Analysis of relative abundances on environmental gradients (#27313)

First submission

Editor guidance

Please submit by **13 May 2018** for the benefit of the authors (and your $200 publishing discount).

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data. Download from the materials page.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

6 Figure file(s)
2 Latex file(s)
1 Raw data file(s)
1 Other file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections.
Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.
- Speculation is welcome, but should be identified as such.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Analysis of relative abundances on environmental gradients

Fiona Chong [1,2] , **Matthew Spencer** Corresp. [2]

[1] TROPIMUNDO Erasmus Mundus Masters Course in Tropical Biodiversity and Ecosystems
[2] School of Environmental Sciences, University of Liverpool, Liverpool, United Kingdom

Corresponding Author: Matthew Spencer
Email address: m.spencer@liverpool.ac.uk

Ecologists often analyze relative abundances, which are compositions (sets of non-negative numbers with a fixed sum). However, they have made surprisingly little use of recent advances in the field of compositional data analysis. Compositions form a vector space in which addition and scalar multiplication are replaced by operations known as perturbation and powering. This algebraic structure makes it easy to understand how relative abundances change along environmental gradients. We illustrate this with an analysis of changes in hard-substrate marine communities along a depth gradient. We show how the algebra of compositions can be used to understand patterns in dissimilarity. We use the calculus of simplex-valued functions to estimate rates of change, and to summarize the structure of the community over a vertical slice. We discuss the benefits of the compositional approach in the interpretation and visualization of relative abundance data.

Analysis of relative abundances on environmental gradients

Fiona Chong[*], Matthew Spencer[†]

School of Environmental Sciences, University of Liverpool,

Liverpool, L69 3GP, UK

April 26, 2018

[*]Current affiliation: TROPIMUNDO Erasmus Mundus Masters Course in Tropical Biodiversity and Ecosystems. Email: fionachong104@gmail.com

[†]Corresponding author: m.spencer@liverpool.ac.uk

Abstract

Ecologists often analyze relative abundances, which are compositions (sets of non-negative numbers with a fixed sum). However, they have made surprisingly little use of recent advances in the field of compositional data analysis. Compositions form a vector space in which addition and scalar multiplication are replaced by operations known as perturbation and powering. This algebraic structure makes it easy to understand how relative abundances change along environmental gradients. We illustrate this with an analysis of changes in hard-substrate marine communities along a depth gradient. We show how the algebra of compositions can be used to understand patterns in dissimilarity. We use the calculus of simplex-valued functions to estimate rates of change, and to summarize the structure of the community over a vertical slice. We discuss the benefits of the compositional approach in the interpretation and visualization of relative abundance data.

1 Introduction

Ecologists often analyze relative abundance data. These are sets of non-negative numbers with a fixed sum (typically 1 or 100), and are therefore examples of compositional data. Compositional data present some special challenges, arising from their constrained multivariate nature, including the absence of an interpretable covariance structure and the inappropriateness of simple parametric models (Aitchison, 1986, chapter 3). Many of these challenges have been addressed in the last few decades, leading to a coherent set of principles for the analysis of compositional data (Pawlowsky-Glahn and Buccianti, 2011). Although some important work on the principals of compositional data analysis was ecological (e.g. Mosimann, 1962; Martin and Mosimann, 1965; Billheimer et al., 2001), ecologists have made surprisingly little use of recent advances in the field (exceptions include Jackson, 1997; Gross and Edmunds, 2015; Yuan et al., 2016). For example, Legendre and Legendre (2012), one of the most important textbooks on analysis of community ecological data, does not cite any papers on compositional data analysis.

The key principle in compositional data analysis is scale invariance (Aitchison, 1992). This means that if $\mathbf{x}$ is a set of abundances, then $a\mathbf{x}$ is equivalent to $\mathbf{x}$, for any positive real number a. To an ecologist, this means treating two communities as equivalent if they have the same relative abundances but different total abundances. It is straightforward to show, using the scale invariance principle, that any meaningful function of a composition can be expressed in terms of ratios of relative abundances (Aitchison, 1992). In addition, in most situations, subcompositional coherence is important. Suppose that two scientists are studying the same community, but one measures the abundances of all taxa, while the other measures the abundances of only some taxa. Subcompositional coherence is the requirement that their results should agree for the subset of taxa measured by both (Aitchison, 1992). Ecologists should care about subcompositional coherence because they are almost always studying only a subset of the taxa present in a community. These seemingly obvious principles can lead to a coherent method of manipulating relative abundance data. In order to understand why this is important, we need to think a little about abstract algebra.

Ecologists make frequent use of some aspects of vector algebra in $\mathbb{R}^n$, a mathematical system which emerged gradually in the late 19th century, primarily driven by the need to solve three-dimensional physical problems in fields such as electricity (Crowe, 1994). In community ecology, the main application of vectors is the representation and manipulation of the abundances of more than one species simultaneously. For such vectors, the operations of addition and scalar multiplication have obvious biological meanings. However, ecologists make little explicit use of the more abstract concept of a real vector space, defined only by the axioms it satisfies, rather than the types of objects involved. This concept, now important in many areas of mathematics, emerged around the same time as vector algebra (Dorier, 1995). A real vector space is a set of objects (vectors) with a binary operation ('addition'), and a scalar operation ('scalar multiplication') by which real numbers act on the objects (Fraleigh and Beauregard, 1995, section 3.1). The addition operation satisfies the familiar algebraic axioms of closure, associativity, commutativity, and the existence of an identity element and of inverse elements. The scalar multiplication operation satisfies the familiar algebraic axioms of closure, distributivity, associativity, and has 1 as the multiplicative identity. This more general concept might be useful in ecology because the ordinary definitions of addition and scalar multiplication for Euclidean vectors do not satisfy the vector space axioms when applied to relative abundances. For example, let $\mathbf{a} = (1/3, 1/3, 1/3)^T$ be a relative abundance vector (throughout, we work with column vectors, so T denotes transpose). Then neither $\mathbf{a}+\mathbf{a}$ nor $2\mathbf{a}$ is a relative abundance vector, so the axiom of closure is not satisfied.

There are in fact operations corresponding to addition and scalar multiplication that make sense for compositions. For a vector of s positive numbers $\mathbf{x}$, let the closure $\mathscr{C}(\mathbf{x})$ of $\mathbf{x}$ be defined by

$$\mathscr{C}(\mathbf{x}) = \frac{1}{\sum_{i=1}^{s} x_i}\mathbf{x}$$

(Aitchison, 1986, p. 31). Now if $\mathbf{a}, \mathbf{b}$ are s-part compositions, then let the perturbation $\oplus$ of $\mathbf{b}$ by $\mathbf{a}$ be defined by

$$\mathbf{a} \oplus \mathbf{b} = \mathscr{C}(a_1 b_1, a_2 b_2, \ldots, a_s b_s)$$

(Aitchison, 1986, p. 42). Also, if $a > 0$, then the powering $\odot$ of $\mathbf{b}$ by a is defined by

$$a \odot \mathbf{b} = \mathscr{C}(b_1^a, b_2^a, \ldots, b_s^a)$$

(Aitchison, 1986, p. 120). The set of s-part compositions with the binary operation of perturbation (corresponding to 'addition') and the scalar operation of powering (corresponding to 'scalar multiplication') satisfies the vector space axioms (Billheimer et al., 2001). Now for any two compositions $\mathbf{a}$ and $\mathbf{b}$, we can transform $\mathbf{a}$ into $\mathbf{b}$ by the closure of the unequal scaling

$$\begin{aligned}\mathbf{b} &= \mathscr{C}\left(\frac{b_1}{a_1}a_1, \frac{b_2}{a_2}a_2, \ldots, \frac{b_s}{a_s}a_s\right) \\ &= \mathbf{b} \oplus ((-1) \odot \mathbf{a}) \oplus \mathbf{a}.\end{aligned}$$

We can thus define the compositional difference $\mathbf{b} \ominus \mathbf{a}$ as

$$\mathbf{b} \ominus \mathbf{a} = \mathbf{b} \oplus ((-1) \odot \mathbf{a}) = \mathscr{C}\left(\frac{b_1}{a_1}, \frac{b_1}{a_1}, \ldots, \frac{b_s}{a_s}\right). \quad (1)$$

This is the only way to define the difference between two compositions, under either one of two additional conditions (Aitchison, 1992). The first and most important for ecology is that the answer must not depend on changes of units for individual components, or equivalently, must not change if detection probabilities differ among taxa. The second is that the ith component of the transformation from one composition to another must depend only on the ith component of the compositions. This is desirable because we would like to identify components of change in relative abundances associated with particular taxa. Adoption of either of these conditions leads immediately to the idea that any measure of dissimilarity between two relative abundance vectors must be perturbation invariant, i.e. it must depend only on the compositional difference between them, defined by the ratios of relative abundances of corresponding taxa.

A common approach to studying variation among communities is to compute some measure d of dissimilarity between pairs of communities, and then carry out graphical or numerical analyses of the resulting distance matrix (Legendre and Legendre, 2012, chapter 7). This has the potential to mislead if the measure of dissimilarity is not perturbation invariant. Consider a series of J communities along an environmental gradient, with compositions $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots, \boldsymbol{\rho}_J$. Suppose that the communities are spaced so that the ratios of relative abundances for each species in successive communities are constant, in other words $\rho_{i,j}/\rho_{i,j+1} = a_i$, where a_i is a constant, for each species $i \in \{1, 2, \ldots, s\}$ and for each community $j \in \{1, 2, \ldots, J-1\}$. Since relative abundances, by definition, are meaningful only in relative terms, there has been the same amount of change in the relative abundance of each species between each pair of communities $j, j+1$. This implies that a meaningful measure of dissimilarity between adjacent pairs of communities must be constant. From the definition of compositional difference (Equation 1), $\boldsymbol{\rho}_{j+1} \ominus \boldsymbol{\rho}_j = \mathbf{a}$, where $\mathbf{a} = (a_1, a_2, \ldots, a_s)$ is a constant perturbation. Then we can write $\boldsymbol{\rho}_{j+1} = \mathbf{a} \oplus \boldsymbol{\rho}_j$, and $\boldsymbol{\rho}_{j+2} = \mathbf{a} \oplus \boldsymbol{\rho}_{j+1}$, and we require that $d(\boldsymbol{\rho}_j, \boldsymbol{\rho}_{j+1}) = d(\mathbf{a} \oplus \boldsymbol{\rho}_j, \mathbf{a} \oplus \boldsymbol{\rho}_{j+1})$. In general, any meaningful dissimilarity measure d for compositions must satisfy the perturbation invariance property $d(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = d(\mathbf{a} \oplus \boldsymbol{\rho}_1, \mathbf{a} \oplus \boldsymbol{\rho}_2)$ for all compositions $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \mathbf{a}$. Most of the popular measures of community dissimilarity are not perturbation invariant, and are therefore misleading. For example, let

$\boldsymbol{\rho}_1 = (\frac{1}{6}, \frac{1}{3}, \frac{1}{2})^T, \boldsymbol{\rho}_2 = (\frac{1}{2}, \frac{1}{3}, \frac{1}{6})^T, \mathbf{a} = (\frac{1}{3}, \frac{1}{6}, \frac{1}{2})^T$. Then using `vegdist(method = ''bray'')` in the R package `vegan 2.4-3` (Oksanen et al., 2017), the Bray-Curtis distance between $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ is 0.333 to three decimal places, but the Bray-Curtis distance between $\mathbf{a} \oplus \boldsymbol{\rho}_1$ and $\mathbf{a} \oplus \boldsymbol{\rho}_2$ is 0.420 to three decimal places. Other popular measures of community dissimilarity are shown not to be perturbation invariant (in the context of temporal change) in Spencer (2015, Appendix B). In contrast, the Aitchison distance (Aitchison, 1992) is a well-established perturbation-invariant measure of dissimilarity between composition. Thus, analyses of dissimilarity between relative abundances should be based on Aitchison distance, rather than the currently-popular measures of community dissimilarity.

Model-based analysis is an increasingly popular alternative way of analyzing differences between communities (Warton et al., 2015). Model-based methods allow appropriate modelling of the observation process, which often leads to mean-variance relationships different from those implicit in widely-used measures of dissimilarity (Warton et al., 2012). Model-based methods are generally more flexible, interpretable and efficient than dissimilarity-based methods (Warton et al., 2015). For example, once a parametric model has been fitted to a set of communities along an environmental gradient, the function that describes expected values can be differentiated to find the rate of change of the community along the gradient, and integration can be used to find the mean community over the entire gradient. Even when dissimilarities are directly of interest, a parametric model is useful in understanding how expected dissimilarity depends on distance along the gradient. However, an overlooked distinction between model-based and dissimilarity-based methods is that most model-based methods (e.g. Wang et al., 2012) are designed for abundance data, while most dissimilarities are designed for relative abundance data. Relative abundances have a different ecological meaning from abundances: communities are often treated as equivalent if they have the same "shape" (relative abundances) regardless of differences in "size" (total abundance). Also, in some cases (e.g. point counts from vegetation and on coral reefs, pollen counts, and environmental sequencing data), only relative abundances are available. Thus, there is a need for model-based analyses of relative abundance data. It seems likely that compositional data analysis, combined with the calculus of simplex-valued functions (Egozcue et al., 20[illegible] will meet this need.

Here, we show how the vector space structure of the simplex provides a coherent way to study changes in community composition along environmental gradients. We show that a low-order polynomial provides a good model for the composition of a community of sessile hard-substrate marine organisms over a depth gradient. We illustrate the use of Aitchison distance as a principled measure of dissimilarity between communities, and use the algebraic structure of the simplex to understand how dissimilarity depends on depth. In particular, we determine the conditions for the same community composition to occur at different depths. We use the derivative of community composition with respect to depth to determine the depth at which the community is changing fastest. We use the integral of community composition over a vertical slice to determine which organisms dominate the mean composition over the entire depth range.

2 Materials and methods

2.1 Location

We studied the community of sessile hard-substrate marine organisms on the walls of Salthouse Dock (53.4006° N, 2.9898° W), Port of Liverpool, United Kingdom. Salthouse Dock is part of the southern dock system on the River Mersey, connected to Wapping Dock to the South, Albert Dock to the West and Canning Dock to the North via Albert Dock. The docks fell into disuse in the 1970s, but were dredged and reopened for recreational use in 1981 (Fielding, 1997, pp. 10-14). Since then, they have been redeveloped as part of a commercial project, and with the completion of the Liverpool Canal Link, are now also connected to the Leeds-Liverpool Canal (Coutts et al., 2012). The regenerated docks are a shallow, semi-enclosed brackish water habitat, with salinity between 22‰ and 33‰ in the South Docks (Fielding, 1997, pp. 17, 70).

PeerJ reviewing PDF | (2018:04:27313:0:1:NEW 26 Apr 2018)

2.2 Video transects

An OpenROV v2.8 remotely-operated vehicle (OpenROV, Berkeley, CA) with an IMU/Depth sensor and the Pro Camera-HD Upgrade (1080p) was used to take 31 approximately vertical transects from surface to bottom, haphazardly spaced along the northern and eastern walls of Salthouse Dock, on 2 February 2017. The distance from the wall was typically around 0.3 m to 0.4 m, giving a field of view with an area of approximately 0.29 m^2 to 0.51 m^2. The field of view was not known exactly because the lasers on the ROV, intended to indicate a known distance on the images, malfunctioned. However, the field of view was always large enough to contain many organisms, so that the relative abundances are unlikely to depend on the exact area sampled. A GoPro HERO3+ Black Edition (GoPro, San Mateo, CA) was also attached to the ROV to provide an extra source of footage with higher resolution but more distortion. The OpenROV videos and telemetry data were recorded in the inbuilt Cockpit software (v30.1.0 with software patch release). The video and data files were downloaded and python scripts were written to overlay depth data on the corresponding videos.

2.3 Image analysis

Four still images were captured per transect at varying depths from 0.11 m to 3.72 m (except one transect where five stills were taken), making 125 still images in total. These stills were selected by the clarity of the image, which is generally when the ROV camera is at an optimum distance away, by advancing the videos one frame at a time. On each image, the taxon present at each of 100 randomly-selected points was recorded using the JMicroVision v1.2.7 image analysis software (Roduit, 2008, Figure 1). Organisms (Table 1) were identified from still images, supplementary GoPro footage, and where possible, specimens collected near the surface, using Hayward and Ryland (1995). For the non-native colonial sea squirt *Botrylloides violaceus*, we used the Marine Life Information Network (Snowden, 2008). Where an organism was growing on top of another, the organism taking up space on the wall was recorded. If positive identification was not possible, the point was skipped and another point drawn. “Bare wall” was recorded if no macroscopic organism was present, or (as often occurred near the bottom) the wall was covered by grey detritus, so that any macroscopic organisms which may have been present were not visible. Point counts were exported from JMicroVision into ASCII text files, which were combined using an R 3.4.0 script (R Core Team, 2017) into a single file with depth data.

2.4 Data analysis

2.4.1 Data aggregation

Due to the rarity of barnacles and *Stomphia coccinea* (one individual of each), these two taxa were excluded from the analysis. The remaining taxa were combined into eight categories, consisting of organisms that were ecologically similar and/or could not be reliably distinguished: algae (red and green), *Aurelia aurita* polyps, *Bugula spp.*, colonial ascidians (*Botryllus schlosseri*, *Botrylloides leachii* and *Botrylloides violaceus*), *Diadumene cincta*, solitary ascidians (*Ciona intestinalis* and *Styela clava*), sponges (*Halichondria spp.* and others), *Mytilus edulis*. We also included the "bare wall" category (for the absence of macroscopic organisms, although usually there was a biofilm of microscopic algae and bacteria, or a layer of detritus).

2.4.2 Statistical model

Let the counts in the ith observation (still image) be $\mathbf{y}_i = (y_{i,1}, y_{i,2} \ldots, y_{i,9})^T$, where $y_{i,j}$ is the observed count of the jth taxon in the ith observation. We assume that $\mathbf{y}_i$ follows a multinomial$(n_i, \boldsymbol{\rho}_i)$ distribution, where n_i is the number of points counted for the ith observation (always 100 in our data) and $\boldsymbol{\rho}_i$ is a vector of expected relative abundances of each taxon.

The vector $\boldsymbol{\rho}_i$ consists of non-negative elements with a fixed sum of 1, and is therefore a composition. The sum constraint, and associated constraints on the covariance structure of compositions, make it difficult to specify sufficiently flexible parametric models for untransformed compositions (Aitchison, 1986, chapter 3). The most popular modern approach to analysis of compositional data is to transform an s-part composition into an unconstrained real space with $s-1$ dimensions. We used an isometric logratio transformation (Egozcue et al., 2003), which is an isomorphism (so that perturbation and powering in the simplex correspond to ordinary vector addition and scalar multiplication in the real space) and an isometry (so that distances under an appropriate norm in the simplex correspond to Euclidean distances in the real space). We used the isometric logratio transformation with the default basis matrix in the R package `compositions`, version 1.40-1 (van den Boogaart and Tolosana-Delgado, 2008), although our results do not depend on this choice of basis.

Let the transformed expected relative abundances for the ith observation be $\mathbf{x}_i = \mathrm{ilr}(\boldsymbol{\rho}_i)$, where $\mathbf{x}$ is an 8-dimensional real vector, and ilr() represents an isometric logratio transformation. We assume that the transformed expected relative abundances can be described by the multivariate regression model

$$\mathbf{x}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_i + \boldsymbol{\beta}_2 z_i^2 + \boldsymbol{\varepsilon}_i, \tag{2}$$

where $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the intercept and linear and quadratic depth coefficients respectively, z_i is the centred and scaled depth for the ith observation, and the errors $\boldsymbol{\varepsilon}$ have an 8-dimensional multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. We fitted this model using Bayesian estimation (Supplemental Information).

Because the isometric logratio transformation is an isomorphism between the simplex with Aitchison geometry and the ordinary real space, we can back-transform the deterministic part of Equation 2 to obtain an expression in terms of perturbation and powering in the simplex:

$$\begin{aligned} M(\boldsymbol{\rho}_i) &= \mathrm{ilr}^{-1}\left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_i + \boldsymbol{\beta}_2 z_i^2\right) \\ &= \boldsymbol{\gamma}_0 \oplus (z_i \odot \boldsymbol{\gamma}_1) \oplus (z_i^2 \odot \boldsymbol{\gamma}_2), \end{aligned}$$

where $\boldsymbol{\gamma}_j = \mathrm{ilr}^{-1}(\boldsymbol{\beta}_j)$, $j = 0,1,2$. The composition $M(\boldsymbol{\rho}_i)$ is the metric centre of the distribution of $\boldsymbol{\rho}_i$, an appropriate measure of location for compositions (Aitchison, 1989).

To make the behaviour of the predictions for rare taxa more obvious, we also examined the predictions on a centred logratio (clr) scale, in which the value on the y-axis is the log of the ratio of the corresponding component to the geometric mean of all components (Aitchison, 1986, p. 79). Thus a constant slope on the clr scale corresponds to constant proportional change in the relative abundance of a given taxon.

2.4.3 Community dissimilarity

As described above, most of the common measures of dissimilarity between communities are not perturbation invariant. In the Aitchison geometry, the obvious perturbation invariant measure of difference between two s-part compositions is the Aitchison norm of the compositional difference, defined by

$$\begin{aligned} d_a(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) &= \|\boldsymbol{\rho}_1 \ominus \boldsymbol{\rho}_2\|_a \\ &= \left[\sum_{i=1}^{s} \log \frac{\rho_{1,i}}{g(\boldsymbol{\rho}_1)} - \log \frac{\rho_{2,i}}{g(\boldsymbol{\rho}_2)}\right]^{1/2}, \end{aligned}$$

where $g(\boldsymbol{\rho})$ denotes the geometric mean of the parts of a composition (Aitchison, 1992; Egozcue et al., 2003). It is immediately obvious that this is perturbation invariant, because $(\mathbf{a} \oplus \boldsymbol{\rho}_1) \ominus (\mathbf{a} \oplus \boldsymbol{\rho}_2) = \boldsymbol{\rho}_1 \ominus \boldsymbol{\rho}_2$, by the associative, commutative and identity properties of the vector space. Under this approach, the dissimilarity between the expected compositions $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2$ is given by

$$\begin{aligned} \|\boldsymbol{\rho}_1 \ominus \boldsymbol{\rho}_2\|_a &= \left\| \left[\gamma_0 \oplus (z_1 \odot \gamma_1) \oplus \left(z_1^2 \odot \gamma_2\right)\right] \ominus \left[\gamma_0 \oplus (z_2 \odot \gamma_1) \oplus \left(z_2^2 \odot \gamma_2\right)\right] \right\|_a \\ &= |z_1 - z_2| \|\gamma_1 \oplus [(z_1 + z_2) \odot \gamma_2]\|_a, \end{aligned} \quad (3)$$

using the identity, commutative, associative and distributive properties of the vector space to simplify.

The Aitchison norm has a biological meaning in terms of population growth. In temporal comparisons, the Aitchison norm of the compositional difference between two sets of relative abundances is proportional to the among-taxon standard deviation of proportional population growth rates (Spencer, 2015). In spatial comparisons, we can therefore think of the Aitchison norm as measuring the among-taxon variability in proportional population growth rates that is needed to transform one set of relative abundances into another, over a given time interval. This property is important because in a closed system, population growth is the only way to transform one set of relative abundances into another. No other measure of community dissimilarity has this interpretation.

The simplex with Aitchison geometry is a normed vector space (Egozcue et al., 2003) and therefore a metric space (Sutherland, 2009, pp. 39-40). Thus $\|\boldsymbol{\rho}_1 \ominus \boldsymbol{\rho}_2\|_a = 0$ if and only if $\boldsymbol{\rho}_1 \ominus \boldsymbol{\rho}_2 = \mathbf{0}$, where $\mathbf{0}$ is the identity element in the simplex. From Equation 3, assuming that $\boldsymbol{\gamma}_1 \neq \mathbf{0}$ and $\boldsymbol{\gamma}_2 \neq \mathbf{0}$, this happens when either $z_1 = z_2$ (the two compositions are at the same depth) or $\boldsymbol{\gamma}_2 = \left(-\frac{1}{z_1+z_2}\right) \odot \boldsymbol{\gamma}_1$ (the coefficient of squared depth is a powering of the coefficient of depth). Thus, if we plot dissimilarity on a grid of depths, there will always be zeros on the main diagonal, because communities at the same depth have the same expected composition. There may also be communities at different depths with the same expected composition, along a counter-diagonal where centred and scaled depth has a constant sum, but only in the special case where $\boldsymbol{\gamma}_2$ is a powering of $\boldsymbol{\gamma}_1$ (or equivalently, where $\boldsymbol{\beta}_2$ is a scalar multiple of $\boldsymbol{\beta}_1$ in ilr coordinates).

We calculated posterior distributions of dissimilarities among 100 equally-spaced expected compositions between the minimum and maximum depths, both including and excluding bare wall. We plotted the posterior mean dissimilarity matrix, and the widths of the 95% highest posterior density intervals. We only report the results including bare wall here, because those excluding bare wall were very similar. Note that it is valid to exclude some parts of the composition if necessary, because the subcompositional coherence property means that such exclusion will not affect relationships among the remaining parts (Aitchison, 1994).

2.4.4 Rate of change of community composition with depth

The community is changing rapidly with respect to depth if a small increase in depth leads to a large difference in composition. In order to correctly evaluate this change, we need an appropriate definition of difference in composition. Given the geometry of the simplex, the difference in composition between depths z and $z+h$ is naturally expressed as $\mathbf{f}(z+h) \ominus \mathbf{f}(z)$. Then letting h go to zero leads to the obvious definition of the derivative $D^{\oplus}\mathbf{f}$ of a simplex-valued function $\mathbf{f}$,

$$D^{\oplus}\mathbf{f}(z) = \lim_{h \to 0} \left(\frac{1}{h} \odot (\mathbf{f}(z+h) \ominus \mathbf{f}(z)) \right),$$

provided this limit exists (Egozcue et al., 2011, section 12.2.2). Using the rules for differentiation of simplex-valued functions (Egozcue et al., 2011, section 12.2.2), in our model, the derivative of community composition with respect to depth, at a depth of z, is

$$D^{\oplus}\mathbf{f}(z) = \gamma_1 \oplus (2z \odot \gamma_2).$$

This is itself a composition. If we want a scalar measure of rate of change, the obvious choice is the norm of this derivative. It is intuitively obvious that the usual Euclidean norm is not appropriate, because the zero element for compositions (with all parts equal, corresponding to no change in composition with respect to depth) does not have zero Euclidean norm. Instead, we use the Aitchison norm $\|D^{\oplus}\mathbf{f}(z)\|_a$ (Egozcue et al., 2003), which is zero in the situation where there is no change in composition with respect to depth, and is used in the definition of a limit in the simplex (Egozcue et al., 2011, Definition 12.2.1). The easiest way to think of this norm is that it is equal to the Euclidean norm of the derivative in isometric logratio coordinates. It is also important to remember that we are measuring proportional change: doubling of relative abundance means the same thing whether the initial relative abundance is low or high. This is an essential property, because relative abundances have meaning only in relative terms.

We evaluated the posterior distribution of this scalar measure of rate of change at 100 equally-spaced depths over the observed depth range.

2.4.5 Depth-integrated relative abundances

Over a vertical slice from surface to bottom, a taxon that has high relative abundance over a small range of depths may be unimportant compared to a taxon that has moderate relative abundance at all depths. We therefore want some measure of the “mean” relative abundances over a vertical slice.

The arithmetic mean is not appropriate for compositional data. For example, with a banana-shaped distribution, the arithmetic mean may lie completely outside the cloud of observations. The metric centre is a more appropriate measure of the centre of a compositional distribution which avoids these problems (Aitchison, 1989). However, taking a sample estimate of the metric centre over all depths is problematic, because sample relative abundances of zero often occur. Zeros are difficult to deal with in compositional data analysis (Martín-Fernandez et al., 2011), and in this context, will lead to the estimate of the centre being undefined. In addition, if the depth distribution of samples is not uniform, the sample estimate of the centre will be biased. Thus, integrating the model-estimated composition over the full range of depths may be a better way to summarize the structure of the community.

The mean of a real function f of one variable over the interval $[a,b]$ is

$$\frac{1}{b-a}\int_a^b f(x)\,\mathrm{d}x,$$

which can be thought of as the value of the constant function whose integral over $[a,b]$ is the same as that of f over the same interval (Riley et al., 2002, pp. 73-74). If we treat community composition as a simplex-valued function of depth, then the analogous mean of this function over the full range of depths gives the composition representing the relative abundance of each part over a vertical slice from top to bottom of the dock wall. Let $[S,D]$ be the depth range, from shallow to deep. Using the rules for integration of simplex-valued functions (Egozcue et al., 2011, section 12.3.2), the required mean value is

$$\frac{1}{D-S}\odot\left[(z\odot\gamma_0)\oplus\left(\frac{z^2}{2}\odot\gamma_1\right)\oplus\left(\frac{z^3}{3}\odot\gamma_2\right)\right]_S^D.$$

We evaluated the posterior distribution of this mean value.

PeerJ reviewing PDF | (2018:04:27313:0:1:NEW 26 Apr 2018)

3 Results

3.1 Trends in composition with depth

Images at different depths often showed large differences in relative abundances (Figure 1). For example, Figure 1a, at 0.19 m, was dominated by green algae. Figure 1b, at 1.33 m, was dominated by bare wall, *Halichondria* spp. and *Ciona intestinalis*, and also had some *Diadumene cincta* and *Bugula* spp. Figure 1c, at 3.02 m, still had fairly high relative abundance of *Halichondria* spp. and *Ciona intestinalis*, and also a moderate relative abundance of *Mytilus edulis*. However, large areas of the lower part of this image were covered by grey detritus and were therefore assigned to bare wall.

Over all the images, there were obvious changes in the relative abundance of bare wall, *Bugula*, solitary ascidians, algae and sponges with depth (Figure 2a-e, circles), while the relative abundances for the rare taxa *Diadumene cincta*, *Mytilus edulis*, *Aurelia aurita* and colonial ascidians had apparently weaker trends (Figure 2f-i, circles). The fitted model (Figure 2, lines) closely tracked the pattern in the observations, indicating that a quadratic model is a plausible description of changes in relative abundance over the depth gradient. The relative abundance of bare wall increased from about 0.1 to 0.4 between 0 m and 1 m, remained fairly constant until 2 m, and increased again to about 0.9 in the deepest samples (Figure 2a). This is a more complicated pattern than could be produced by a quadratic function in an unrestricted space. The cover of algae dropped dramatically from around 0.8 at the surface to almost nothing just after 1 m (Figure 2c). The remaining three taxa with moderately high relative abundances at some depths (*Bugula*, solitary ascidians and sponges: Figure 2b, c, e) showed similar patterns, being absent at the surface and rare in the deepest samples, with peaks at intermediate depths (around 1 m for sponges, 2 m for *Bugula* and solitary ascidians).

For the rare taxa, centred logratio plots showed that although the predicted relative abundances were everywhere low, there were large proportional changes in predicted relative abundance (Figure 2f to i, insets). All the rare taxa had lower predicted relative abundances near the surface, with

Diadumene cincta (Figure 2f) showing little change at mid depths, *Mytilus edulis* (Figure 2g) and colonial ascidians (Figure 2i) decreasing in abundance in the deepest samples, and *Aurelia aurita* (Figure 2h) increasing steadily with depth. The centred logratio trends are in accordance with the observations. For example, *A. aurita* was only observed occasionally. However, when it was observed, it was below 3 m and in dense aggregations of small polyps, especially on downward-facing parts of the dock wall. The fitted trend ensures that the probability of a non-zero count is very low except for images deeper than 3 m.

3.2 Community dissimilarity

Dissimilarity between expected composition, measured as the Aitchison norm of the compositional difference (Equation 3) was small for small differences in depth (Figure 3a, dark colours), and increased with increasing difference in depth. The uncertainty in dissimilarity behaved in a similar way (Figure 3b). There was no counter-diagonal pattern of similar communities at widely-separated depths. This implies that the squared depth coefficient γ_2 is not a powering of the depth coefficient γ_1. Figure 4 confirms this. For the subcomposition consisting of bare wall, algae and sponges, the set of powerings of γ_1 can be represented as a compositional straight line in the simplex (Figure 4, lines). The point in the simplex representing γ_2 does not lie on this line. Thus γ_2 is not a powering of γ_1, and dissimilarity cannot be zero for communities with a non-zero difference in depth. Although expected relative abundance may be the same at widely-separated depths for individual taxa (e.g. sponges, Figure 2e), this pattern does not coincide across taxa.

3.3 Rate of change of community composition with depth

The posterior mean rate of change of community composition with respect to depth was highest at the surface, decreased with increasing depth until just below 2 m, and increased again until the bottom was reached (Figure 5, white line). Although the 95% credible band for the rate of change (Figure 5, grey band) was wide, the majority of the rates of change for individual Monte Carlo iterations (Figure 5, black lines) had the same shape, with a minimum in the middle (between depths 1 m and 3 m). The overall pattern of rate of change makes intuitive sense, given that on the centred logratio scale, all taxa had substantial changes in posterior mean predicted relative abundance near the surface, all but algae (Figure 2d, inset) and *Aurelia aurita* (Figure 2h, inset) had flatter relationships at mid depths, and all but *Diadumene cincta* (Figure 2f, inset) had substantial changes near the bottom.

3.4 Mean composition of organisms over the entire depth

Over the entire depth range, bare wall had the highest relative abundance of around 0.5 (Figure 6). This means that over half the area of the dock walls was not covered by any macroscopic organism. The macroscopic taxa with the highest relative abundances were sponges and solitary ascidians, with relative abundance around 0.2, followed by *Bugula*, with relative abundance around 0.05. These taxa, especially *Bugula*, did not have very high relative abundance at any depth (Figure 2b-c, e), but had moderately high relative abundance at all depths, resulting in fairly high mean relative abundances. All other taxa had low mean relative abundances, including algae, which was very abundant at the surface but decreased quickly with depth (Figure 2d).

4 Discussion

We showed that the vector space structure of the simplex leads naturally to tangible, functional and intuitive summaries of the changes in community compositions with depth in a subtidal marine system. A quadratic model was a plausible description of these changes. This is important because needing a complicated model to describe real data is often a sign of some fundamental misspecification. Although a regression analysis cannot reveal the causes of the pattern we observed, it can hint at possible explanations. For example, integrating the composition over depth showed that bare wall had much higher relative abundance than any taxon, suggesting that the classical picture of intense competition for space determining the structure of subtidal marine communities may need revision (Ferguson et al., 2013; Svensson and Marshall, 2015). A major strength of the compositional data approach is the logical connection between statistical modelling and ecology. For example, we showed that the community was changing fastest at the surface and near the bottom, and that we would not find the same community composition at different depths. These results were based on a measure of dissimilarity that has both a strong statistical justification, based on the requirement for perturbation invariance (Aitchison, 1992) and a natural biological interpretation as the amount of among-taxon variability in proportional population growth rates needed to transform one community into another. We therefore believe that compositional data analysis deserves to be more widely used by ecologists.

An observational study alone cannot determine the causes of the patterns in relative abundance with depth in our data. However, although space is thought to be a limiting resource in many hard-substrate subtidal communities (Witman and Dayton, 2001, p. 356), it seems unlikely that space is limiting at our study site, because of the high relative abundance of bare wall (Figure 6). Our surveys were done in winter, but relative abundance of bare wall remained high in summer (Edney, 2017), so it is unlikely that space is even seasonally limiting. Also, competition for space alone cannot explain the change in community composition with depth. Three other factors that may contribute to the depth effect are recruitment, food and oxygen availability.

Recruitment may regulate population dynamics of sessile marine organisms (Caley et al., 1996). For example, in a simple model for the dynamics of open populations of the bryozoan *Cellepora pumicosa*, equilibrium population size was proportional to recruitment rate (Hughes, 1990). At our site, settlement panels at 3 m typically had fewer than half as many new organisms as those at 1 m after five weeks in summer (Edney, 2017). Thus, changes in recruitment with depth are likely to contribute to the depth effect on community composition.

Competition for food may also be important. Increasing phytoplankton supply increased species richness and reduced free space on settlement panels (Svensson and Marshall, 2015). Field measurements showed reduced phytoplankton density close to the walls of a dock adjacent to our site (Fielding, 1997, p. 118). Thus, phytoplankton abundance may be limiting. However, it is not clear whether light levels will decrease with depth rapidly enough to generate a strong depth effect on phytoplankton production, and thus for phytoplankton limitation to generate a depth effect on community composition. For example, chlorophyll *a* concentrations in the Liverpool docks were little different between surface and bottom water (Fielding, 1997, p. 106).

Oxygen depletion may occur in the low-flow, topographically complex environment typical of fouling communities (Ferguson et al., 2013). Summer oxygen levels in the Liverpool docks may be much lower near the bottom than the surface (Fielding, 1997, pp. 74-75). Thus exploitative competition for oxygen may become more intense as depth increases, potentially contributing to the depth effect on community composition, at least in summer.

The compositional regression approach taken here is closely related to multinomial logistic regression, but offers some advantages in flexibility and interpretability. Multinomial logistic regression is another approach to the analysis of count data derived from an underlying continuous model for relative abundances on a gradient (e.g. Qian et al., 2012). In multinomial logistic regression, the linear predictor is expressed in terms of logs of ratios of relative abundances, exactly as in a compositional linear model. In its basic form, multinomial logistic regression does not allow for overdispersion, which in a compositional linear model such as Equation 2 is captured by the random errors ε_i (Xia et al., 2013). Overdispersion is important for describing patterns in organisms that tend to occur in aggregations, such as the cnidarian *A. aurita* in our data.

More importantly, treating the simplex as a vector space with perturbation and powering operations makes it easy to do algebra and analysis on compositions. This can simplify interpretation compared to the multinomial regression approach, where coefficients are expressed on the log-odds scale (Billheimer et al., 2001). For example, we were able to determine why, in algebraic terms, we did not see communities with high similarity at widely separated depths, even though such an outcome is possible under a quadratic model. Such outcomes are related to the "double-zero problem" in the design of measures of ecological dissimilarity (Legendre and Legendre, 2012, p. 271). A given taxon may have low expected relative abundance at both ends of a gradient because of unsuitable conditions. In our data, this pattern occurred for taxa including solitary ascidians and sponges (Figure 2c and e). With finite sampling effort, this may lead to zeros at both ends of the gradient. However, unless the quadratic coefficient is an exact powering of the linear coefficient, the predicted dissimilarity will not be exactly zero. We therefore do not think that similarity resulting from similar relative abundance patterns is ecologically misleading, even if it does not arise from similar environments.

The algebra of perturbation and powering is central to visualization and interpretation of experiments and observational studies on compositional response variables. For example, Billheimer et al. (2001) expressed the effects of vegetation removal and addition of specialist predators on arthropod community composition, relative to a control treatment, using a perturbation. Similarly, Billheimer et al. (1997) used a perturbation to visualize the effect of salinity on relative abundances of stress-tolerant taxa, intolerant taxa and palp worms in a benthic habitat. In a regression study, Xia et al. (2013) visualized the estimated effects of changes in nine different nutrients on the relative abundances of three bacterial genera in the human gut microbiome as compositional straight lines, using the perturbation and powering operators. In all these cases, the necessary algebra is very straightforward if the simplex is treated as a vector space. Less obviously, knowing that a statistic has the perturbation invariance property (Aitchison, 1992) guarantees that differences in detection probabilities among taxa will not affect the results. For example, because we used the perturbation-invariant Aitchison distance as a measure of dissimilarity, our estimates of rate of change will not be biased by large, conspicuous organisms such as the solitary ascidians *Ciona intestinalis* and *Styela clava* being easier to detect than small, inconspicuous organisms such as the cnidarian *A. aurita*. In contrast, widely-used dissimilarity measures such as the Bray-Curtis index, which is not perturbation invariant, would lead to artefacts.

5 Conclusions

In conclusion, we believe that ecologists working with relative abundance data would benefit from making more use of compositional data analysis. There has been substantial progress in compositional data analysis since the 1980s, but as yet, it has had little influence on ecology. In areas such as the analysis of environmental gradients, compositional data analysis provides a simple, coherent approach that is in keeping with the current preference for model-based analyses. With only a small shift in perspective, techniques such as differentiation and integration can be used to answer ecological questions in ways that have meaning for relative abundances.

Acknowledgements

We are grateful to the 2017 ENVS271 class for ROV piloting.

References

Aitchison, J. (1986). *The statistical analysis of compositional data*. Chapman and Hall, London.

Aitchison, J. (1989). Measures of location for compositional data sets. *Mathematical Geology*, 21:787–790.

Aitchison, J. (1992). On criteria for measures of compositional difference. *Mathematical Geology*, 24(4):365–379.

Aitchison, J. (1994). Principles of compositional data analysis. In Anderson, T. W., Olkin, I., and Fang, K., editors, *Multivariate analysis and its applications*, volume 24, pages 73–81. Institute of Mathematical Statistics, Hayward, CA.

Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H.-W., and Silkey, M. (1997). Natural variability of benthic species composition in Delaware Bay. *Environmental and Ecological Statistics*, 4:95–115.

Billheimer, D., Guttorp, P., and Fagan, W. F. (2001). Statistical interpretation of species composition. *Journal of the American Statistical Association*, 96:1205–1214.

Caley, M. J., Carr, M. H., Hixon, M. A., Hughes, T. P., Jones, G. P., and Menge, B. A. (1996). Recruitment and the local dynamics of open marine populations. *Annual Review of Ecology and Systematics*, 27:477–500.

Coutts, R., Pellizzon, R., and Alderdice, S. (2012). A Waterspace Strategy for the Sustainable Development of Liverpool South Docks. `http://www.liverpoolvision.co.uk/wp-content/uploads/2014/01/Liverpool-South-DocksWaterspace-2012.pdf`. Accessed 19 June 2017.

Crowe, M. J. (1994). *A history of vector analysis: the evolution of the idea of a vectorial system*. Dover Publications, Inc., New York.

Dorier, J.-L. (1995). A general outline of the genesis of vector space theory. *Historia Mathematica*, 22:227–261.

Edney, S. (2017). The influences of larval dispersal and competition on the colonisation of sessile communities at different depths in Salthouse docks. Master's thesis, University of Liverpool.

Egozcue, J. J., Jarauta-Bragulat, E., and Díaz-Barrero, J. L. (2011). Calculus of simplex-valued functions. In Pawlowsky-Glahn, V. and Buccianti, A., editors, *Compositional data analysis: theory and applications*, pages 158–175. John Wiley & Sons, Ltd, Chichester.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3):279–300.

Ferguson, N., White, C. R., and Marshall, D. J. (2013). Competition in benthic marine invertebrates: the unrecognized role of exploitative competition for oxygen. *Ecology*, 94:126–135.

Fielding, N. J. (1997). *Fish and benthos communities in regenerated dock systems on Merseyside*. PhD thesis, University of Liverpool.

Fraleigh, J. B. and Beauregard, R. A. (1995). *Linear algebra*. Addison-Wesley, Reading, Massachusetts, third edition.

Gross, K. and Edmunds, P. J. (2015). Stability of Caribbean coral communities quantified by long-term monitoring and autoregression models. *Ecology*, 96:1812–1822.

Hayward, P. J. and Ryland, J. S., editors (1995). *Handbook of the marine fauna of North-West Europe*. Oxford University Press, Oxford.

Hughes, T. P. (1990). Recruitment limitation, mortality, and population regulation in open systems: a case study. *Ecology*, 71:12–20.

Jackson, D. A. (1997). Compositional data in community ecology: the paradigm or peril of proportions. *Ecology*, 78:929–940.

Legendre, P. and Legendre, L. (2012). *Numerical Ecology*. Elsevier, Amsterdam, 3rd English edition.

Martin, P. S. and Mosimann, J. E. (1965). Geochronology of pluvial Lake Cochise, southern Arizona, III. Pollen statistics and Pleistocene metastability. *American Journal of Science*, 263:313–358.

Martín-Fernandez, J. A., Palarea-Albaladejo, J., and Olea, R. A. (2011). Dealing with zeros. In Pawlowsky-Glahn, V. and Buccianti, A., editors, *Compositional data analysis: theory and applications*, pages 43–58. John Wiley & Sons, Ltd, Chichester.

Mosimann, J. E. (1962). On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika*, 49:65–82.

Neal, K. J. (2007). *Diadumene cincta* A sea anemone. `http://www.marlin.ac.uk/species/detail/2052`. In Tyler-Walters H. and Hiscock K. (eds) Marine Life Information Network: Biology and Sensitivity Key Information Reviews, [on-line].

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H. (2017). *vegan: Community Ecology Package*. R package version 2.4-3.

Pawlowsky-Glahn, V. and Buccianti, A., editors (2011). *Compositional data analysis: theory and applications*. Wiley, Chichester.

Qian, S. S., Cuffney, T. F., and McMahon, G. (2012). Multinomial regression for analyzing macroinvertebrate assemblage composition data. *Freshwater Science*, 31:681–694.

R Core Team (2017). R: A language and environment for statistical computing. `https://www.R-project.org/`.

Riley, K. F., Hobson, M. P., and Bence, S. J. (2002). *Mathematical methods for physics and engineering*. Cambridge University Press, Cambridge, second edition.

Roduit, N. (2008). JMicrovision: Image analysis toolbox for measuring and quantifying components of high-definition images. `http://www.jmicrovision.com/`. Accessed 19 June 2017.

Snowden, E. (2008). A colonial sea squirt *Botrylloides violaceus*. `http://www.marlin.ac.uk/species/detail/2186`. In Tyler-Walters H. and Hiscock K. (eds) Marine Life Information Network: Biology and Sensitivity Key Information Reviews, [on-line].

Spencer, M. (2015). Size change, shape change, and the growth space of a community. *Journal of Theoretical Biology*, 369:23–41.

Sutherland, W. A. (2009). *Introduction to metric and topological spaces*. Oxford University Press, Oxford, second edition.

Svensson, J. R. and Marshall, D. J. (2015). Limiting resources in sessile systems: food enhances diversity and growth of suspension feeders despite available space. *Ecology*, 96:819–827.

van den Boogaart, K. G. and Tolosana-Delgado, R. (2008). "`compositions`": a unified R package to analyze compositional data. *Computers and Geosciences*, 34:320–338.

Wang, Y., Naumann, U., Wright, S. T., and Warton, D. I. (2012). `mvabund` - an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution*, 3:471–474.

Warton, D. I., Foster, S. D., De'ath, G., Stoklosa, J., and Dunstan, P. K. (2015). Model-based thinking for community ecology. *Plant Ecology*, 216:669–682.

Warton, D. I., Wright, S. T., and Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution*, 3:89–101.

Witman, J. D. and Dayton, P. K. (2001). Rocky subtidal communities. In Bertness, M. D., Gaines, S. D., and Hay, M. E., editors, *Marine community ecology*, pages 339–366. Sinauer Associates, Inc., Sunderland, Massachusetts.

Xia, F., Chen, J., Fung, W. K., and Li, H. (2013). A logistic normal multinomial regression model for microbiome compositional data analysis. *Biometrics*, 69:1053–1063.

Yuan, Y., Buckland, S. T., Harrison, P. J., Foss, S., and Johnston, A. (2016). Using species proportions to quantify turnover in biodiversity. *Journal of Agricultural, Biological, and Environmental Statistics*, 21:363–381.

Table 1: List of species identified from stills and samples.

Aurelia aurita
Botryllus schlosseri
Botrylloides leachii
Botrylloides violaceus
Bugula spp.
Ciona intestinalis
Diadumene cincta (some individuals may be *Metridium senile* (Neal, 2007))
Green algae
Halichondria spp.
Mytilus edulis
Other sponges
Red algae
Stomphia coccinea
Styela clava
Unidentified barnacle

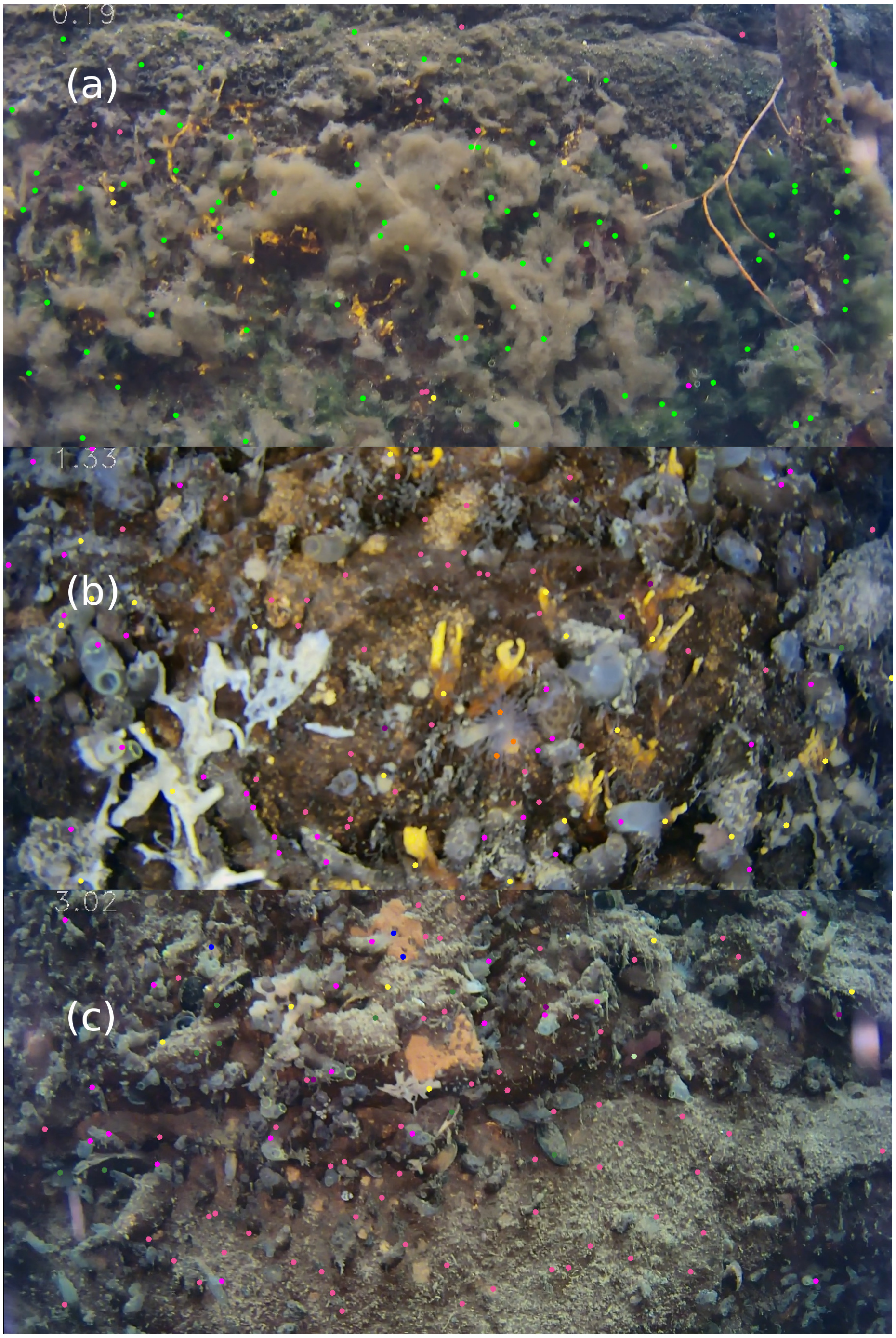

Figure 1: Still images from (a) 0.19 m, (b) 1.33 m and (c) 3.02 m, with 100 point counts each. Bright green dots correspond to green algae, pink dots to bare wall, violet to *Ciona intestinalis*, yellow to *Halichondria spp.*, purple to *Bugula spp.*, orange to *Diadumene cincta*, green to *Mytilus edulis*, blue to other sponges and off-white to *Botrylloides violaceus*.

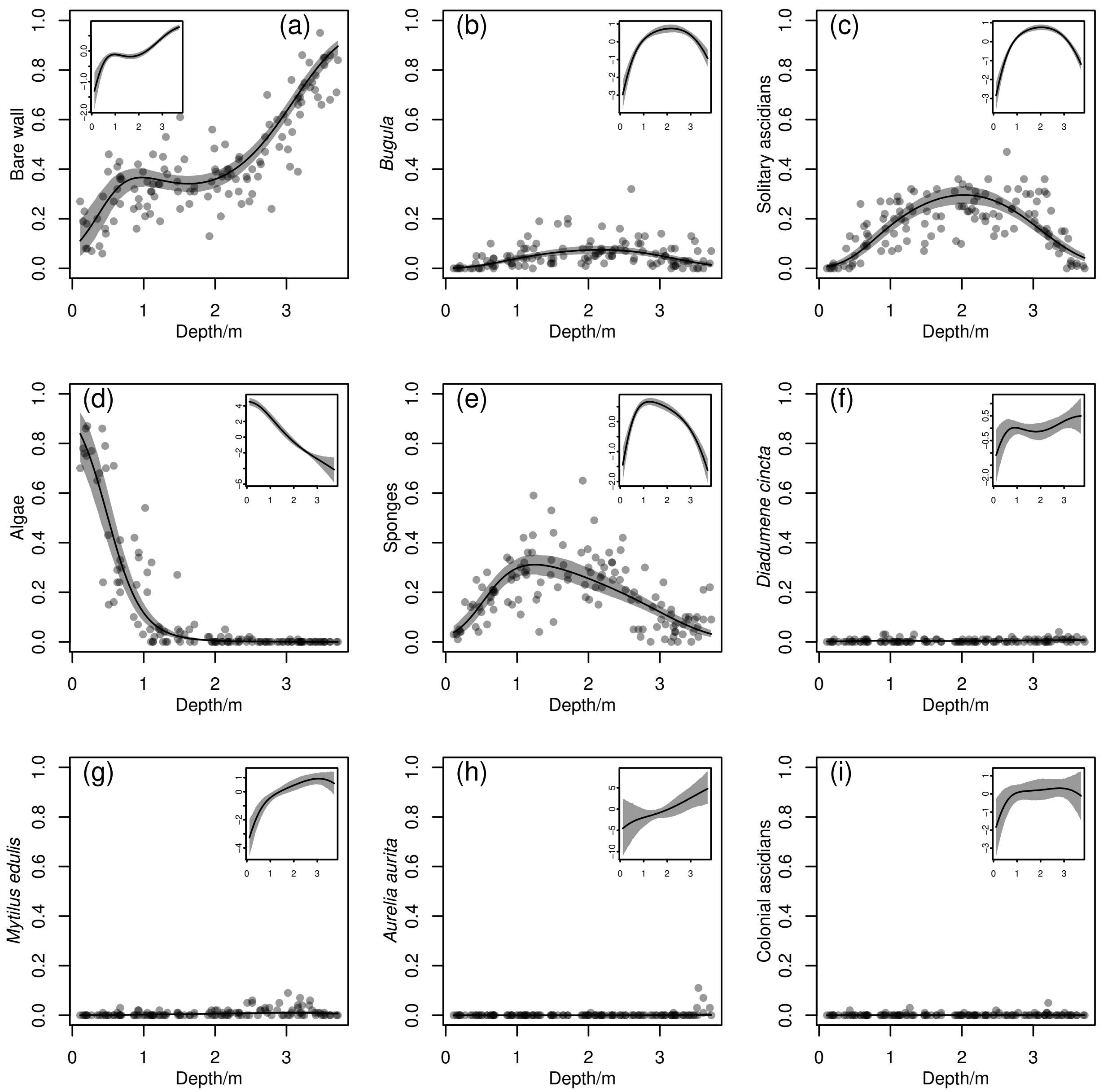

Figure 2: Estimated relationships between relative abundance and depth for bare wall and eight taxa. Circles are sample estimates of relative abundance from point counts. Grey bands are 95% credible bands, and black lines are posterior means. Insets: posterior means and 95% credible bands on a centred logratio scale, in which the value on the *y*-axis is the log of the ratio of the corresponding component to the geometric mean of all components (note the difference in *y*-axis scales among insets).

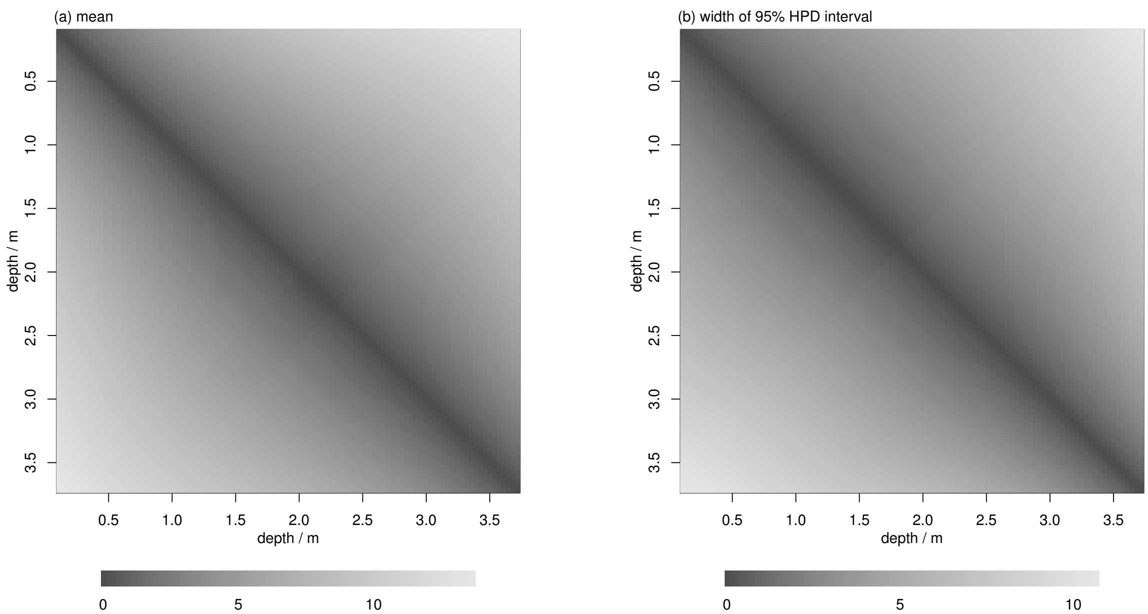

Figure 3: Dissimilarity matrices based on Aitchison distance between expected composition at different depths. Posterior mean (a) and width of 95% highest posterior density intervals (b).

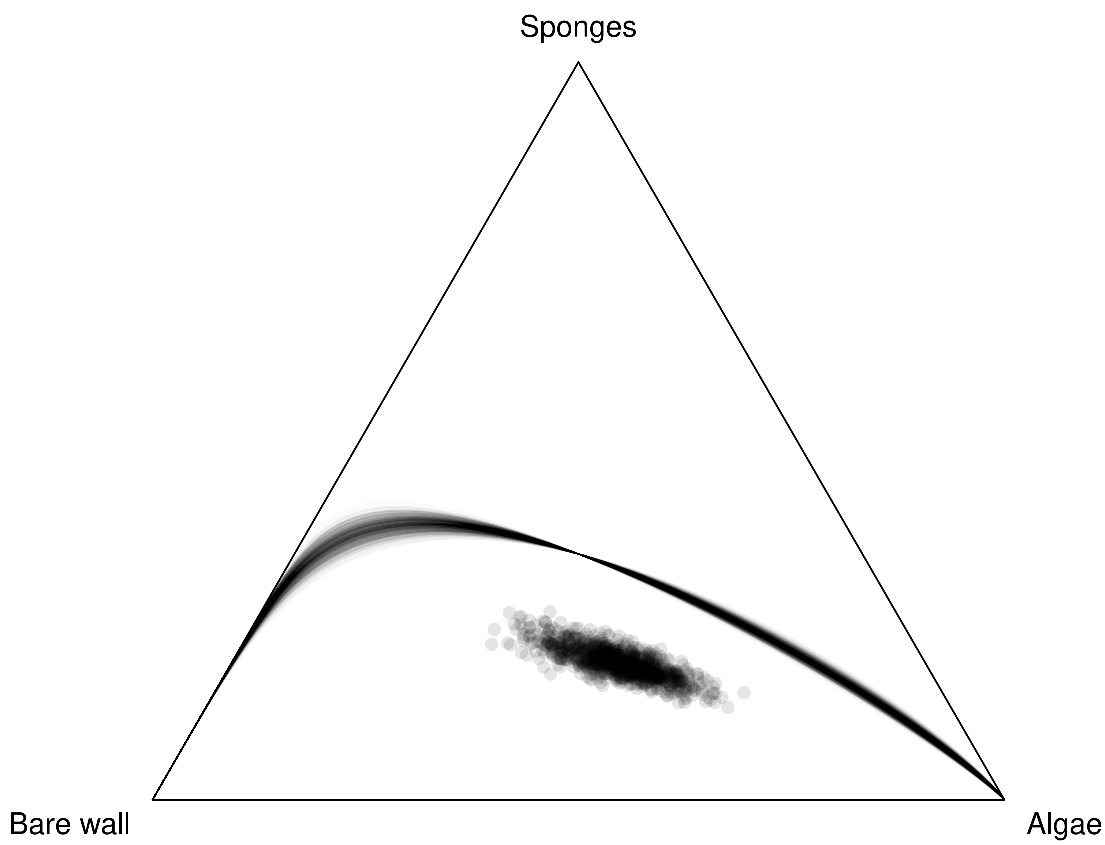

Figure 4: The set of powerings of the depth coefficient γ_1 (lines, sample of 1000 Monte Carlo iterations), and the squared depth coefficient γ_2 (dots: sample of 1000 Monte Carlo iterations), for the subcomposition consisting of bare wall, sponges and algae.

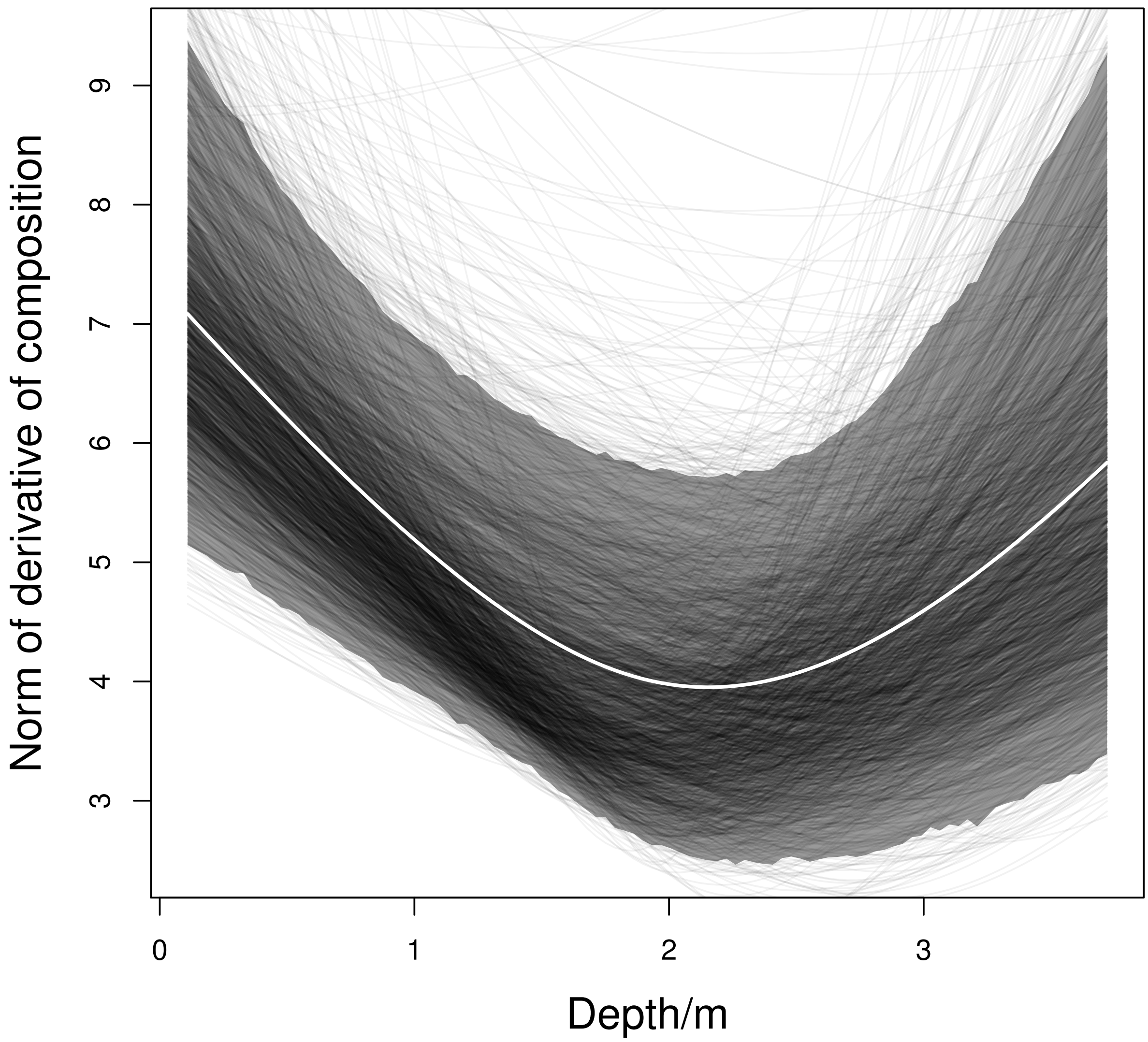

Figure 5: Relationship between rate of change of community composition with respect to depth (the norm of the derivative with respect to depth) and depth. White line: posterior mean. Grey band: 95% credible band. Black lines: norms of derivatives for a subsample of 2000 Monte Carlo iterations.

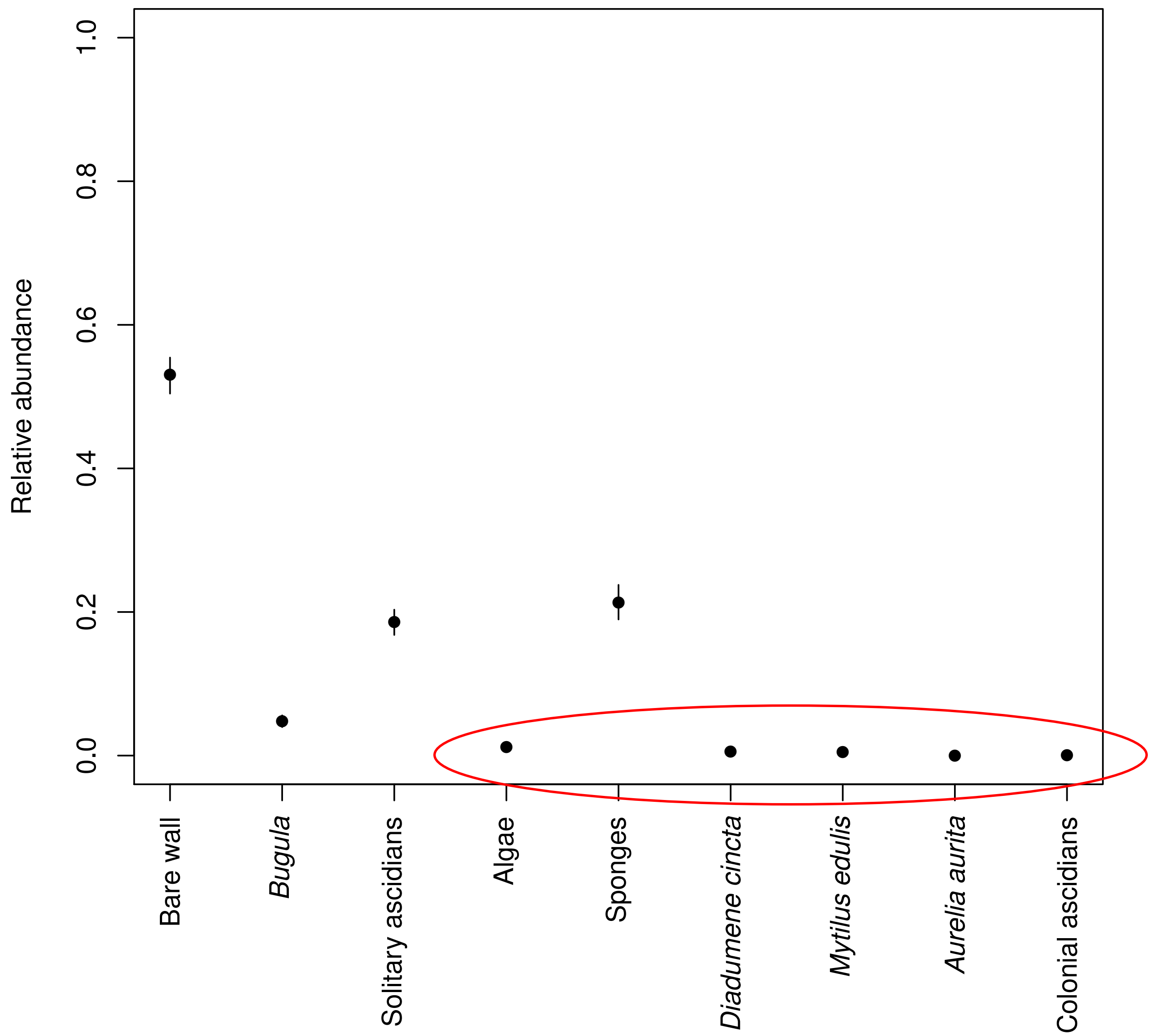

Figure 6: Mean relative abundance of the eight taxa and bare wall, obtained by integration over the entire depth range. Dots: posterior means. Black lines: 95% HPD intervals.