.e. density), could provide a more vigorous assessment of the models applicability to microbial TARs.

When validating my model fitting procedures, error between true and estimated results increased with increasing parameter values. As I increased simulation parameters in par-

allel with each other (e.g. an increase in θ, was coupled by an increase in m_0 and K), the source of the increasing error is difficult to discern. Parameter ranges of speciation rate, m_0 and k given to the simulations were not inferred from microbial ecological theory, but were selected for ease of computation. A more thorough exploration of the parameter space, with ecologically relevant parameter ranges, to further validate the model fitting procedure and the areas of parameter space where there may be errors in fitting would be desirable in further work.

There remains to be a thorough synthesis between biogeography and microbial ecology. Here I have gone some way to evaluate the influence of immigration on microbial TARs, however more work is needed to examine dispersal barriers. Dormancy is a widespread microbial response that may allow OTUs to overcome dispersal barriers and increase immigration to new habitats. However, it is a slow, passive process that will not necessarily lead the individual to a viable habitat (Locey, 2010). To fully elucidate the interplay of microbial ecology and biogeographic patterns, work is needed to incorporate dormancy as a biogeographic response.

An implication of this work is that if we can identify the niches within a habitat and the taxonomic groups that tend to occupy those niches, we may be able to better predict OTU richness at a range of spatial scales. This presents a complex challenge that requires the integration of environmental variables and habitat stratification. If these challenges could be overcome, it would be a particularly useful tool in predicting colonisation of new habitats such as soil exposed by glacier retreat, thus helping us model the biogeochemical processes this colonisation will produce.

This study has demonstrated that microbial communities in isolated 'island' habitats can be subject to deterministic biogeographic mechanisms such as niche-structuring, before a critical area of transition (A_{crit}), to stochastic mechanisms of colonisation and extinction. I have also shown that taxonomic group is significant in predicting A_{crit}, but habitat type is not. The overwhelming number and complexity of microbial life, as well as the vital role these organisms play in ecosystem functioning, illustrates the importance of elucidating their biogeographic patterns. I hope that my study will lead to further research into the presence of deterministic and stochastic mechanisms in microbial biogeography, as well as the importance of taxonomic group on the relative influence of these processes. The

synthesis of microbial ecology and biogeography will be of increasing interest as climate change alters habitats, creating and removing barriers, extending the range of pathogenic OTUs and leading to climate feedback loops of mineral and nutrient cycling. Microbial biogeography is an essential area of study in our global challenge to predict and mitigate the impacts of climate change. Everything is *not* everywhere, and everything is changing.

Data and Code Availability

Name a data (e.g., Dropbox, FigShare, Zenodo, etc) and a code (e.g., Dropbox, GitHub, etc.) archive from where the data and code can be obtained that will allow replication of your results. The code may be in the form of a single script file. You will be taught the principles of reproducible analyses in the R week of your coursework. If the data cannot be made available publicly (e.g., because it is yet to be formally published), or if there are some other confidentiality issues with submitting the data, speak with your course director and supervisor, and include a clear statement about why the data cannot be made available under the same Code and Data Availability header. Note that most data repositories allow timed embargos on data (e.g., Zenodo; see http://about.zenodo.org/policies/).

Acknowledgements

Thank you to my supervisors, Ryan, James and Tom. Thank you to all of the people that took the time to share their data with me.

Bibliography

Abbott, I. (1974). Numbers of plant, insect and land bird species on nineteen remote islands in the southern hemisphere. *Biological Journal of the Linnean Society*, 6(2):143–152.

Anderson, W. and Wait, D. (2001). Subsidized island biogeography hypothesis: another new twist on an old theory. *Ecology Letters*, 4(4):289–291.

Antón, J., Rosselló-Mora, R., Rodríguez-Valera, F., and Amann, R. (2000). Extremely halophilic bacteria in crystallizer ponds from solar salterns. *Applied and environmental microbiology*, 66(7):3052–3057.

Baas-Becking, L. G. M. (1934). *Geobiologie; of inleiding tot de milieukunde*. WP Van Stockum & Zoon NV.

Barberán, A. and Casamayor, E. O. (2011). Euxinic freshwater hypolimnia promote bacterial endemicity in continental areas. *Microbial Ecology*, 61(2):465–472.

Barrett, K., Wait, D., and Anderson, W. (2003). Small island biogeography in the gulf of california: lizards, the subsidized island biogeography hypothesis, and the small island effect. *Journal of Biogeography*, 30(10):1575–1581.

Battes, K. P., Cimpean, M., Momeu, L., ȘUTEU, A. M., Pauliuc, G., Stermin, A. N., and David, A. (2019). Species-area relationships for aquatic biota in several shallow lakes from the fizeș valley (transylvania, romania). *North-Western Journal of Zoology*, 15(2).

Bell, T., Ager, D., Song, J.-I., Newman, J. A., Thompson, I. P., Lilley, A. K., and Van der Gast, C. J. (2005). Larger islands house more bacterial taxa. *Science*, 308(5730):1884–1884.

Bolgovics, Á., Ács, É., Várbíró, G., Görgényi, J., and Borics, G. (2016). Species area relationship (sar) for benthic diatoms: a study on aquatic islands. *Hydrobiologia*, 764(1):91–102.

Bradley, J. A., Anesio, A. M., and Arndt, S. (2017). Microbial and biogeochemical dynamics in glacier forefields are sensitive to century-scale climate and anthropogenic change. *Frontiers in Earth Science*, 5:26.

Brown, J. H. and Kodric-Brown, A. (1977). Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology*, 58(2):445–449.

Bueno, C. G. and Moora, M. (2019). How do arbuscular mycorrhizal fungi travel? *New Phytologist*, 222(2):645–647.

Cameron, K. A., Hodson, A. J., and Osborn, A. M. (2012). Structure and diversity of bacterial, eukaryotic and archaeal communities in glacial cryoconite holes from the arctic and the antarctic. *FEMS microbiology ecology*, 82(2):254–267.

Caruso, T., Chan, Y., Lacap, D. C., Lau, M. C. Y., McKay, C. P., and Pointing, S. B. (2011). Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *The ISME Journal*, 5(9):1406–1413.

Cashdan, E. (2014). Biogeography of human infectious diseases: A global historical analysis. *PLoS One*, 9(10):e106752.

Chase, J. M. and Myers, J. A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical transactions of the Royal Society B: Biological sciences*, 366(1576):2351–2363.

Chisholm, R. A., Fung, T., Chimalakonda, D., and O'Dwyer, J. P. (2016). Maintenance of biodiversity on islands. *Proceedings of the Royal Society B: Biological Sciences*, 283(1829):20160102.

Cole, J. J., Pace, M. L., Caraco, N. F., and Steinhart, G. S. (1993). Bacterial biomass and cell size distributions in lakes: more and larger cells in anoxic waters. *Limnology and Oceanography*, 38(8):1627–1632.

Connor, E. F. and McCoy, E. D. (1979). The statistics and biology of the species-area relationship. *The American naturalist*, 113(6):791–833.

Darcy, J. L., Gendron, E., Sommers, P., Porazinska, D. L., and Schmidt, S. K. (2018). Island biogeography of cryoconite hole bacteria in antarctica's taylor valley and around the world. *Frontiers in Ecology and Evolution*, 6:180.

Davis, A. M. and Glick, T. F. (1978). Urban ecosystems and island biogeography. *Environmental Conservation*, 5(4):299–304.

Davison, J., Moora, M., Öpik, M., Ainsaar, L., Ducousso, M., Hiiesalu, I., Jairus, T., Johnson, N., Jourand, P., Kalamees, R., et al. (2018). Microbial island biogeography: isolation shapes the life history characteristics but not diversity of root-symbiotic fungal communities. *The ISME journal*, 12(9):2211–2224.

Delgado-Baquerizo, M., Eldridge, D. J., Hamonts, K., Reich, P. B., and Singh, B. K. (2018). Experimentally testing the species-habitat size relationship on soil bacteria: A proof of concept. *Soil Biology and Biochemistry*, 123:200–206.

Duarte, S., Cássio, F., Pascoal, C., and Bärlocher, F. (2017). Taxa-area relationship of aquatic fungi on deciduous leaves. *PloS one*, 12(7):e0181545.

Dunn, R. R., Davies, T. J., Harris, N. C., and Gavin, M. C. (2010). Global drivers of human pathogen richness and prevalence. *Proceedings of the Royal Society B: Biological Sciences*, 277(1694):2587–2595.

Eadie, J. M., Hurly, T. A., Montgomerie, R. D., and Teather, K. L. (1986). Lakes and rivers as islands: species-area relationships in the fish faunas of ontario. *Environmental Biology of Fishes*, 15(2):81–89.

Elloumi, J., Carrias, J.-F., Ayadi, H., Sime-Ngando, T., Boukhris, M., and Bouaïn, A. (2006). Composition and distribution of planktonic ciliates from ponds of different salinity in the solar saltwork of sfax, tunisia. *Estuarine, Coastal and Shelf Science*, 67(1-2):21–29.

Feinstein, L. M. and Blackwood, C. B. (2012). Taxa–area relationship and neutral dynamics influence the diversity of fungal communities on senesced tree leaves. *Environmental Microbiology*, 14(6):1488–1499.

Fenchel, T. and Finlay, B. J. (2004). The ubiquity of small species: patterns of local and global diversity. *Bioscience*, 54(8):777–784.

Foissner, W. (2006). Biogeography and dispersal of micro-organisms: a review emphasizing protists. *Acta protozoologica*, 45(2):111–136.

Glassman, S. I., Lubetkin, K. C., Chung, J. A., and Bruns, T. D. (2017). The theory of island biogeography applies to ectomycorrhizal fungi in subalpine tree "islands" at a fine scale. *Ecosphere*, 8(2):e01677.

Godon, J.-J., Arulazhagan, P., Steyer, J.-P., and Hamelin, J. (2016). Vertebrate bacterial gut diversity: size also matters. *BMC ecology*, 16(1):12.

Golan, J. J. and Pringle, A. (2017). Long-distance dispersal of fungi. *The Fungal Kingdom*, pages 309–333.

Gooriah, L. D. and Chase, J. M. (2019). Sampling effects drive the species area relationship in lake zooplankton. *Oikos*, 129(1):124–132.

Green, J. and Bohannan, B. J. (2006). Spatial scaling of microbial biodiversity. *Trends in ecology & evolution (Amsterdam)*, 21(9):501–507.

Green, J. L., Holmes, A. J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., Gillings, M., and Beattie, A. J. (2004). Spatial scaling of microbial eukaryote diversity. *Nature*, 432(7018):747–750.

Griffiths, R. I., Thomson, B. C., James, P., Bell, T., Bailey, M., and Whiteley, A. S. (2011). The bacterial biogeography of british soils. *Environmental microbiology*, 13(6):1642–1654.

Grossmann, L., Jensen, M., Pandey, R., Jost, S., Bass, D., Psenner, R., and Boenigk, J. (2016). Molecular investigation of protistan diversity along an elevation transect of alpine lakes. *Aquatic Microbial Ecology*, 78.

Haila, Y. (2002). A conceptual genealogy of fragmentation research: from island biogeography to landscape ecology. *Ecological applications*, 12(2):321–334.

Henebry, M. S. and Cairns Jr, J. (1980). The effect of island size, distance and epicenter maturity on colonization in freshwater protozoan communities. *American Midland Naturalist*, pages 80–92.

Ho, Y.-N., Mathew, D. C., and Huang, C.-C. (2017). Plant-microbe ecology: interactions of plants and symbiotic microbial communities. *Plant Ecology–Traditional Approaches To Recent Trends*, pages 93–119.

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Princeton University Press.

Humayoun, S. B., Bano, N., and Hollibaugh, J. T. (2003). Depth distribution of microbial diversity in mono lake, a meromictic soda lake in california. *Applied and environmental microbiology*, 69(2):1030–1042.

Jean, K., Burnside, W. R., Carlson, L., Smith, K., and Guégan, J.-F. (2016). An equilibrium theory signature in the island biogeography of human parasites and pathogens. *Global Ecology and Biogeography*, 25(1):107–116.

Karatayev, A. Y., Burlakova, L. E., and Dodson, S. I. (2005). Community analysis of belarusian lakes: relationship of species diversity to morphology, hydrology and land use. *Journal of Plankton Research*, 27(10):1045–1053.

Kavazos, C. (2016). Small-scale biogeographic patterns of benthic bacterial and ciliate communities in the saline ponds of lake macleod, north-western australia.

Kleinteich, J., Hildebrand, F., Bahram, M., Voigt, A. Y., Wood, S. A., Jungblut, A. D., Küpper, F. C., Quesada, A., Camacho, A., Pearce, D. A., et al. (2017). Pole-to-pole connections: similarities between arctic and antarctic microbiomes and their vulnerability to environmental change. *Frontiers in Ecology and Evolution*, 5:137.

Kulkarni, S., Dhakar, K., and Joshi, A. (2019). Alkaliphiles: diversity and bioprospection. In *Microbial Diversity in the Genomic Era*, pages 239–263. Elsevier.

Lanzén, A., Simachew, A., Gessesse, A., Chmolowska, D., Jonassen, I., and Øvreås, L. (2013). Surprising prokaryotic and eukaryotic diversity, community structure and biogeography of ethiopian soda lakes. *PloS one*, 8(8):e72577.

Lasken, R. S. and McLean, J. S. (2014). Recent advances in genomic dna sequencing of microbial species from single cells. *Nature Reviews Genetics*, 15(9):577–584.

Lepère, C., Domaizon, I., Taïb, N., Mangot, J.-F., Bronner, G., Boucher, D., and Debroas, D. (2013). Geographic distance and ecosystem size determine the distribution of smallest protists in lacustrine ecosystems. *FEMS Microbiology Ecology*, 85(1):85–94.

Li, S.-p., Wang, P., Chen, Y., Wilson, M. C., Yang, X., Ma, C., Lu, J., Chen, X.-y., Wu, J., Shu, W.-s., et al. (2020). Island biogeography of soil bacteria and fungi: similar patterns, but different mechanisms. *The ISME Journal*, pages 1–11.

Locey, K. J. (2010). Synthesizing traditional biogeography with microbial ecology: the importance of dormancy: Synthesizing traditional biogeography with microbial ecology. *Journal of biogeography*, pages no–no.

Lomolino, M. and Weiser, M. (2001). Towards a more general species-area relationship: diversity on all islands, great and small. *Journal of biogeography*, pages 431–445.

Lomolino, M. V. (1982). Species-area and species-distance relationships of terrestrial mammals in the thousand island region. *Oecologia*, 54(1):72–75.

MacArthur, R. H. (1967). *The theory of island biogeography*. Monographs in population biology ; 1. Princetown University Press, Princetown.

MacDonald, Z. G., Anderson, I. D., Acorn, J. H., and Nielsen, S. E. (2018). The theory of island biogeography, the sample-area effect, and the habitat diversity hypothesis: complementarity in a naturally fragmented landscape of lake islands. *Journal of Biogeography*, 45(12):2730–2743.

Malard, L. A. and Pearce, D. A. (2018). Microbial diversity and biogeography in arctic soils. *Environmental microbiology reports*, 10(6):611–625.

Martiny, J. B. H., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C. R., et al. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4(2):102–112.

McCormick, P., Pratt, J., Jenkins, D., and Cairns Jr, J. (1988). A comparison of protozoan, algal, and metazoan colonization of artificial substrates of differing size. *Transactions of the American Microscopical Society*, pages 259–268.

Nicolaides, C., Avraam, D., Cueto-Felgueroso, L., González, M. C., and Juanes, R. (2020). Hand-hygiene mitigation strategies against global disease spreading through the air transportation network. *Risk Analysis*, 40(4):723–740.

Olive, M., Gan, C., Carratalà, A., and Kohn, T. (2020). Control of waterborne human viruses by indigenous bacteria and protists is influenced by temperature, virus type, and microbial species. *Applied and Environmental Microbiology*, 86(3).

Orrock, J. L., Allan, B. F., Drost, C. A., Rohani, A. E. P., and Bronstein, E. J. L. (2011). Biogeographic and ecological regulation of disease: Prevalence of sin nombre virus in island mice is related to island area, precipitation, and predator richness. *The American naturalist*, 177(5):691–697.

Pasztaleniec, A. and Poniewozik, M. (2010). Phytoplankton based assessment of the ecological status of four shallow lakes (eastern poland) according to water framework directive–a comparison of approaches. *Limnologica-Ecology and Management of Inland Waters*, 40(3):251–259.

Patrick, R. (1967). The effect of invasion rate, species pool, and size of area on the structure of the diatom community. *Proceedings of the National Academy of Sciences of the United States of America*, 58(4):1335.

Peay, K. G., Bruns, T. D., Kennedy, P. G., Bergemann, S. E., and Garbelotto, M. (2007). A strong species–area relationship for eukaryotic soil microbes: island size matters for ectomycorrhizal fungi. *Ecology letters*, 10(6):470–480.

Pepper, I. (2019). Biotic characteristics of the environment. In *Environmental and Pollution Science*, pages 61–87. Elsevier.

Prevost-Boure, N. C., Christen, R., Dequiedt, S., Mougel, C., Lelievre, M., Jolivet, C., Shahbazkia, H. R., Guillou, L., Arrouays, D., and Ranjard, L. (2011). Validation and application of a pcr primer set to quantify fungal communities in the soil environment by real-time quantitative pcr. *PloS one*, 6(9):e24166.

Price, J. P. and Wagner, W. L. (2011). A phylogenetic basis for species–area relationships among three pacific island floras. *American Journal of Botany*, 98(3):449–459.

Reche, I., Pulido-Villena, E., Morales-Baquero, R., and Casamayor, E. O. (2005). Does ecosystem size determine aquatic bacterial richness? *Ecology*, 86(7):1715–1722.

Rengefors, K., Laybourn-Parry, J., Logares, R., Marshall, W. A., and Hansen, G. (2008). Marine-derived dinoflagellates in antarctic saline lakes: Community composition and annual dynamics 1*[link]. *Journal of phycology*, 44(3):592–604.

Rengefors, K., LOGARES, R., and LAYBOURN-PARRY, J. (2012). Polar lakes may act as ecological islands to aquatic protists. *Molecular ecology*, 21(13):3200–3209.

Ricklefs, R. E. and Lovette, I. J. (1999). The roles of island area per se and habitat diversity in the species–area relationships of four lesser antillean faunal groups. *Journal of Animal Ecology*, 68(6):1142–1160.

Rivett, D. W. and Bell, T. (2018). Abundance determines the functional role of bacterial phylotypes in complex communities. *Nature microbiology*, 3(7):767–772.

Rosenzweig, M. L. et al. (1995). *Species diversity in space and time*. Cambridge University Press.

Rosindell, J., Wong, Y., and Etienne, R. S. (2008). A coalescence approach to spatial neutral ecology. *Ecological Informatics*, 3(3):259–271.

Rosselli, R., Fiamma, M., Deligios, M., Pintus, G., Pellizzaro, G., Canu, A., Duce, P., Squartini, A., Muresu, R., and Cappuccinelli, P. (2015). Microbial immigration across the mediterranean via airborne dust. *Scientific reports*, 5:16306.

Samson, F. B. (1980). Island biogeography and the conservation of prairie birds. In *Proceedings of the North American Prairie Conference*, volume 7, pages 293–305.

Sfenthourakis, S. and Triantis, K. A. (2009). Habitat diversity, ecological requirements of species and the small island effect. *Diversity and Distributions*, 15(1):131–140.

Simberloff, D. S. (1974). Equilibrium theory of island biogeography and ecology. *Annual review of Ecology and Systematics*, 5(1):161–182.

Stegen, J. C., Lin, X., Konopka, A. E., and Fredrickson, J. K. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME Journal*, 6(9):1653–1664.

Stewart, K. and Schlichting Jr, H. (1966). Dispersal of algae and protozoa by selected aquatic insects. *The Journal of Ecology*, pages 551–562.

Stibal, M., Bradley, J. A., Edwards, A., Hotaling, S., Zawierucha, K., Rosvold, J., Lutz, S., Cameron, K. A., Mikucki, J. A., Kohler, T. J., et al. (2020). Glacial ecosystems are essential to understanding biodiversity responses to glacier retreat. *Nature Ecology & Evolution*, 4(5):686–687.

Stomp, M., Huisman, J., Mittelbach, G. G., Litchman, E., and Klausmeier, C. A. (2011). Large-scale biodiversity patterns in freshwater phytoplankton. *Ecology (Durham)*, 92(11):2096–2107.

Triantis, K., Vardinoyannis, K., Tsolaki, E., Botsaris, I., Lika, K., and Mylonas, M. (2006). Re-approaching the small island effect. *Journal of Biogeography*, 33(5):914–923.

Triantis, K. A., Mylonas, M., and Whittaker, R. J. (2008). Evolutionary species–area curves as revealed by single-island endemics: insights for the inter-provincial species–area relationship. *Ecography*, 31(3):401–407.

Turner, S., Mikutta, R., Meyer-Stüve, S., Guggenberger, G., Schaarschmidt, F., Lazar, C. S., Dohrmann, R., and Schippers, A. (2017). Microbial community dynamics in soil depth profiles over 120,000 years of ecosystem development. *Frontiers in microbiology*, 8:874.

Van Der Gast, C. J., Jefferson, B., Reid, E., Robinson, T., Bailey, M. J., Judd, S. J., and Thompson, I. P. (2006). Bacterial diversity is determined by volume in membrane bioreactors. *Environmental microbiology*, 8(6):1048–1055.

Van Der Gast, C. J., Lilley, A. K., Ager, D., and Thompson, I. P. (2005). Island size and bacterial diversity in an archipelago of engineering machines. *Environmental microbiology*, 7(8):1220–1226.

Várbíró, G., Görgényi, J., Tóthmérész, B., Padisák, J., Hajnal, É., and Borics, G. (2017). Functional redundancy modifies species–area relationship for freshwater phytoplankton. *Ecology and evolution*, 7(23):9905–9913.

Wildman, H. (1987). Fungal colonization of resources in soil—an island biogeographical approach. *Transactions of the British Mycological Society*, 88(3):291–297.

Woodcock, S., Curtis, T. P., Head, I. M., Lunn, M., and Sloan, W. T. (2006). Taxa–area relationships for microbes: the unsampled and the unseen. *Ecology letters*, 9(7):805–812.

Wright, D. H. (1983). Species-energy theory: an extension of species-area theory. *Oikos*, pages 496–506.

Wurzbacher, C. M., Bärlocher, F., and Grossart, H.-P. (2010). Fungi in lake ecosystems. *Aquatic Microbial Ecology*, 59(2):125–149.

Zacharias, D. and Brandes, D. (1990). Species area-relationships and frequency—floristical data analysis of 44 isolated woods in northwestern germany. *Vegetatio*, 88(1):21–29.

Supplementary Material

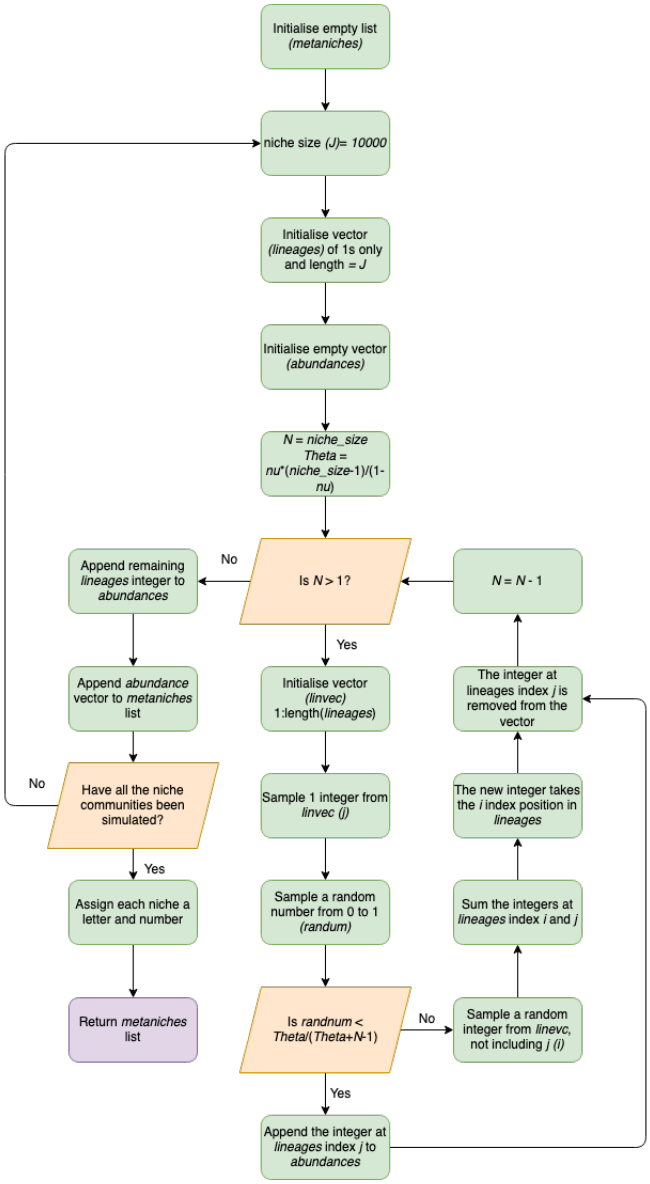
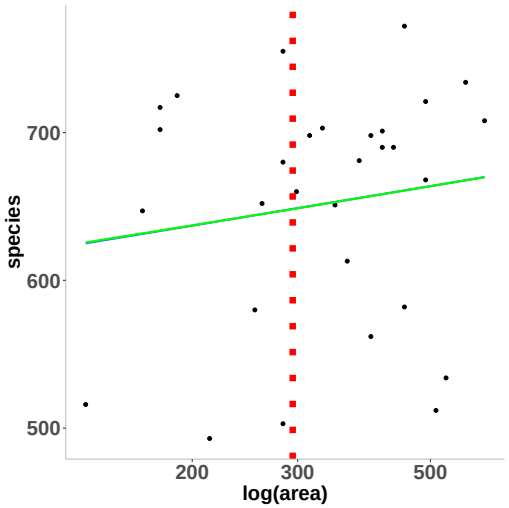
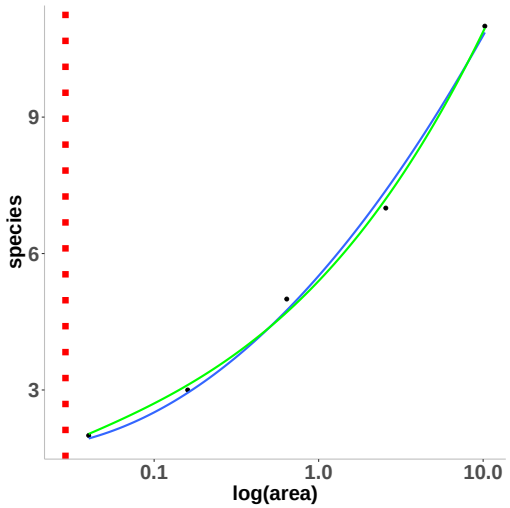


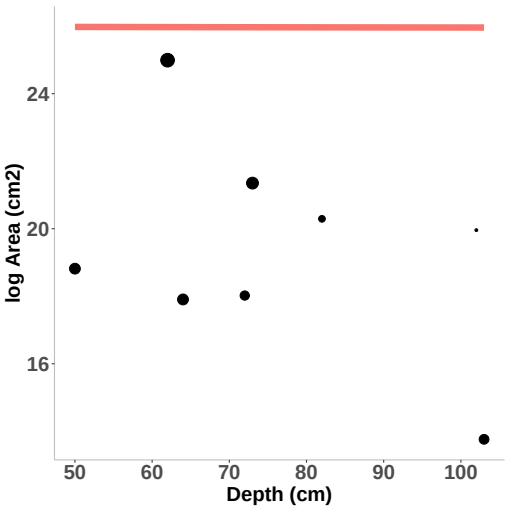
Figure 7.1: Flowchart of metacommunity function



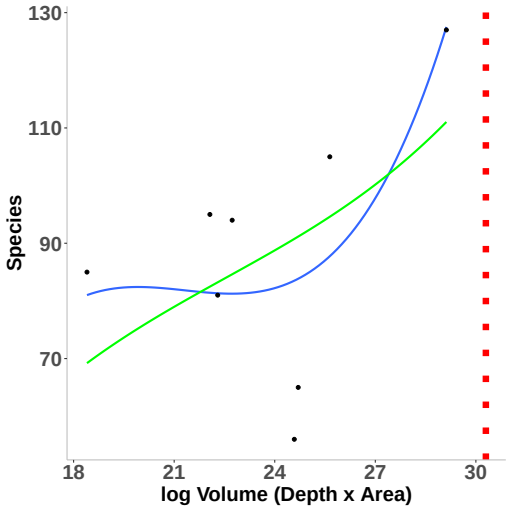
(a) Dataset 4, bacteria in Antarctic Croconite Holes



(b) Dataset 54, fungi in a soil based laboratory experiment



(c) Dataset 22, benthic bacteria in saline lakes (log area plotted with depth)



(d) Dataset 22, benthic bacteria in saline lakes (log volume plotted with OTU richness)

Figure 7.2: Selection of plots of datasets that failed to be fit using either of the three model variants (Classic, Depth, Perimeter) or the power-law model, where the blue line is the model fit, the green line is the power-law model fit and the red dotted line is the A_{crit} estimation. A) Dataset 4, bacteria in cryoconite holes with the Classic Model (R_2=0.02, adjusted R_2=-0.13, θ=29, m_0=0.208, K=354, A_{crit}=294 cm^2). B) Fungi in soil based laboratory experiment plotted with the Perimeter Model (R_2=0.099, adjusted R_2=Inf, θ=6, m_0=6.24 x 10^{-6}, K=1, A_{crit}=2.89 x 10^{-2} cm^2). C) Benthic bacteria in saline lakes plotted with the Depth Model (R_2=0.51, adjusted R_2=-0.15, θ=14, m_0=1.98 x 10^{-16}, K=81, A_{crit}=1.89 x 10^{11} cm^2)

Table 7.1: Summary of Datasets

Study and Dataset ID	**Author and Year**	**Habitat**	**Taxa**
Study 1, Datasets 1 & 2	(Li et al., 2020)	Terrestrial	Bacteria & Fungi
Study 2, Dataset 3	(Battes et al., 2019)	Lacustrine	Algae
Study 3, Datasets 4-15	(Darcy et al., 2018)	Lacustrine	Bac, Alg, Fungi & Proto
Study 4, Datasets 16 & 17	(Delgado-Baquerizo et al., 2018)	Terrestrial	Bacteria
Study 5, Dataset 18	(Davison et al., 2018)	Terrestrial	Fungi
Study 6, Datasets 19 & 20	(Glassman et al., 2017)	Plant	Fungi
Study 7, Dataset 21	(Várbíró et al., 2017)	Lacustrine	Algae
Study 8, Datasets 22 & 23	(Kavazos, 2016)	Lacustrine	Bac & Proto
Study 9, Dataset 24	(Bolgovics et al., 2016)	Lacustrine	Algae
Study 10, Datasets 25-27	(Grossmann et al., 2016)	Lacustrine	Alg, Proto & Fun
Study 11, Datasets 28-32	(Jean et al., 2016)	Terrestrial	Bac, Path, Fun & Proto
Study 12, Dataset 33	(Cashdan, 2014)	Terrestrial	Pathogens
Study 13, Dataset 34	(Lepère et al., 2013)	Lacustrine	Protozoa
Study 14, Datasets 35 & 36	(Lanzén et al., 2013)	Lacustrine	Bacteria & Archaea
Study 15, Dataset 37	(Rengefors et al., 2012)	Lacustrine	Algae
Study 16, Datasets 38-39	(Feinstein and Blackwood, 2012)	Plant	Fungi
Study 17, Dataset 40	(Stomp et al., 2011)	Lacustrine	Algae
Study 18, Dataset 41	(Orrock et al., 2011)	Terrestrial	Pathogens
Study 19, Datasets 42 & 43	(Barberán and Casamayor, 2011)	Lacustrine	Bacteria
Study 20, Dataset 44	(Peay et al., 2007)	Plant	Fungi
Study 21, Dataset 45	(Van Der Gast et al., 2006)	Machine	Bacteria
Study 22, Dataset 46	(Bell et al., 2005)	Lacustrine	Bacteria
Study 23, Dataset 47	(Reche et al., 2005)	Lacustrine	Bacteria
Study 24, Datasets 48-50	(Van Der Gast et al., 2005)	Machine	Bacteria
Study 25, Dataset 51	(Karatayev et al., 2005)	Lacustrine	Algae
Study 26, Datasets 52 & 53	(McCormick et al., 1988)	Riverine	Protozoa and Algae
Study 27, Dataset 54	(Wildman, 1987)	Terrestrial	Fungi
Study 28, Dataset 55	(Henebry and Cairns Jr, 1980)	Lacustrine	Protozoa
Study 29, Datasets 56 & 57	(Patrick, 1967)	Riverine	Algae

Table 7.2: Rho Estimation Methods (S & D ID = Study & Dataset ID)

Study/Data	**Taxa**	**Habitat**	ρ **(cm^3)**	**Method**
S1, D1	Bacteria	Soil	$1.48\text{x}10^{12}$	Gene seq num from paper
S1, D2	Fungi	Soil	$7.41\text{x}10^{3}$	Gene seq num from paper
S2, D3	Algae	Freshwater	$3.56\text{x}10^{3}$	Proxy (Pasztaleniec and Poniewozik, 2010)
S3, D4-6	Bacteria	Cryo Holes	$4.50\text{x}10^{4}$	Proxy (Cameron et al., 2012)
S3, D7-9	Algae	Cryo Holes	1	Gene seq num from paper
S3, D10-12	Fungi	Cryo Holes	1	Gene seq num from paper
S3, D13-15	Protozoa	Cryo Holes	1	Gene seq num from paper
S4, D16	Bacteria	Soil	$3.50\text{x}10^{12}$	Gene seq num from paper
S4, D17	Bacteria	Soil	$1.48\text{x}10^{13}$	Gene seq num from paper
S5, D18	Fungi	Soil	1	1 as area includes entire island
S6, D19 & 20	Fungi	Soil	$2.98\text{x}10^{3}$	Gene seq num from paper
S7, D21	Algae	Freshwater	$3.56\text{x}10^{3}$	Same proxy as D3
S8, D22	Bacteria	Sal Water	$3.00\text{x}10^{6}$	Proxy (Antón et al., 2000)
S8, D23	Protozoa	Sal Water	$1.84\text{x}10^{2}$	Proxy (Elloumi et al., 2006)
S9, D24	Algae	Freshwater	$3.56\text{x}10^{3}$	Same proxy as D3
S10, D25	Algae	Freshwater	$3.56\text{x}10^{3}$	Same proxy as D3
S10, D26	Protozoa	Freshwater	$1.92\text{x}10^{3}$	Proxy (Olive et al., 2020)
S10, D27	Fungi	Freshwater	1	Proxy (Wurzbacher et al., 2010)
S11, D28-32	Bac, Pat, Fun, Pro	Hosts	1	1 as area includes entire island
S12, D33	Pathogens	Hosts	1	1 as area includes entire island
S13, D34	Protozoa	Freshwater	$5.72\text{x}10^{3}$	Nanoflag count taken from paper
S14, D35	Archaea	Sal Water	$1.00\text{x}10^{7}$	Book (Kulkarni et al., 2019)
S14, D36	Bacteria	Sal Water	$6.00\text{x}10^{6}$	Proxy (Humayoun et al., 2003)
S15, D37	Algae	Sal Water	$4.85\text{x}10^{3}$	Proxy (Rengefors et al., 2008)
S16, D38-39	Fungi	Plant	10	No proxy found so 10 est
S17, D40	Algae	Freshwater	$3.56\text{x}10^{3}$	Same proxy as D3
S18, D41	Pathogens	Hosts	1	1 as area includes entire island
S19, D42 & 43	Bacteria	Freshwater	$1.16\text{x}10^{7}$	Proxy (Cole et al., 1993)
S20, D44	Fungi	Soil	$6.90\text{x}10^{6}$	Proxy (Prevost-Boure et al., 2011)
S21, D45	Bacteria	Machine	$2.29\text{x}10^{10}$	Cell abund taken from paper
S22, D46	Bacteria	Tree Holes	$4.90\text{x}10^{5}$	Proxy (Rivett and Bell, 2018)
S23, D47	Bacteria	Freshwater	$1.16\text{x}10^{7}$	Same proxy as D43
S24, D48-50	Bacteria	Machine	$2.29\text{x}10^{10}$	Same proxy as D46
S25, D51	Algae	Freshwater	$3.56\text{x}10^{3}$	Same proxy as D3
S26, D52	Protozoa	River	$5.72\text{x}10^{3}$	Same proxy as D34
S26, D53	Algae	River	$3.56\text{x}10^{3}$	Same proxy as D3
S27, D54	Fungi	Soil	$1.00\text{x}10^{5}$	Book (Pepper, 2019)
S28, D55	Protozoa	Freshwater	$5.72\text{x}10^{3}$	Same proxy as D34
S29, D56 & 57	Algae	River	$3.56\text{x}10^{3}$	Same proxy as D3

Table 7.3: Results of successful Power-Law Model fitting to the 24 positive TAR datasets with R^2, adjusted R^2, z values (model exponent), c values (model constant) (S & D ID = Study & Dataset ID)

S & D ID	**Author & Year**	**R^2**	**Adj R^2**	z	c
S1, D1	(Li et al., 2020)	0.56	0.49	0.09	1294.63
S1, D2	(Li et al., 2020)	0.60	0.53	0.10	332.03
S3, D6	(Darcy et al., 2018)	0.23	0.10	0.32	70.02
S3, D11	(Darcy et al., 2018)	0.27	0.15	0.42	4.68
S4, D16	(Delgado-Baquerizo et al., 2018)	0.82	0.75	0.11	340.96
S4, D17	(Delgado-Baquerizo et al., 2018)	0.70	0.54	0.08	451.61
S6, D19	(Glassman et al., 2017)	0.23	0.02	0.27	0.48
S6, D20	(Glassman et al., 2017)	0.32	0.13	0.16	1.93
S7, D21	(Várbíró et al., 2017)	0.02	0.0001	0.01	23.10
S9, D24	(Bolgovics et al., 2016)	0.62	0.48	0.05	32.25
S10, D25	(Grossmann et al., 2016)	0.38	0.28	0.06	48.28
S10, D26	(Grossmann et al., 2016)	0.29	0.17	0.14	0.29
S11, D28	(Jean et al., 2016)	0.23	0.18	0.01	39.29
S11, D29	(Jean et al., 2016)	0.34	0.30	0.02	19.31
S11, D30	(Jean et al., 2016)	0.24	0.19	0.02	4.27
S11, D31	(Jean et al., 2016)	0.35	0.31	0.03	4.60
S11, D32	(Jean et al., 2016)	0.33	0.29	0.05	4.26
S20, D44	(Peay et al., 2007)	0.87	0.78	0.18	0.64
S21, D45	(Van Der Gast et al., 2006)	0.94	0.82	0.27	0.88
S22, D46	(Bell et al., 2005)	0.46	0.38	0.33	1.74
S24, D48	(Van Der Gast et al., 2005)	0.73	0.63	0.36	1.25
S24, D49	(Van Der Gast et al., 2005)	0.84	0.76	0.32	1.80
S24, D50	(Van Der Gast et al., 2005)	0.80	0.73	0.36	1.34
S25, D51	(Karatayev et al., 2005)	0.10	0.09	0.11	19.00

Table 7.4: Results of Power-Law Model AIC score - Classic, Depth and Perimeter AIC scores. There must be a difference of at least 2 to be statistically significant (positive results favour the Classic, Depth and Perimeter Models, negative results favour the Power-Law Model) (S & D ID = Study & Dataset ID)

S & D ID	**Author & Year**	**Classic**	**Depth**	**Perimeter**
S1, D1	(Li et al., 2020)	7.25	7.25	6.91
S1, D2	(Li et al., 2020)	10.37	10.18	9.49
S3, D6	(Darcy et al., 2018)	0.23	0.23	0.08
S3, D11	(Darcy et al., 2018)	0.005	0.02	0.02
S4, D16	(Delgado-Baquerizo et al., 2018)	0.005	0.005	0.005
S4, D17	(Delgado-Baquerizo et al., 2018)	0.002	0.002	0.002
S6, D19	(Glassman et al., 2017)	0.03	0.03	0.01
S6, D20	(Glassman et al., 2017)	0.72	0.72	0.52
S7, D21	(Várbíró et al., 2017)	1.20	1.22	1.03
S9, D24	(Bolgovics et al., 2016)	2.64	2.62	3.33
S10, D25	(Grossmann et al., 2016)	0.91	1.03	0.92
S10, D26	(Grossmann et al., 2016)	0.05	0.11	0.04
S11, D28	(Jean et al., 2016)	0.49	0.49	-0.61
S11, D29	(Jean et al., 2016)	0.33	0.34	0.11
S11, D30	(Jean et al., 2016)	2.77	2.78	3.16
S11, D31	(Jean et al., 2016)	0.76	0.75	-0.05
S11, D32	(Jean et al., 2016)	2.22	2.21	0.75
S20, D44	(Peay et al., 2007)	-0.05	-0.05	-0.27
S21, D45	(Van Der Gast et al., 2006)	2.67	2.67	2.25
S22, D46	(Bell et al., 2005)	0.45	1.34	0.17
S24, D48	(Van Der Gast et al., 2005)	2.84	2.87	0.47
S24, D49	(Van Der Gast et al., 2005)	4.28	4.33	1.80
S24, D50	(Van Der Gast et al., 2005)	3.38	3.42	0.59
S25, D51	(Karatayev et al., 2005)	0.23	-1.18	0.17

Table 7.5: Best-fit results of the Classic, Depth and Perimeter Model fittings, with best-fit parameters (θ, m_0, K), A_{crit} and best-fit model (S & D ID = Study & Dataset ID)

S & D ID	R^2	Adj R^2	θ	m_0	K	A_{crit}	Best-fit Model
S1, D1	0.65	0.60	1.73×10^{3}	3.67×10^{-5}	1	5.00×10^{-2}	All
S1, D2	0.72	0.67	3.56×10^{2}	1.80×10^{-8}	1	4.21×10^{2}	Classic & Depth
S3, D6	0.23	0.11	1.60×10^{5}	2.43×10^{-5}	204	1.98×10^{2}	All
S3, D11	0.27	0.15	5.89×10^{4}	5.64×10^{-1}	25	1.92×10^{2}	All
S4, D16	0.82	0.75	1.42×10^{2}	1.53×10^{-12}	315	5.52x10	All
S4, D17	0.70	0.54	1.08×10^{2}	5.12×10^{-13}	424	5.02×10^{2}	All
S6, D19	0.23	0.02	13	5.88×10^{-7}	4	1.91×10^{4}	All
S6, D20	0.34	0.17	2	5.09×10^{-8}	1	4.91×10^{2}	Classic & Depth
S7, D21	0.02	0.01	1	1.39×10^{-4}	21	1.03×10^{20}	Classic & Depth
S9, D24	0.69	0.58	16	2.79×10^{-8}	60	4.83×10^{10}	Perimeter
S10, D25	0.41	0.31	10	1.19×10^{-6}	2	2.38x10	Depth
S10, D26	0.29	0.18	1	2.86×10^{-13}	3	1.33×10^{9}	Classic & Depth
S11, D28	0.23	0.18	1	1.70×10^{-12}	51	1.07×10^{48}	Classic & Depth
S11, D29	0.34	0.30	1	3.75×10^{-6}	28	1.89×10^{31}	All
S11, D30	0.28	0.23	2.43×10^{3}	5.87×10^{-10}	7	8.52×10^{16}	Perimeter
S11, D31	0.36	0.32	1	3.31×10^{-13}	9	2.23×10^{24}	Classic
S11, D32	0.35	0.31	3	1.44×10^{-15}	17	4.24×10^{16}	Classic & Depth
S20, D44	0.85	0.77	5	6.15×10^{-11}	2	4.08×10^{4}	All
S21, D45	0.96	0.88	9	4.97×10^{-16}	7	2.62×10^{4}	Classic & Depth
S22, D46	0.49	0.40	8	3.75×10^{-9}	6	2.19×10^{2}	Depth
S24, D48	0.78	0.69	7	7.97×10^{-14}	1	1.95x10	Classic & Depth
S24, D49	0.89	0.83	6	1.15×10^{-13}	1	1.38x10	Classic & Depth
S24, D50	0.84	0.78	7	8.74×10^{-14}	1	1.78x10	Classic & Depth
S25, D51	0.10	0.09	12	3.25×10^{-5}	28	4.46×10^{3}	All