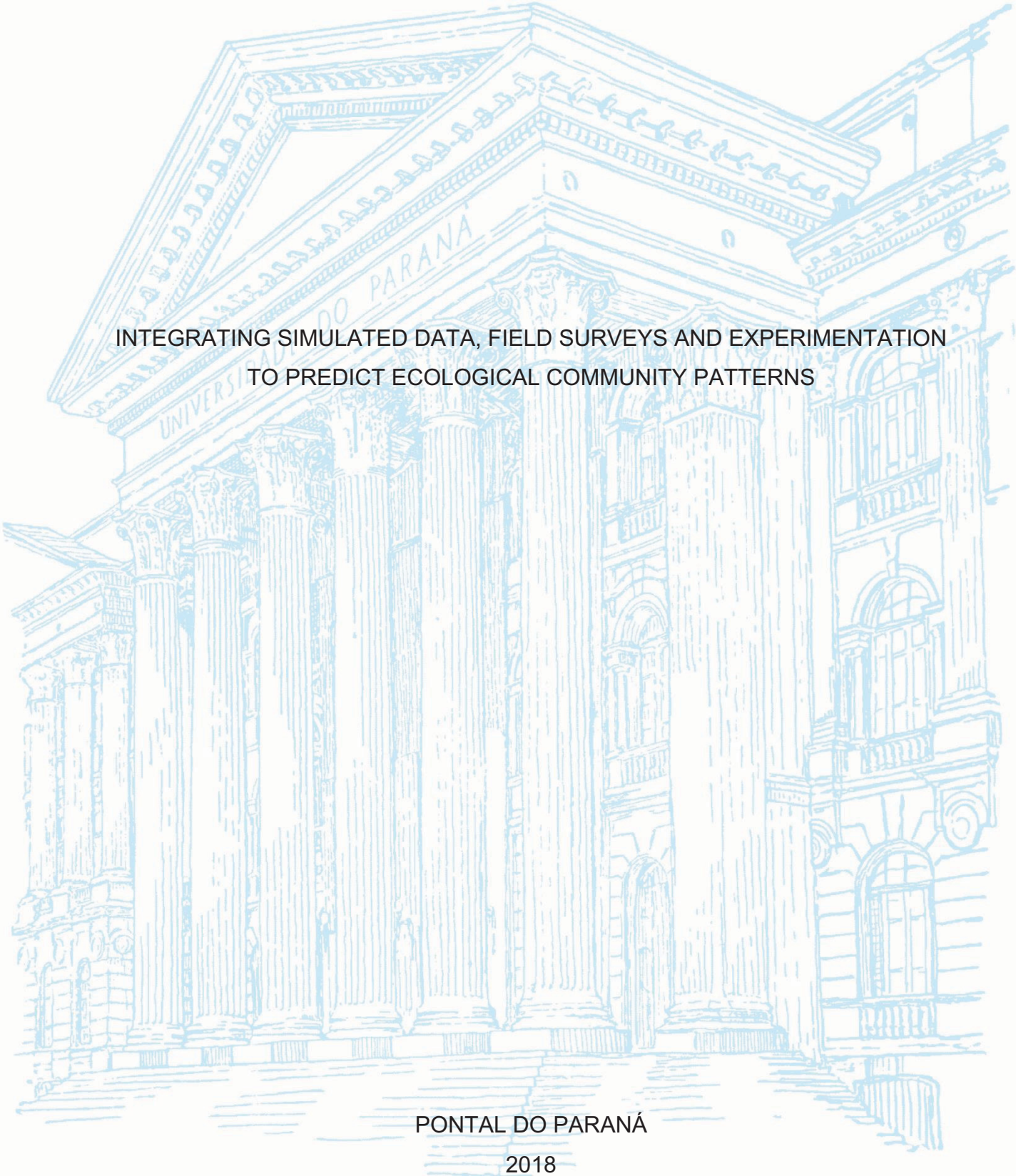
UNIVERSIDADE FEDERAL DO PARANÁ

DANILO CANDIDO VIEIRA



INTEGRATING SIMULATED DATA, FIELD SURVEYS AND EXPERIMENTATION TO PREDICT ECOLOGICAL COMMUNITY PATTERNS

PONTAL DO PARANÁ

2018

DANILO CANDIDO VIEIRA

INTEGRATING SIMULATED DATA, FIELD SURVEYS AND EXPERIMENTATION TO PREDICT ECOLOGICAL COMMUNITY PATTERNS

Tese apresentada ao curso de Pós-Graduação em Sistemas Costeiros e Oceânicos, Centro de Estudos do Mar, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutor em Sistemas Costeiros e Oceânicos

Orientador: Prof. Gustavo Fonseca

PONTAL DO PARANÁ

2018

CATALOGAÇÃO NA FONTE:
UFPR / SiBi - Biblioteca do Centro de Estudos do Mar
Elda Lopes Lira – CRB 9/1295

Vieira, Danilo Candido
V657i Integrating simulated data, field surveys and experimentation to predict ecological community patterns Pontal do Paraná 2018 . / Danilo Candido Vieira. – Pontal do Paraná, 2018.
181 f.; il.; map.; graf.; 29 cm.

Orientador: Profº. Dr. Gustavo Fonseca

Tese (Doutorado) – Programa de Pós-Graduação em Sistemas Costeiros e Oceânicos, Centro de Estudos do Mar, Universidade Federal do Paraná.

1.Oceanografia. 2.Ecologia. 3.Nematóides. I.Título. II.Fonseca, Gustavo. III. Universidade Federal do Paraná.

CDD 551.46

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós- Graduação em SISTEMAS COSTEIROS E OCEÂNICOS da Universidade Federal do Paraná forma convocados para realizar a arguição da tese de Doutorado de **DANILO CÂNDIDO VIEIRA** intitulada: **Integrating simulated data, field surveys and experimentation to predict ecological community patterns**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita a homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-graduação.

Pontal do Paraná, 28 de setembro de 2018.

GUSTAVO FERNANDES CAMARGO FONSECA
Presidente da Banca Examinadora

THIAGO FERNANDO LOPES V. RANGEL
Avaliador externo

LEONARDO SANDRINI NETO
Avaliador interno

PAULO INACIO DE KNEGT LOPEZ DE PRADO
Avaliador externo

PAULO DA CUNHA LANA
Avaliador interno

RESUMO

Estudos ecológicos que integram simulações, observações em campo e experimentos manipulativos são ferramentas poderosas na validação de teorias e no aprimoramento de predições de biodiversidade. O objetivo geral desta tese foi integrar essas ferramentas usando modelagem inversa. Na primeira parte desta tese, introduzimos duas novas ferramentas para modelar comunidades empíricas ao longo de gradientes ambientais: o pacote R “segRDA” e o framework “EB-NB” (Capítulos 2 e 3, respectivamente). SegRDA modela respostas não-contínuas de comunidades ecológicas a gradientes ambientais. Já o framework EB-NB é uma ferramenta para explorar a importância relativa de processos de seleção (baseados em nichos) e não-seleção (baseados em neutralidade) na estruturação de comunidades ecológicas. Após a apresentação desses resultados metodológicos, confrontamos teorias com dados simulados, através de observações de campo (Capítulo 4) e de manipulações experimentais (Capítulo 5). Dados de campo mostraram que a importância relativa dos processos de seleção e não-seleção depende do grupo de organismos em estudo: os nematoides têm um componente de não-seleção maior que a macrofauna. Já as manipulações experimentais mostraram que a capacidade suporte é o fator chave na predição de respostas de comunidades sob diferentes regimes de imigração e perturbação. Juntos, estes estudos mostraram que a integração entre modelagem, observações de campo e experimentos manipulativos pode levar a uma melhor compreensão dos processos que regem a dinâmica de comunidades.

Palavras-chave: Ecologia teórica. Simulação. Estudo de campo. Experimentação. Nematoides.

ABSTRACT

Studies based on theoretical assumptions designed to integrate data from observational surveys, computer simulation and manipulative experiments are powerful tools in validating the theory and enhancing our predictive power of biodiversity patterns. The overall purpose of this thesis was to test the predictions of theoretical models on simulated data, field surveys and on experimentally manipulated communities using inverse modelling designs and individual-based models. In the first part of this thesis, we introduced two new tools for modeling empirical communities along environmental gradients: the segRDA and the EB-NB framework (Chapters 2 and 3, respectively). SegRDA was designed to model non-continuous responses of ecological communities to environmental gradients, whereas EB-NB framework was designed to explore the relative importance of neutral and niche-based assembly processes in structuring ecological communities. Following the methodological results, we confront the theory and simulated data using data from empirical surveys (Chapter 4) and from experimental manipulations (Chapter 5). The empirical survey showed that the relative importance of selection and non-selection processes depends on the group of organisms under study: nematodes have a larger non-selective component compared to macrofauna. The experiments showed that carrying capacity is the key factor in predicting the community responses under different immigration regimes. Together, these chapters showed that the tight interplay between modeling, field observations and experiments can lead to a better understanding of the processes governing community dynamics.

Keywords: Theoretical ecology. Simulation. Field surveys. Experimentation. Nematodes.

SUMÁRIO

1 GENERAL INTRODUCTION AND OUTLINE OF THE THESIS

The overall purpose of my Ph.D. was to test the predictions of theoretical models on simulated data, field surveys and experimentally manipulated communities. Within the broad scope of theoretical ecology, the present thesis has focus on coupling modeling and empirical data using free-living nematodes as a model system. The general introduction of this work will first focus on the importance of conducting integrative research, followed by an overview of concepts of model, modeling and simulations. Subsequently, I introduce the main theories addressed along the thesis and explore the advantages of using nematode assemblages in integrative approaches. I will conclude this overall introduction with the outline of my Ph.D. thesis.

1.1 LINKING EMPIRICAL AND THEORETICAL ECOLOGY

Understanding and predicting the dynamics of organisms in a changing world is a central issue in ecology. Ecology has long recognized the need to work alongside theoretical models, nevertheless the majority of ecological studies are still based either on empirical evidence or on model simulations, theoretical hypothesis-driven studies are lacking (Logue et al. 2011, Codling and Dumbrell 2012). Currently, much of the integration between the empirical and theoretical disciplines of ecology is mostly made conceptually. On one hand, empirical ecologists apply statistical models and numerical approaches to the analysis of their data in attempt to better describe the observed patterns and the underlying theory generally comes *a posteriori*. At the other hand, theoretical ecologists often make biologically unrealistic simplifying assumptions in the pursuit of analytically tractable mathematical models (Codling and Dumbrell 2012).

Collecting data without ecological theory is a rather unfocused endeavor; in the same way, producing mathematical models of ecological systems without validation against real data may be thought-provoking but lacks the necessary link with real communities (Codling and Dumbrell 2012). Hence, studies based on theoretical assumptions and designed to integrate data from field surveys, computer simulations and manipulative experiments are powerful tools in validating theoretical models and enhancing our predictive power of biodiversity patterns (Logue et al. 2011, Vellend et al. 2014, Brown et al. 2017).

1.2 MODELS, MODELING AND SIMULATION

In the literature, the terms model, modeling and simulation are used with multiple meanings. The definition of these terms is particularly important to provide a precise understanding of the context of this research. **Model** is theoretical simplification of a complex reality (Regan et al. 2005) and has three main purposes: **description**, **explanation,** and **prediction** (Hallam and Levin 1986). According to these authors, a descriptive model synthesizes the available information on a process with no real attempt to explain the underlying mechanism. For example, the regression fit to a data is a model in the statistical sense. A model can also be explanatory in a way that makes underlying assumptions about processes and derives logical implications of those assumptions. An example would be the effect of environmental factors on a community in which the abiotic factors are used to explain the community structure. Lastly, a model may be constructed for the purpose of predicting the response of the system to factors which haven't been observed. An example would be determining the effects of the decrease in resources, caused by the increase in density, in experimental units.

The word "**modeling"** is used in ecological research in different ways, ranging from mathematical models used to develop theoretical ecology to statistical models used to analyse empirical data (Ovaskainen et al. 2016). A way of classifying different modelling approaches is that of "forward" (from mechanisms to patterns) and 'inverse' (from patterns to mechanisms) approaches. **Forward approaches** are those where **assumptions are made about the underlying mechanisms**, **and mathematical or simulation tools are used to study the consequences of those assumptions**. The forward approaches operate on global laws defined by the equations (equation-based) (Hallam and Levin 1986) and requires the simplification of assumptions to define and solve the equations (Codling and Dumbrell 2012). The main advantage of this approach is that the system need not be physically built in order to predict its behavior (Reddy 2011). However, when describing a number of interacting and structured populations, purely mathematical techniques are overcome (Marilleau et al. 2018). For example, although forward models may recognize the interlocking individual variation of organisms, these variations have no explicit representation in the models (Van Dyke Parunak et al. 1998). Consequently, when the dynamics are noncontinuous or

nonlinear, local variations from the averages can lead to significant deviations in overall system behavior (Marilleau et al. 2018).

A broad description of **inverse approaches** is that they pertain to the case when the system under study already exists, and one **uses measured or observed system behavior to aid in the model building and/or refinement** (Reddy 2011)**.** The recent advances in ecological modeling techniques have allowed the development of models in a manner that combines elements from both the forward and the inverse approaches. Like the forward approach, the researcher uses prior knowledge about the ecological context to formulate a model, with the **hypotheses on** the **parameters and processes constrained by the observed patterns** (Wiegand et al. 2003)**.** 'Pattern-oriented modelling' (POM, *sensu* Grimm 2005) is an inverse modelling technique in ecology that considers the use of multiple field data pattern simultaneously to filter the parameterizations which were successfully tested against all available data on system dynamics (Kramer-Schadt et al. 2007). The modeling steps must be considered as a cycle (Grimm and Railsback, 2005; Schmolke et al., 2010), with the idea that each step can be done several times, possibly using methods of increasing complexity (Courbaud et al. 2015).

S**imulation** is defined in the frame of modeling as a set of techniques that allow an examination of the dynamic behaviour of models (Matějíček 2002). The model can be reconfigured and experimented with; usually, this is impossible or impractical to do in the system it represents (Maria 1997). Simulations are particularly valuable when they incorporate detailed information of the system through direct observation or experiments rising higher-level features and patterns (Latombe et al. 2011, Grimm et al. 2017). In this context**, individual-based models (IBM)** are increasingly used to simulate ecological communities (Railsback 2008). In these models, the actions of unique individuals are simulated as they interact with each other and with the environment. In spatially explicit IBMs, species can vary according to their environmental fitness and the environment can be characterized by key ecological drivers (Smith and Lundholm 2010).

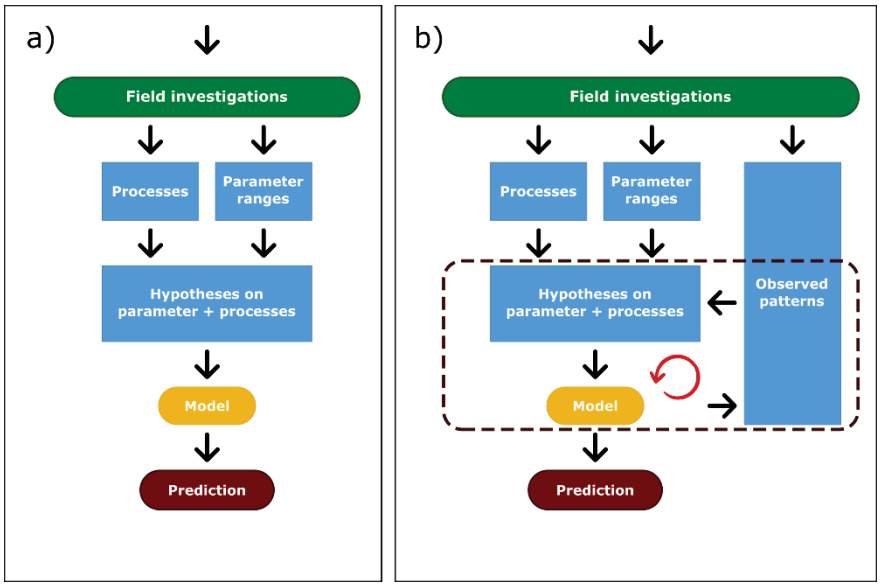


Fig.1. Forward (a) and inverse (b) modelling. Figure adapted from Wiegand et al. 2003.

1.3 NICHE AND NEUTRAL-BASED MODELS

In community ecology, niche and neutral theories are the two competing families of theoretical models that aim at explaining community biodiversity and structure patterns (Gravel et al. 2006, Adler 2011). **Niche-based** models state that each species is adapted to exploit a partly unique niche as a result of environmental filtering and biotic interactions (Leibold 1995, Chase and Myers 2011). On the other hand, the key factors for the **neutral theory** are stochastic processes such as colonization (which is often at least partly determined by chance because it depends, among other things, on distance among patches, on passive dispersal and on the availability of habitat patches), random extinctions and ecological drift in determining such patterns (Hubbell 2001). In recent years, ecologists have observed that both niche and neutral processes can drive native community patterns, suggesting that community composition at any point should be the result of **an interplay of stochastic processes, dispersal limitation, environmental filtering and species interactions** (Gravel et al. 2006, Latombe et al. 2015)**.**

1.4 FREE-LIVING NEMATODES AS A MODEL SYSTEM

In general, nematodes are the **dominant meiofaunal** (organisms ranging from 0.045 and 0.5 mm) taxon in marine sediments (Lambshead et al. 2003). For a number

of reasons, the use of nematodes as model system can facilitate the task of anchoring theoretical predictions with observed data: (1) they are characterized by a high local species richness and abundance (Coull, 1999; Heip et al., 1985), at a scale which is amenable to control and adequately replicate in laboratory conditions (tens of species and hundreds to thousands of individuals in patches as small as 10 cm2); (2) they exhibit short life cycles in the order of days to few months (Heip et al., 1985), which allows relevant data on population dynamics; and finally (4) because species-specific characteristics can be inventoried, which is an essential step for the design and implementation of niche-based models and for the comparison of empirical and model results.

1.5 AIMS AND OUTLINE OF THE THESIS

The overall purpose of this thesis was to test the predictions of theoretical models on simulated data, field surveys and experimentally manipulated communities. All studies from this thesis share common approaches: inverse modeling, individual-based simulations, and free-living nematodes as a study case.

Chapter 2 introduces the segRDA package for modeling non-continuous responses in ecological communities. The assumption of linearity is unrealistic in most ecological communities and is only done because more appropriate methods of analysis are not available (Makarenkov and Legendre 2002). We proposed the combined use of the split-moving-window analysis (SMW, Ludwig and Cornelius 1987) and the piecewise redundancy analysis (pwRDA, described here) to identify ecological discontinuities and model species-habitat associations that may differ between two communities that occur along à environmental gradient. This analytical routine was explored with three datasets; two data sets generated from spatially-explicit IBMs and one coastal dataset of benthic communities. The proposed method was implemented as an R package (segRDA), available at https://github.com/DaniloCVieira/segRDA.

Chapter 3 describes the EB-NB framework for exploring the relative importance of deterministic and stochastic processes in structuring ecological communities. The framework is based on two species-level metrics: environmental boundary (EB) and niche breadth (NB). In this chapter, EB is defined as the entire environmental range occupied by a given species, whereas NB reflects the variance of resources used by a species along an environmental gradient (MacArthur 1972,

Swihart et al. 2006, Hirzel and Le Lay 2008). The EB-NB framework proceeds in comparing empirical measurements of EB and NB with model generated-expectations based on spatially-explicit IBMs.

Chapter 4 uses the EB-NB framework to test the importance of deterministic and stochastic processes across groups of organisms with different dispersal capabilities and environmental tolerances: meiofauna and macrofauna. These benthic groups are distinct from each other in terms of reproduction (all meiofauna are in situ breeders), dispersal (no meiofaunal organisms have a specific dispersal phase), and life histories (most meiofaunal juveniles resemble the adults). Given these differences in lifestyles and dispersal strategies, we hypothesized that macrofauna species would be more environmentally-restricted (i.e., species with narrow EB and NB) than meiofaunal species.

Chapter 5 tests the ability of IBMs to predict the response of free-living marine nematodes to events of disturbance and immigration. Disturbance is the major cause of local extinction, whereas immigration acts in the opposite direction, preventing the local extinction of species. This study used inverse modeling techniques to model the hypothesis that disturbance associated with low immigration rates decreases abundance and diversity, whereas disturbance associated with high immigration result in a more abundant, diverse and evenly distributed assemblage.

Finally, in the last chapter I integrate and discuss the obtained results to build a comprehensive framework on how the tight interplay between modeling, field observations and experiments can lead to a better understanding of the processes governing community dynamics.

2 CHAPTER 2 - A PIECEWISE REDUNDANCY ANALYSIS FOR MODELING NON-CONTINUOUS ECOLOGICAL COMMUNITIES

Danilo Cândido Vieira[1,2]; Marco Colossi Brustolin[1,2], Fábio Cop[2], Gustavo Fonseca[2]

[1]Centro de Estudos do Mar - Universidade Federal do Paraná - Caixa Postal 50.002, 83255-000, Pontal do Paraná, PR, Brazil.

[2] Universidade Federal de São Paulo – Av. Dona Ana Costa, 95 – CEP 11060-001, Santos, SP, Brazil.

Corresponding author: vieiradc@yahoo.com.br

phone: +55 16 993 626 902

Keywords: community discontinuity, canonical analysis, piecewise redundancy analysis

2.1 RESUMO

A análise de redundância (RDA) é amplamente usada para correlacionar matrizes de espécies e de descritores ambientais. Uma limitação da RDA é que a variável resposta (i.e. a matriz de espécies) assume uma resposta linearmente contínua ao longo de todo gradiente. No entanto, em grandes gradientes, respostas não contínuas são comuns em comunidades ecológicas. Neste contexto, este artigo propõe a análise de redundância segmentada (pwRDA). Esta análise foi explorada com três dados simulados e um conjunto de dados costeiros de comunidades bentônicas. A pwRDA permite a quebra da relação entre as variáveis resposta e explicativas. O ponto de quebra entre as comunidades é avaliado através da análise split moving-window (SMW). Ambas as análises são apresentadas no pacote segRDA para o ambiente R. Em todos os exemplos, os modelos pwRDA explicaram maiores porcentagens da variância total nas variáveis de resposta que o tradicional RDA. Em comunidades com acentuada transição entre as partes, os modelos pwRDA diferiram significativamente do tradicional modelo de RDA. Para transições mais suaves, o pwRDA não diferiu do tradicional RDA. O uso combinado de SMW e pwRDA permite identificar pontos de quebra e zonas de transição entre comunidades ecológicas ao longo de um gradiente. Esta abordagem é particularmente relevante quando as associações de habitats de espécies diferem entre as comunidades. Para gradientes com comunidades descontínuas, essa abordagem aumenta nosso poder preditivo.

2.2 ABSTRACT

Redundancy analysis is widely used to correlate species matrix against environmental descriptors. One restriction of redundancy analysis is that the response variable (i.e., the species matrix) assumes the same continuous linear trend over the whole gradient. However, over large gradients, non-continuous responses are common in ecological communities. In order to tackle this issue, this paper proposes the piecewise redundancy analysis (pwRDA). The analysis is explored with three simulated data and a coastal dataset of benthic communities. The pwRDA allows the relationship between the response and explanatory variables to break into pieces. The breakpoint between communities is assessed by the split-moving window (SMW) approach. Both analyses are presented within segRDA package for R environment. In all the examples, pwRDA models accounted for greater percentages of the total variance in the response variables compared with traditional RDA. In communities with a sharp transition between pieces, the pwRDA models significantly differed from the traditional RDA model. For smooth transitions, the pwRDA did not differ from the traditional RDA.The combined use of SMW and pwRDA allows one to identify breakpoints and transition zones among ecological communities along a gradient. This approach is particularly relevant when species-habitat associations may differ among communities. For gradients with discontinuous communities, this approach enhances our predictive power.

2.3 INTRODUCTION

One of the most challenging tasks of ecological studies is to model the response of communities to environmental conditions, which often requires the use of canonical multivariate analyses (Legendre and Legendre 2012). Among the multitude of canonical methods, the redundancy analysis (RDA, Rao 1964) is widely used to reduce the dimensions of species multivariate data along environmental gradients while retaining the trends and patterns. However, such analysis assumes a linear and constant trend of each species over the whole gradient (Legendre et al. 2011), which is arguably often not the case (Vergnon et al. 2012). Ecological responses can change from one state to another once a boundary is crossed in space (Beckage et al. 2007). As such, if there are two or more contrasting boundaries along a gradient and distinct community responses associated to it, the traditional RDA will fail to detect such discontinuities, returning a model with high unexplained variance (Peres-Neto et al. 2006, Shi et al. 2015).

Non-continuous responses in univariate data are traditionally analyzed by means of piecewise (or segmented) regression (Lerman 1980, Toms and Lesperance 2003). By breaking the original regression model in pieces, it minimizes the sum of squares of the differences between observed and predicted values of the response variable. The significance of the pieced model is assessed by comparing its sum of squares with those estimated from the original regression model. The same principles of the piecewise simple linear regression can be applied to multiple regression models (Ertel & Fowlkes 1976), and lastly extended for building piecewise-RDA (pwRDA) models. The pwRDA analysis has much of the simplicity of the classical RDA methodology but allows explanatory-response relationships to vary in segments.

This paper describes the pwRDA for analyzing discontinuous communities along gradients. Because the procedure depends on locating the breakpoints (i.e., the community boundaries), we preceded the pwRDA with the split-moving window analysis (SMW, Ludwig & Cornelius 1987), a simple and powerful method widely used to detect discontinuities in ordered datasets. Both analyses are included within `segRDA` package for R environment. We illustrate the use of `segRDA` using simulated and empirical examples. As others R packages, `segRDA` can be combined with other tools available in R and investigators can customize the routine for their own purposes.

2.4 METHODOLOGY

Since the pwRDA requires prior identification of the community's breakpoints, we start by describing the SMW analysis. Subsequently, we introduce the pwRDA analysis followed by a short overview of the segRDA package. Finally, we describe the simulated and the empirical data used to illustrate the application of segRDA.

2.4.1 Split-moving-window analyses (SMW)

The pwRDA requires a prior identification of the community's breakpoints. Statistically, a breakpoint can be defined as the location where the highest rate of change occurs (Burrough 1986). The SMW is a non-parametric method that identifies discontinuities within multivariate data ordered in one dimension (Ludwig and Cornelius 1987). The application of SMW applied to a species matrix consists of (1) placing a window of even-numbered size at the beginning of the data series, (2) splitting the window into two equal halves, (3) calculating the community centroids within each half, (4) computing a dissimilarity metric between the two halves, (5) shifting window one position along the series, and (6) repeating the procedure till the end of the data series (Cornelius & Reynolds 1991). The significance of the dissimilarity values within each window is tested through multi-response permutations and includes two types of procedures: the "random-plot" and "random-shift". The first randomizes sites along the series while preserving species composition and abundance structure of the sites. The second randomizes patterns of each species relative to each other while preserving the spatial structure of the species abundance. Therefore, choosing between random plot and random shift relies on considering sites or species as a fixed aspect of the null distribution. Both procedures include computing an expected mean dissimilarity $\overline{\mathbf{DS}}$ and standard deviation **SD** for each window midpoint. Confidence limits have been suggested as one or two **SD** above $\overline{\mathbf{DS}}$ or estimated from one-tailed 95% confidence intervals (Erdos et al. 2014). Lastly, dissimilarity profile graphs are constructed by plotting the dissimilarity values vs. the location of the window midpoint. Peaks in dissimilarity suggest the locations of breakpoints.

The choice of window size affects the results of the SMW analysis. Small windows result in many peaks that represent small-scale variation. In contrast, large windows reduce the number of peaks, smoothing the small-scale variation. Computing multiple dissimilarity profiles with increasing window sizes reduce the scale-effect (Körmöczi et al. 2016), assuming that breakpoints corresponding to ecologically meaningful boundaries would persist. Dissimilarity profiles of the mean Z-score (standardized values of dissimilarity) can be used to detect the community breakpoints. The SWM algorithm used in `segRDA` package is presented in Cornelius & Reynolds (1991).

2.4.2 Piecewise redundancy analysis

Once the breakpoints have been chosen, the pwRDA is implemented. Based on the same principles of the multiple piecewise regression (Liu et al. 1997), the pwRDA aims to maximize the explained variance of the response variables by a linear combination of the explanatory variables. Let **Y** (with *n* samples and *p* species) be the response matrix, and **X** (with *n* samples and *m* environmental variables) the explanatory matrix. The classical RDA ("full" model) can be summarized in three main steps: (1) computing a matrix of fitted values $\widehat{\boldsymbol{Y}}_{\boldsymbol{full}}$ through independent multiple regression of each column in **Y** on **X** (eq. 1.1, Table 1); (2) performing a principal component analysis (PCA) of $\hat{Y}_{full}$ to produce the canonical eigenvalues, eigenvectors and axes (object ordination scores) (eq. 1.2, Table 1); and (3) generating a triplot diagram using the ordination scores (eq. 1.3, Table 1).

The pwRDA replaces the step 1 of the classic multiple regression with a piecewise multiple regression (Ertel & Fowlkes 1976). Let x_{ij} represent the jth independent variable for the ith sample. The piecewise multiple regression divides the explanatory matrix into *k* segments, structuring **X** as a block-diagonal matrix **Xb**, in which the number of columns equals the number of explanatory variables times the number of segments (Fig. 1).

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \\ X_{41} & X_{42} & X_{43} \\ X_{51} & X_{52} & X_{53} \\ X_{61} & X_{62} & X_{63} \end{bmatrix} \longrightarrow \mathbf{X_b} = \begin{bmatrix} X_{11} & X_{12} & X_{13} & 0 & 0 & 0 \\ X_{21} & X_{22} & X_{23} & 0 & 0 & 0 \\ X_{31} & X_{32} & X_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & X_{41} & X_{42} & X_{43} \\ 0 & 0 & 0 & X_{51} & X_{52} & X_{53} \\ 0 & 0 & 0 & X_{61} & X_{62} & X_{63} \end{bmatrix}$$

Fig. 1: An example of a Block-diagonal matrix in which the first 3 rows are in segment one and the last 3 rows in the second segment.

The remainder of the analysis does not differ from the traditional RDA (steps 2 and 3). Another additional difference involves the computation of the scores of explanatory variables **X** for the biplots. The traditional RDA model obtains the species scores **U** through PCA of $\hat{Y}$ and produces a triplot diagram using three ordination elements: species score **U** (eq. 2.1), site scores **F** (eq. 3.1) and, a biplot **BP** produced from the correlations between the explanatory variables **X** and the linear constraints **Z** (eqs. 4.1 and 5.1). In the pwRDA, a biplot $\mathbf{BP}_{pw}$ produced from the block diagonal matrix generates as many axes as the number of columns in **Xb**. In order to represent each explanatory variable by a single arrow in the triplot, the pwRDA estimates $\mathbf{BP}_{pw}$ from the multiple correlation of **X** and its block diagonal form **Xb** with the linear constraints $\mathbf{Z}_{pw}$ (eq. 5.2, Table 1). This method has also been used for polynomial piecewise redundancy analysis (Makarenkov and Legendre 2002).

The adjusted R^2_{adj} (Appendix 1, eqs. 7 to 10) measures the strength of each RDA model (*full* and *pw*) and the permutational procedures allows to test their significance (Legendre & Legendre 2012). The pwRDA model considers the following statistical tests:

The permutation procedure tests the significance of the pwRDA model. Permutations generate *N* random samples from the original data. The pwRDA model is then refitted using the permuted samples and the resulting $R^2_{(adj|pw)}$ tabulated. Significance is evaluated with 95% one-tailed test.

Table 1: Main equations for performing a traditional (full) and a piecewise (pw) redundancy analyses. **X –** explanatory variables; **Y –** response variables; $\mathbf{Y}_c$ **–** centered response variable; **Xb** – block diagonal matrix; $\hat{Y}$ – the predicted values; **U** – the species scores; **F** – the site scores; **Z** – the linear constraints; **BP** – biplot; PCA - principal component analysis.

Eqs.		$\mathbf{1.\ RDA}_{full}$		$\mathbf{2.\ RDA}_{pw}$
1	(1.1)	$\widehat{\mathbf{Y}} = \beta\mathbf{X}$	(1.2)	$\widehat{\mathbf{Y}}_{pw} = \beta\mathbf{Xb}$
		$\beta = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Y}$		$\beta_{pw} = [\mathbf{Xb}'\mathbf{Xb}]^{-1}\mathbf{Xb}'\mathbf{Y}$
2	(2.1)	$\mathbf{U} = \mathrm{PCA}(\widehat{\mathbf{Y}})$	(2.2)	$\mathbf{U}_{pw} = \mathrm{PCA}(\widehat{\mathbf{Y}}_{pw})$
3	(3.1)	$\mathbf{F} = \mathbf{Y_c U}$	(3.2)	$\mathbf{F}_{pw} = \mathbf{Y_c U}_{pw}$
4	(4.1)	$\mathbf{Z} = \widehat{\mathbf{Y}}\mathbf{U}$	(4.2)	$\mathbf{Z}_{pw} = \widehat{\mathbf{Y}}_{pw}\mathbf{U}_{pw}$
5	(5.1)	$\mathbf{BP} = cor\ (\mathbf{X}, \mathbf{Z})$	(5.2)	$\mathbf{BP}_{pw} = cor\ (\mathbf{cor}(\mathbf{X}, \mathbf{Xb}), \mathbf{Z}_{pw})$

The F-ratio tests if the difference between the residual sum of squares for the full RDA model and a pwRDA model is significantly different (Appendix 1, eq. 11). A non-significative F-ratio means that dividing the community into more than one segment does not decrease the residual terms of the model and thus a traditional RDA analysis would be the most appropriate.

2.4.3 The segRDA package

The `segRDA` package is organized straightforward through three steps: (a) data ordering, (b) split-moving-window analysis and (c) piecewise redundancy analysis.

1. The analysis of community changes along a given gradient begins by sequentially ordering the data in question (i.e., the community table). In `segRDA`, the function `OrdData` defines the order of samples using one of the axes of a classical RDA model. The recommendation is to use the first and most explanatory axis.
2. The function `SMW.test` implements the algorithms for performing the SMW analysis and includes arguments allowing users to choose a dissimilarity metric (i.e., all distances metrics available in `vegan::vegdist`), the type of randomization for testing the significance of the dissimilarity values (“random shift” or “random plot”) and the statistical test to detect significant dissimilarities (peaks standing 1SD or 2SD above the expected dissimilarity, 95% one-tailed test and Z-score higher than a critical z-value). The function `PoolSMW` creates a pooled

profile by averaging together dissimilarities from different window sizes (for each window midpoint location). In order to avoid scale-dependency effects, we recommend the pooled version of the analysis with the Z-scores higher than 1.85 as significant (Erdos et al. 2014).

3. Lastly, the function `pwRDA` implements the algorithms for performing the pwRDA analysis. This function generates objects from both the "full" (`rda.0`) and the "pw" (`rda.pw`) RDA models. `pwRDA` function also generates a data frame (`pw$summ`) containing the summarized statistics $R^2_{(adj|full)}$, $R^2_{(adj|pw)}$, F-ratio (full/pw) and their respective p-values. The p-value from $R^2_{(adj|full)}$ is calculated using `anova.cca` function from the vegan package with defaults parameters.

All functions in `segRDA` the package depend on `vegan` (Oksanen et al. 2017). The vignette presented in Appendix 2 shows a more detailed analytical demonstration of the functions available in `segRDA`, including auxiliary functions for plotting the results.

2.4.4 Simulated and empirical examples

This paper explores the performance of `segRDA` using three simulated communities (`sim1`, `sim2`, and `sim3`) from niche-based, spatially-explicit models (package `neutral.vp`, Smith and Lundholm 2010). Simulations were idealized to contain 100 samples and two potential communities with distinct optimal distributions over a simulated environmental gradient. The simulations varied in terms of overlapping sites. This was achieved by controlling species fitness along the environmental gradient, dispersion and immigration rates (Appendix 3). The breakpoints of each simulated dataset were identified using pooled dissimilarity profiles from five different window sizes (10, 20, 30, 40 and 50).

Additionally, an empirical example is provided. This dataset consists of 141 benthic samples with 9 environmental variables (dataset `nema)` and 194 free-living marine nematode species (dataset `nema)` of the Araçá Bay (southeastern coast of Brazil, Corte *et al.* 2017; Appendix A1). Nematode data were transformed using a Hellinger transformation prior to the analyses (Legendre & Legendre 2012). Dissimilarity profiles were obtained using a pooling from seven different windows sizes (10, 20, 30, 40, 50, 60 and 70).

For all the SMW analyses, the order of the sites was based on the first RDA axis and, significance tests were based on random-shift permutations. Random-shift is considered the best choice for two reasons: (a) because z-scores are less scale-dependent than in random plot and (b) because it preserves the community patterns within the sites (Erdos et al. 2014).

2.5 RESULTS

The resulting Z-score profiles for the three simulated datasets provided consistent estimates for both, number and location of community breakpoints along the gradient (Fig. 2). `sim1` showed one breakpoint along a sharp unimodal dissimilarity profile. `sim2` presented one breakpoint but with a diffuse transition between communities. `sim3` exhibited two distinct breakpoints with communities largely overlapping each other. For the empirical data `nema`, SMW analysis revealed two statistically significant breakpoints along the gradient.

The four full-RDA models showed that the environmental variables significantly explain changes in community structure (Table 1). All the pwRDA analyses produced significant models (after 999 permutation tests) and increased the R^2_{adj} when compared to the traditional RDA model. The increased pwRDA explanation was not statistically significant in the simulation with a diffuse transition between communities (sim2; Fig 2). In contrast, pwRDA significantly improved the model in simulations with sharper transitions between communities (sim1 and sim3). For the empirical data, the pwRDA analysis explained 38.9% of the variance which was significantly higher when compared to the full model (Table 1). The stations arranged between the two peaks of the z-score profile represented the transition between the two communities (nema; Fig 2).

Table 2 – Summary statistics from piecewise redundancy analysis. R^2_{full} and R^2_{pw}: adjusted R squares from the full and pieced RDA model, respectively. sumAxis12: percentage of variance explanation accounted for the first two canonical axis.

Dataset	Statistic		P.value	sumAxis12
sim1	$R^2_{adj\|full}$	0.646	0.001	0.277
	$R^2_{adj\|pw}$	0.825	<0.001	0.348
	$\mathrm{F}_{pw/full}$	3.11	0.008	
sim2	$R^2_{adj\|full}$	0.667	0.001	0.2
	$R^2_{adj\|pw}$	0.768	0.002	0.229
	$\mathrm{F}_{pw/full}$	2.012	0.072	
sim3	$R^2_{adj\|full}$	0.547	0.001	0.187
	$R^2_{adj\|pw}$	0.887	<0.001	0.291
	$\mathrm{F}_{pw/full}$	2.504	0.007	
nema	$R^2_{adj\|full}$	0.269	0.001	0.163
	$R^2_{adj\|pw}$	0.389	<0.001	0.198
	$\mathrm{F}_{pw/full}$	2.387	0.003	

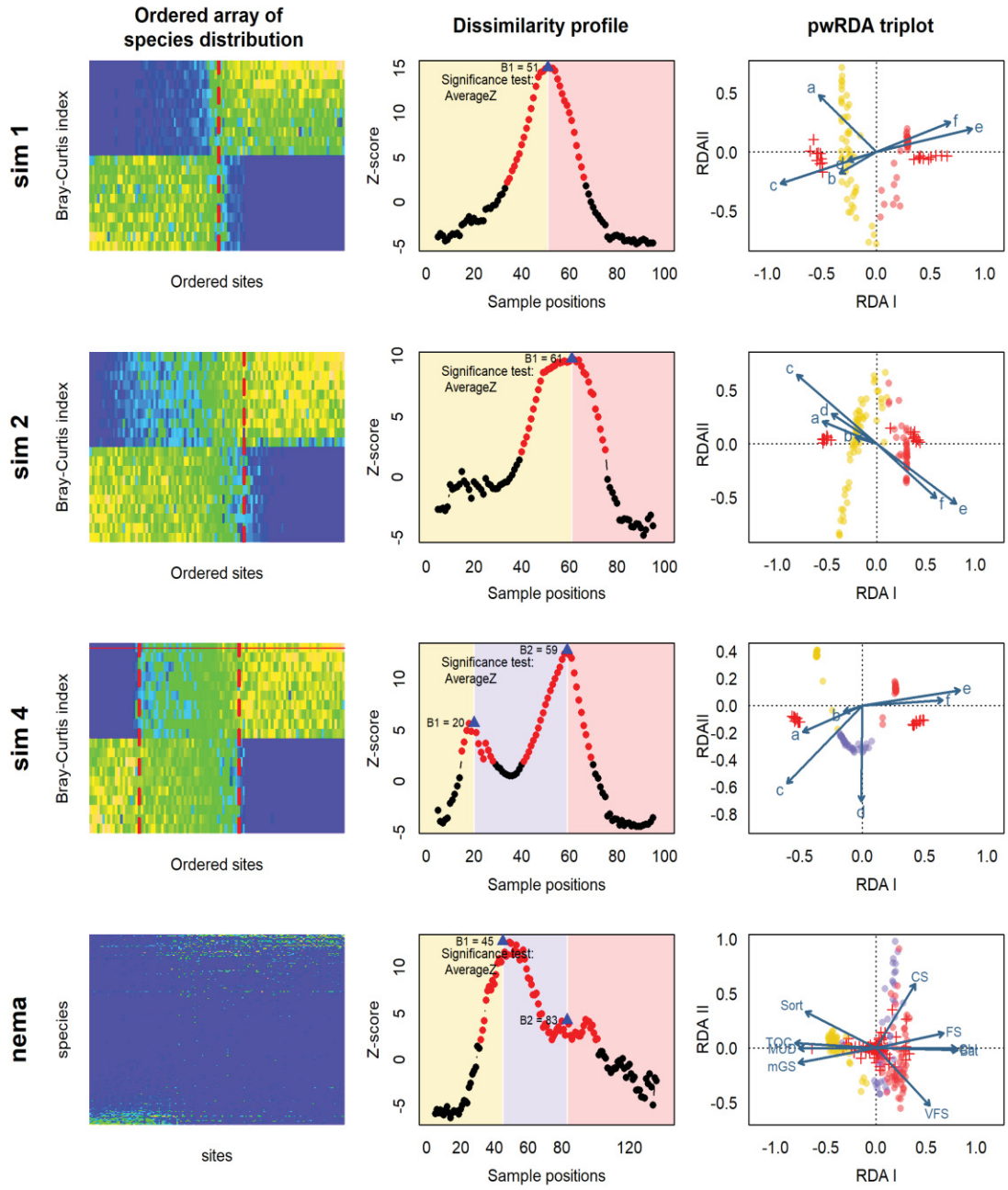


Fig. 2: Ordered array of species distribution, dissimilarity profiles, and pwRDA triplot from three simulated datasets - sim1, sim2 and sim3; and from the empirical dataset nema. The ordered arrays are displayed as a heatmap of abundance values, in which the dashed red line indicates breakpoints locations. Dissimilarity profiles and pwRDA triplots are colored according to their breakpoints locations. In pwRDA triplots: a-f: simulated environmental variables; Chl - chlorophyll a; TOC - total organic carbon; Bat - bathymetry; MUD- mud content; VFS - very fine sand; FS - fine sand; CS - coarse sands; mGS - mean grain size; Sort – sorting.

2.6 DISCUSSION

The present paper proposes the piecewise redundancy analysis for modeling non-continuous responses of communities along environmental gradients. The pwRDA decreases the residual terms of the traditional RDA by dividing the community matrix into groups according to selected breakpoints. The core of pwRDA is the use of piecewise multiple regression during the first stage of RDA, instead of the traditional multiple regression. Makarenkov & Legendre (2002) previously extend the traditional RDA using polynomial multiple regressions. Both polynomial and piecewise RDA models deal with the non-linearity problem, but while the polynomial RDA assumes non-linear continuous responses, the pwRDA assumes non-continuous linear responses. Therefore, the suitability of the pwRDA analysis relies on the nature of change in the relationship between response and explanatory variables. The simulated results from this paper evince that abrupt changes in ecological communities are better modeled by pwRDA models than by the traditional RDA model.

Because the regression coefficients of each species change over the range of the ordering variable, it is important to note that defining the number and the location of breakpoints that best represents the data are a critical step before pwRDA analysis. The R package `segRDA` offers a single interface for ordering the target community along the first RDA axis and to identify relevant breakpoints. Important considerations regarding the `segRDA` routine are:

Data ordering: Data-ordering is a user-defined situation. In complex data structures, for example, more than one axis might be important to explain the community variation (Schmera et al. 2018). This means that same community matrix reordered along different axes might reveal contrasting community structures.

SMW analysis: Ecologists have a range of methods for detecting ecological discontinuities. Alternative methods to the SMW include Mantel tests, wombling analysis, semi-variograms, analysis for spatial autocorrelation among others (see Hufkens, Scheunders & Ceulemans 2009 for review). As in SMW, all these methods include a description of spatial rates of change in the target community to determine the locations (or zones) of greatest change (Barbujani et al. 1989, Kent et al. 2006). Compared to these methodologies, the SMW is relatively easier to implement and interpret, and allows to measure the strength and width of “changing zones”. A possible

drawback of SMW is that it assumes samples to be equally distributed along the environmental gradient. Therefore, the scale-dependence of the algorithm together with an arbitrary choice in sample spacing can introduce bias (Hufkens et al. 2009).

pwRDA analysis. As in the traditional RDA, overfitting in pwRDA occurs when a response variable is fitted using a number of explanatory variables larger than the number of samples (Legendre and Legendre 2012). Note that, the number of explanatory variables in pwRDA is multiplied by the number of groups within the community. Therefore, reducing the number of explanatory variables may be necessary to avoid overfitting the response variables.

The pwRDA is so far the unique canonical multivariate approach for modeling non-continuous community responses to environmental data and can be applied to a variety of situations, such as landscape ecology, impact assessments, time series, etc. The `segRDA` package offers a comprehensive way to implement the pwRDA analyses and allows the investigator to test a wide range of hypotheses involving species-habitat associations that may differ along gradients. More importantly, this approach enhances the predictive power in modeling community discontinuities along environmental gradients.

2.7 APPENDIX 1 – MAIN EQUATIONS

Main equations for performing a traditional ($full$) and a piecewise (pw) redundancy analyses. $\mathbf{X}$ – explanatory variables; $\mathbf{Y}$ – response variables; $\mathbf{Y_c}$ – centred response variable; $\mathbf{Xb}$ – block diagonal matrix; $\widehat{\mathbf{Y}}$ – the predicted values; $\mathbf{U}$ – the species scores; $\mathbf{F}$ – the site scores; $\mathbf{Z}$ – the linear constrains; **BP** – biplot; PCA – principal component analysis; **RSS** – residual sum-of-squares; **TSS** – Total sum-of-squares; $\mathbf{R^2}$– R squared;

$\mathbf{R^2_{adj}}$– adjusted R squared; $\mathbf{F}_{stat}$ – F statistics.

		$\mathbf{1.\ RDA}_{full}$		$\mathbf{2.\ RDA}_{pw}$
1	(1.1)	$\widehat{\mathbf{Y}} = \beta\mathbf{X}$ $\beta = [\mathbf{X'X}]^{-1}\mathbf{X'Y}$	(1.2)	$\widehat{\mathbf{Y}}_{pw} = \beta\mathbf{Xb}$ $\beta_{pw} = [\mathbf{Xb'Xb}]^{-1}\mathbf{Xb'Y}$
2	(2.2)	$\mathbf{U} = \mathrm{PCA}(\widehat{\mathbf{Y}})$	(2.2)	$\mathbf{U}_{pw} = \mathrm{PCA}(\widehat{\mathbf{Y}}_{pw})$
3	(3.1)	$\mathbf{F} = \mathbf{Y_cU}$	(3.2)	$\mathbf{F}_{pw} = \mathbf{Y_cU}_{pw}$
4	(4.1)	$\mathbf{Z} = \widehat{\mathbf{Y}}\mathbf{U}$	(4.2)	$\mathbf{Z}_{pw} = \widehat{\mathbf{Y}}_{pw}\mathbf{U}_{pw}$
5	(5.1)	$\mathbf{BP} = cor\ (\mathbf{X}, \mathbf{Z})$	(5.2)	$\mathbf{BP_{pw}} = cor\ (\mathbf{cor}(\mathbf{X}, \mathbf{Xb}), \mathbf{Z}_{pw})$
6			(6.4)	$\mathbf{Xb} = \begin{matrix} \mathbf{X}_{(1,1)} & 0 & 0 & (for\ x_0 \leq x_i \leq x_1) \\ 0 & \mathbf{X}_{(2,2)} & 0 & (for\ x_1 \leq x_i \leq x_2) \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \mathbf{X}_{(k,k)} & (for\ x_{k-1} \leq x_i \leq x_k) \end{matrix}$
7		$\mathbf{RSS} = \sum(\widehat{\mathbf{Y}} - \overline{\mathbf{Y}})^2$		
8		$\mathbf{TSS} = \sum(\mathbf{Y} - \overline{\mathbf{Y}})^2$		
9		$\mathbf{R^2} = \dfrac{\mathbf{RSS}}{\mathbf{TSS}}$		
10		$\mathbf{R^2_{adj}} = 1 - (1 - R^2_{Y\|X})\dfrac{(n-1)}{(n-m-1)}$		
11		$\mathbf{F}_{stat} = \dfrac{\dfrac{\mathbf{RSS}_{full} - \mathbf{RSS}_{pw}}{(p_f - p_r)}}{\dfrac{\mathbf{RSS}_{full}}{(n-p_f)}}$		

2.8 APPENDIX 2 - VIGNETTE

segRDA: An R package for performing piecewise redundancy analysis
Danilo Cândido Vieira; Marco Colossi Brustolin; Fabio Cop; Gustavo Fonseca.
August 31, 2018

The `segRDA` requires the `vegan` package (Oksanen et al. 2017). `segRDA`compiles three functions: (1) data ordering, (2) breakpoints detection and (3) piecewise redundancy analysis.

The package includes the following functions: - `OrdData`: designed to order the target data; - `SMW.test` and `PoolSMW`: designed for performing SMW analyses; -`SMW.test`: designed to locate the breakpoints along a dissimilarity profile; - `pwRDA`: designed to perform the pwRDA analysis; - `Dprofile` and `Dprofile2`: designed to plot the SMW outputs.

The code below demonstrates how to use the `segRDA` package. Throughout this document, we will use the dataset `sim2` from the package, which is composed of two matrices: `envi` and `comm`.

```
data(sim2) ##Simulated data
x<-sim2$envi ## matrix of explanatory variables
y<-sim2$comm ## matrix of response variables
```

A transformation of y is recommended to ensure the homoscedasticity of the response variables. One option is to use the function "*decostand*" from `vegan` package.

```
y<-decostand(y,"hell")
```

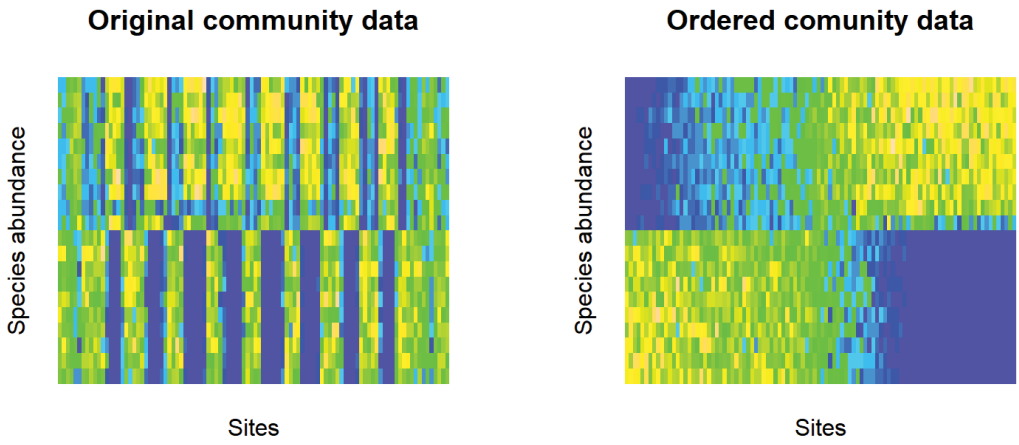
Data ordering

Both SMW and pwRDA analysis depends on ordered datasets. The ordering can be user-defined or generated using `OrdData` function. The latter orders both response and explanatory matrices using one of the axes of a full RDA model. Defaults to using the first RDA axis.

```
sim.ord<-OrdData(x=x,y=y, axis=1)
x.ord<-sim.ord$x
y.ord<-sim.ord$y
```

The resulting ordered community can be represented by a species-site interaction matrix using the function `image`, which is available from the stats package of R.

```
par(mfrow=c(1,2), mgp=c(1,1,0))
image(y, main="Original community data", col=topo.colors(100), axes=F, xlab="Sites", ylab="Species abundance")
image(y.ord, main="Ordered community data", col=topo.colors(100), axes=F, xlab="Sites", ylab="Species abundance")
```



Breakpoints detection

SMW with randomization tests

The function `SMW.test` generates dissimilarity profiles with randomization tests using a single window size `w`.This function includes arguments allowing users to choose the dissimilarity metric (e.g. all distances metrics available in `vegan::vegdist`), the type of randomization (`"shift"` or `"plot"`) and the statistical test used to detect significant discontinuities (`"Z"`,`"SD"`,`"2SD"` or `"Tail1"`). Further details are in `help(SWM.test)`. Note that the running time of the analysis depends on the number of randomizations (argument `"n.rand"`). In the following examples, we settled a low number of randomizations to speed up the analyses. By setting `progress=TRUE` the progress of the analysis is shown while running is exhibited.

```
smw.10<-SMW.test(y=y.ord,w=10, n.rand=10, progress=TRUE) ## window size 4
smw.26<-SMW.test(y=y.ord,w=26, n.rand=10, progress=TRUE) ## window size 20
```

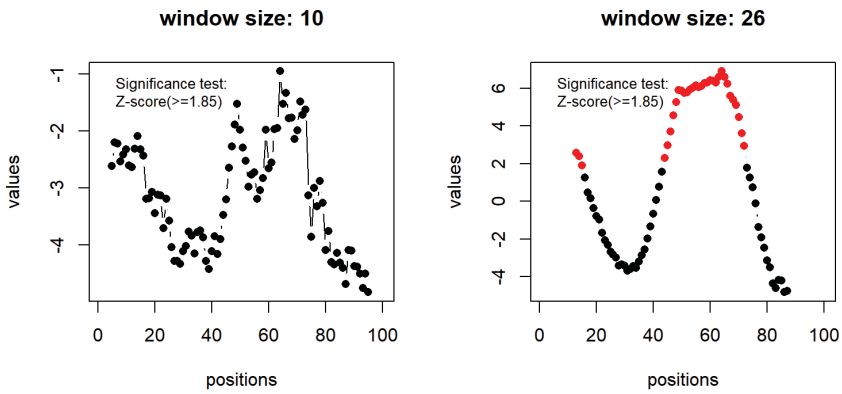
The main output of `SMW.test` is an object of class `smw`, which contains the resulting dissimilarity profile table (`$DP`). For a full description on `SMW.test` outputs please go to `help(SMW.test)`.

```
ls(smw.26) ## List of objects in SMW.test output
## [1] "D.overall"  "Dmean"      "DP"         "DP.rand"    "params"
## [6] "SD"         "SD.overall"
head(smw.26$DP) ## Dissimilarity profile table
##   positions sampleID      diss    zscore sig:Z
## 1        13       62 0.1499717 2.5808025  TRUE
## 2        14       12 0.1468783 2.3994498  TRUE
## 3        15       78 0.1384266 1.9039579  TRUE
## 4        16        8 0.1273701 1.2557588 FALSE
## 5        17       51 0.1139244 0.4674857 FALSE
## 6        18       20 0.1089361 0.1750446 FALSE
```

Dissimilarity profile graphs can be constructed by plotting the location of the window midpoint `$DP$position` vs. the dissimilarity values (`$DP$diss`). This plot can be automatically generated by applying the function `Dprofile` on the object returned by `SMW.test`.

```
par(mfrow=c(1,2))
Dprofile(smw=smw.10, main="window size: 10")
Dprofile(smw=smw.26, main="window size: 26")
```

`Dprofile` automatically displays the significant dissimilarity values (in “red”). It can be customized through other graphical parameters. By default, a legend informing the significance test performed is displayed (`legend=TRUE`).



The function `BPlocation` is an auxiliary tool for locating the breakpoints in the dissimilarity profile table.

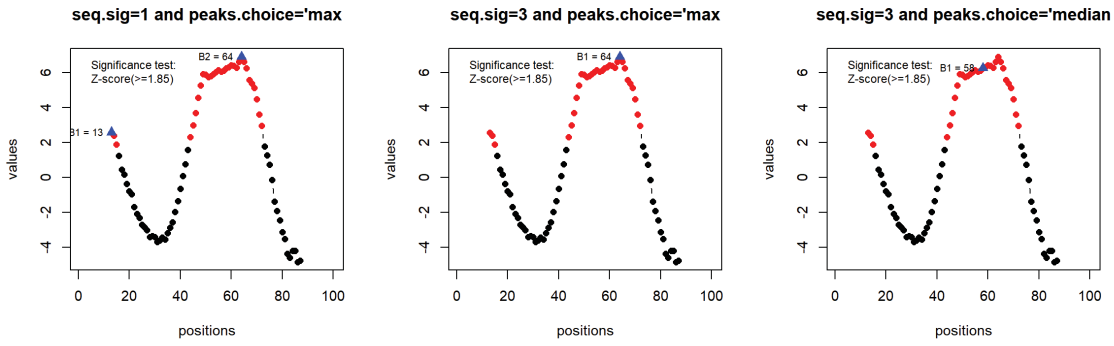
`BPlocation` allows users to define the criteria that will be used to locate the community breakpoints. It is based on two arguments: `"seq.sig"` and `"peaks.choice"`. The argument `seq.sig` specifies the length of a sequence of significant dissimilarity values that will be considered in defining the community breakpoints. The argument `peak.choice` defines if the breakpoints should be chosen as those site positions corresponding to the maximum dissimilarity in the sequence (`peak.choice="max"`) or

as those site positions corresponding to the median position of the sequence. The function returns a subset of the original `smw` table containing only the samples suggested as breakpoints.

```
BP.1<-BPlocation(smw=smw.26, seq.sig = 1, peaks.choice = "max")
BP.2<-BPlocation(smw=smw.26, seq.sig = 10, peaks.choice = "max")
BP.3<-BPlocation(smw=smw.26, seq.sig = 10, peaks.choice = "median")
BP.1
##    positions sampleID      diss   zscore sig:Z
## 1         13       62 0.1499717 2.580802  TRUE
## 52        64       83 0.2237242 6.904636  TRUE
BP1<-BP.1$positions
BP2<-BP.2$positions
BP3<-BP.3$positions
```

Breakpoints can be displayed in the dissimilarity profiles using the function `Dprofile` and specifying the argument `BPs`.

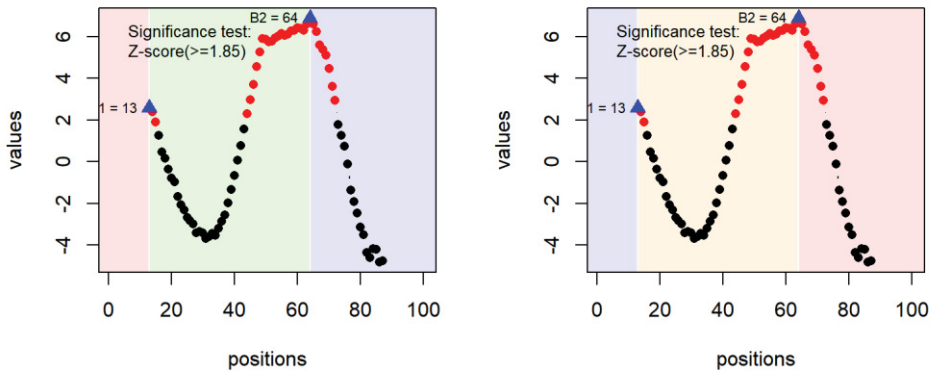
```
par(mfrow=c(1,3), cex=.9)
Dprofile(smw.26,BPs=BP1, main="seq.sig=1 and peaks.choice='max")
Dprofile(smw.26,BPs=BP2, main="seq.sig=3 and peaks.choice='max")
Dprofile(smw.26,BPs=BP3, main="seq.sig=3 and peaks.choice='median")
```



An optional customization of SMW plot is available for colouring the background area of the graph according to the breakpoints location. This is done by the argument `bg`,

which specifies the transparency of the background (between 0 and 1). The colours are specified as a vector in the argument `bg.colours`. When `bg` is specified and `bg.colors` is null, the function uses the colour palette `rainbow` from R package base.

```
par(mfrow=c(1,2))
Dprofile(smw.26,BPs=BP1, bg=.1,bg.colors=NULL)
Dprofile(smw.26,BPs=BP1, bg=.1, bg.colors  =c("blue","orange","red"))
```



Pooled SMW with randomization tests

The function `PoolSMW` creates a pooled SMW dissimilarity profile by averaging together dissimilarities profiles from several different window sizes (for each window midpoint location). The window sizes to be analysed are specified as a vector in the argument `windows`. All windows sizes must be even. Because results of SMW are scale-dependent, dissimilarity values for each window are transformed into standardized variables (Z scores). Note that the running time of this analysis will be proportional to the number of randomizations (`n.rand`) times the number of window sizes analyzed (`length(windows`). The progress of the analysis while running is automatically displayed.

```
pooled.sim<-PoolSMW(y=y.ord,windows=c(10,20,30,40,50), n.rand=10)
## Progress 20 %
## Progress 40 %
## Progress 60 %
## Progress 80 %
## Progress 100 %
```

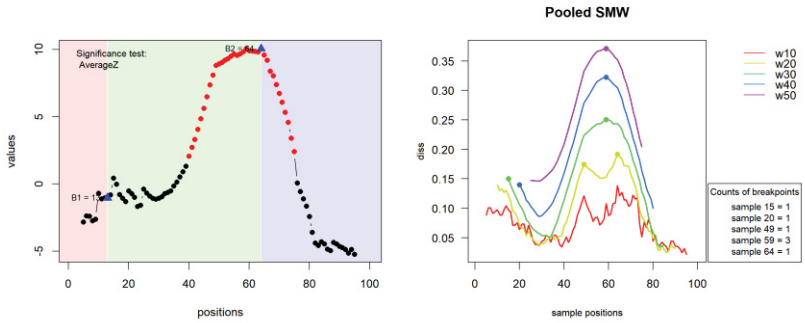
This function returns an smw object of class `pool` similar to the output from `SMW.test`.

```
ls(pooled.sim) ## List of objects in SMW.test output
## [1] "DP"            "params"        "result.pooled"
head(pooled.sim$DP) ## Dissimilarity profile table
##   positions SiteID   AverageZ sig:AverageZ
## 1         5     29 -2.8195057        FALSE
## 2         6     39 -2.3722139        FALSE
## 3         7     88 -2.3933331        FALSE
## 4         8     69 -2.7349501        FALSE
## 5         9     72 -2.6086584        FALSE
## 6        10     18 -0.6874934        FALSE
```

The object returned by `PoolSMW` can also be used by both `Dprofile` and `Dprofile2`. The latter drawns together the profiles obtained from the different sizes of windows. When `count='TRUE'`, the function uses `BPlocation` with defaults parameters to count the number of breakpoints. `BPlocation` arguments can be passed to `Dprofile2`.

```
par(mfrow=c(1,2),cex=1.2)
Dprofile(pooled.sim, BP=BP1, bg=.1)
```

```
Dprofile2(pooled.sim)
```



Piecewise redundancy analysis (pwRDA)

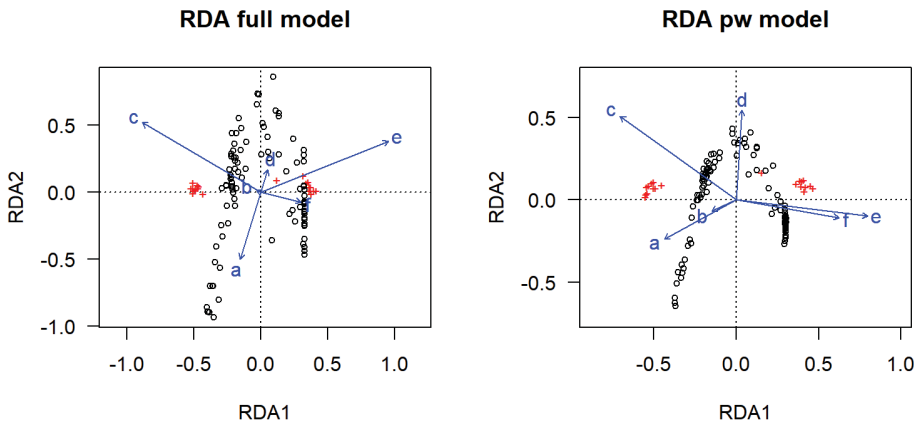
Once the breakpoints have been defined and their significance tested, the pwRDA can be implemented. It is done with the function `pwRDA`. The function has three arguments: `x` explanatory variables, `y` response variables and `cps` the positions of the breakpoints. The function displays a message with the main result of the analysis:

```
pw.sim<-pwRDA(x.ord=x.ord,y.ord=y.ord, BPs=BP1, progress=F)
##
##      pwRDA analysis
##  ------------------------
##  ** Summary statistics **
##  ------------------------
##      Statistic      P.value
## FULL 0.6669929 0.0010000000
## PW   0.8137186 0.0000011983
## F    1.2932628 0.2384139295
##  ------------------------
##
```

The function returns an invisible list with following descriptors: - `$summ`: a vector containing the summarized statistics. - `$rda.0` and `$rda.pw`: the respectives `cca`

objects from the full and “pw”" RDA models. These objects can be used in all functions that applies to `cca.object`. More details are in `help(cca.object){vegan}`.

```
head(pw.sim$summ)
##      Statistic      P.value
## FULL 0.6669929 0.0010000000
## PW   0.8137186 0.0000011983
## F    1.2932628 0.2384139295
par(mfrow=c(1,2))
plot(pw.sim$rda.0, main="RDA full model", las=1)
plot(pw.sim$rda.pw, main="RDA pw model", las=1)
```



2.9 APPENDIX 3 - SIMULATED DATA

Simulated data were generated using the “neut.simulate” function (*neutral.vp* package, Smith and Lundholm 2010) . This package implements a spatially explicit, individual-based model in which individuals are assumed to have different probabilities of establishment according to the environment. This type of simulation allows creating communities with well-defined boundaries, which is an essential for testing the pwRDA performance.

The environmental gradient was simulated using the function “habitat.builder”. This function returns a list of matrices with different environmental configuration: linear gradient, linear gradient orthogonal to the first, a unimodal hump, a four unimodal hump, a linear wave, and a random configuration. Then, we ran a PCA on these variables (previously normalized) and used the first PC scores as the environmental template for all simulations.

In order to create two optimal distributions of species, two values of fitness were attributed to each half of the community, correspondent to 0.25 and 0.75 percentiles of the PC1. The remainder of the parameterization is detailed in the table below.

Table 1 - Parameters used in the *neut.simulate* function.

Number of species (*S*)	20
Carrying capacity (*K*)	200
Time steps (*cycles*time*)	10000
Base birth rate (*b*)	0.2
Base death rate (*d*)	0.00005
Dispersal (*u*)	0.5
Immigration (*m*)	0.1
Ef. Strength (*ef.strength*)	0.026

3 A SIMULATION-BASED FRAMEWORK TO EXPLORE THE IMPORTANCE OF NON-SELECTION AND SELECTION PROCESSES IN STRUCTURING ECOLOGICAL COMMUNITIES

Danilo Cândido Vieira[1,2]; Gustavo Fonseca[2]

[1]Centro de Estudos do Mar - Universidade Federal do Paraná - Caixa Postal 50.002, 83255-000, Pontal do Paraná, PR, Brazil.

[2] Universidade Federal de São Paulo – Av. Dona Ana Costa, 95 – CEP 11060-001, Santos, SP, Brazil.

Corresponding author: vieiradc@yahoo.com.br

phone: +55 16 993 626 902

Keywords: dispersal, environmental boundary, environmental filtering, niche breadth, pattern-oriented modeling.

3.1 RESUMO

O objetivo deste framework é identificar a importância relativa dos processos de seleção e dispersão na estruturação de comunidades ecológicas. Baseado em uma abordagem de “modelagem orientada a padrões”, o framework consiste em cinco etapas: (a) agregação das informações da comunidade empírica e seu ambiente, (b) simulação das comunidades sob diferentes níveis de dispersão e seleção, (c) seleção das melhores simulações para compor um modelo único (modelo composto) usando a borda ambiental (EB) e a amplitude de nicho (NB) observada para cada espécie, (d) validação do modelo composto através da comparação entre resultados observados e previstos de três padrões da comunidade e (e) classificação de cada espécie observada ao longo do continuum seleção / não seleção. Dados de nematoides marinhos de vida livre de uma baía costeira foram usados como exemplo empírico. Um total de 20 simulações foram realizadas, variando os níveis de seleção e dispersão. Na ausência de seleção, as espécies dos modelos de alta dispersão apresentaram elevados EBs e NB, enquanto que os modelos baseados em seleção geraram espécies com valores mais estreitos de EB e NB. Os valores de EB e NB diminuíram com a dispersão. O modelo composto incluiu 96% das 194 espécies de nematoides e predisse três padrões observados: (1) distribuição do rank de abundâncias, as estruturas de assembleias ao longo do gradiente (2) espacial e (3) ambiental. Os modelos de não-seleção e seleção representaram 34% e 85% das espécies observadas, respectivamente. A principal vantagem dessa abordagem é que as medidas empíricas de nicho são colocadas no contexto das expectativas geradas pelo modelo, permitindo uma compreensão muito mais profunda dos processos estruturadores da comunidade, e de como estes processos variam de espécie para espécie.

3.2 ABSTRACT

The purpose of the current framework is to identify the relative importance of selection and dispersion processes in structuring ecological communities. Based on a "pattern-oriented modeling" approach, it consists of five steps: (a) aggregate information from the empirical community and its environment, (b) simulate communities under different degrees of dispersal and selection, (c) select the best set of simulations into a composite model using the environmental boundary (EB) and niche breadth (NB) of each observed species, (d) validate the composite model by comparing expected and observed results from three additional community patterns and (e) classify each observed species along the selection/non-selection continuum. A free-living marine nematodes dataset from a coastal bay was used as an empirical example. A total of 20 simulations were performed varying selection and dispersion levels. In the absence of selection, species from high-dispersal models showed maximum EBs and NBs, whilst selection-based models generated species with narrower EB and NB values. EB and NB values declined with decreasing dispersal. The composite model encompassed 96% of the 194 nematode species and predicted all the three patterns evaluated: (1) abundance-rank distribution, the assemblage structures along both the (2) spatial and (3) environmental gradients. Non-selection and selection models accounted for 34% and 85% of the observed species, respectively. The main advantage of this approach is that empirical niche measurements are placed in the context of model-generated expectations, enabling a much deeper understanding of community assembly processes and how they vary from species to species.

3.3 INTRODUCTION

With the growing recognition that the relative importance of dispersal and environmental filtering is crucial in predicting communities' dynamics, several studies have also sought to quantify these processes both empirically and theoretically (e.g.,Bonthoux and Balent 2015, Gascón et al. 2016). In empirical studies, a common approach is the use of variance-partitioning methods to decompose variation in community composition explained by spatial, environmental, and spatially structured environmental processes (Peres-Neto et al. 2006). However, in such analysis, environmental filtering and dispersal may be confounded when dispersal limitation correlates with the spatial arrangement of the environment (Smith and Lundholm 2010). Besides, it assumes that species have linear responses to ecological gradients, which is arguably often not the case (Vergnon et al. 2012). Yet, in theoretical studies, simulations are powerful tools in predicting communities' dynamics under different assumptions of dispersal and environmental filtering (Brown et al. 2017). Even so, there is still a gap between theoretical predictions and empirical data, thus limiting us to place real communities on a continuum between two extremes of environmental filtering and dispersal.

Species response curves along environmental gradients can provide important insight into mechanisms underlying community dynamics (Basille et al. 2008, Hirzel and Le Lay 2008). The relation between two metrics, environmental boundary (EB) and niche breadth (NB), may be especially useful to explore the relative importance of dispersal and environmental filtering. EB represents the complete environmental range accounting for species occurrences including sites in which the species' presence is the result of random dispersion (Swihart et al. 2006), while NB is defined as the variance of resources used by a species along a gradient (MacArthur 1972). Species under the extreme influence of environmental filtering are unlikely to survive outside their optimal habitat (Keddy 1992), resulting in narrow EB and NB (Hirzel and Le Lay 2008). As the environmental filter loses strength and dispersal increases, the species abundance decreases with increasing distance from their optimum habitat (Mouquet and Loreau 2003), resulting in wider EB compared to its NB. Finally, upon a non-selection influence scenario, dispersal is the predominant force shaping species displacement, irrespective of the environmental gradient (Biswas and Wagner 2012). *Accordingly*, dispersal-driven patchy distributions are expected under low dispersal

rates. However, if dispersal is strong enough to homogenize the system, EB and NB would be as large as the sampling area.

Based on these assumptions, the present study integrates several analytical tools into a multi-step framework to quantify the relative importance of dispersal and environmental filtering in empirical communities. Here, we examine the role of such drivers using a pattern-oriented modelling approach, in which multiple observed patterns are used to evaluate model's proficiency in capturing real-community dynamics (Grimm 2005). We generated and explored two families of models (non-selection and selection-based models) and compared how well they perform at predicting the empirical distribution of EBs and NBs. The proposed approach builds on the "outlying mean index" (OMI) analysis, which is an ordination technique robust enough for unimodal, linear, or a mixture of species response curves (Dolédec et al. 2000). In addition, this framework offers a direct graphical interpretation and allows inferences at species levels. We used a free-living marine nematodes' data set from a coastal bay as an empirical example. Marine nematodes are benthic organisms that inhabit the interstitial matrix of marine sediments and are extremely abundant and species-rich (Giere 2009); two important advantages for testing the theoretical framework of community ecology.

3.4 METHODOLOGY

The proposed framework combines empirical measurements of environmental boundary and niche breadth (Fig. 1) into a two-dimensional EB-NB space to quantify the relative importance of dispersal and environmental filtering of empirical communities. Measurements of EB and NB were both determined by means of the "Outlying Mean Index" (OMI) analysis using the R package "ade4" (Dray and Dufour 2007). It consists of running a principal component analysis (PCA) on the environmental table and associating the resulting table of row profiles with the respective faunistic table, therefore giving the average position (i.e., niche position) of each species along the ordination axes. The OMI analysis also provides a niche breadth value, which represents the total variance of the environmental table weighted by the species' abundances. Hence, in the present study, EB and NB were respectively defined as the 95% of the interquartile range and the standard deviation of the distribution of each species along the first PCA score. Interquartile range often

provides the measure of variability for distributions that are less sensitive to extreme scores (Gravetter and Wallnau 2016).

A sensitivity analysis based on randomized subsamples was performed to examine how robust the observed EB-NB results are to variation of the sampling effort. PC1 explanation and EB-NB relationship were summarized using profile plots of increasing sample effort together with their bootstrapped 95 % confidence intervals (0.025 and 0.975 percentiles; 100 bootstraps). The statistics used to estimate the EB-NB relationship was based on generalized least squares models (GLS, package "nlme") corrected for heterogeneity of variances (exponential variance structure). Akaike Information Criterion (AIC) was used as a measure of the GLS model strength.

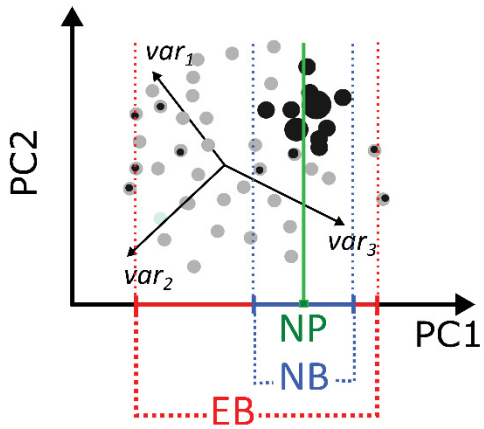


Fig. 1. The OMI analysis is used to obtain the niche position (NP), environmental boundaries (EBs) and niche breadths (NBs) of species along the first principal component (PC1). The black circles represent species' presence, which is scaled to their abundances. Note that the EB is always equal to or larger than the NB.

A sensitivity analysis based on randomized subsamples was performed to examine how robust the observed EB-NB results are to variation of the sampling effort. PC1 explanation and EB-NB relationship were summarized using profile plots of increasing sample effort together with their bootstrapped 95 % confidence intervals (0.025 and 0.975 percentiles; 1,000 bootstraps). The statistics used to estimate the EB-NB relationship was based on generalized least squares models (GLS, package "nlme") corrected for heterogeneity of variances (exponential variance structure). Akaike Information Criterion (AIC) was used as a measure of the GLS model strength.

The development framework is based on pattern-oriented modeling (Grimm 2005) and can be broken into five key steps (Fig. 2):

1. **Empirical data:** aggregate information from the empirical community and its environment.
2. **Simulations:** simulate communities under different degrees of dispersal and environmental filtering using the information collected in the first step.
3. **Model evaluation and refinement:** use EB and NB of the simulated and observed species to identify the combination of parameters to which the models are particularly sensitive. The set of simulations that maximizes the correspondence between the observed and the predicted EB-NB space are assembled into a composite model.
4. **Validation:** validate the composite model output by comparing the observed and the predicted patterns at different levels of the system, such as species rank distribution and the multivariate assemblage structure along both the spatial and environmental gradients.

5. **Implementation:** classify the observed species along the selection/non-selection continuum using the two-dimensional EB-NB space of the composite model.

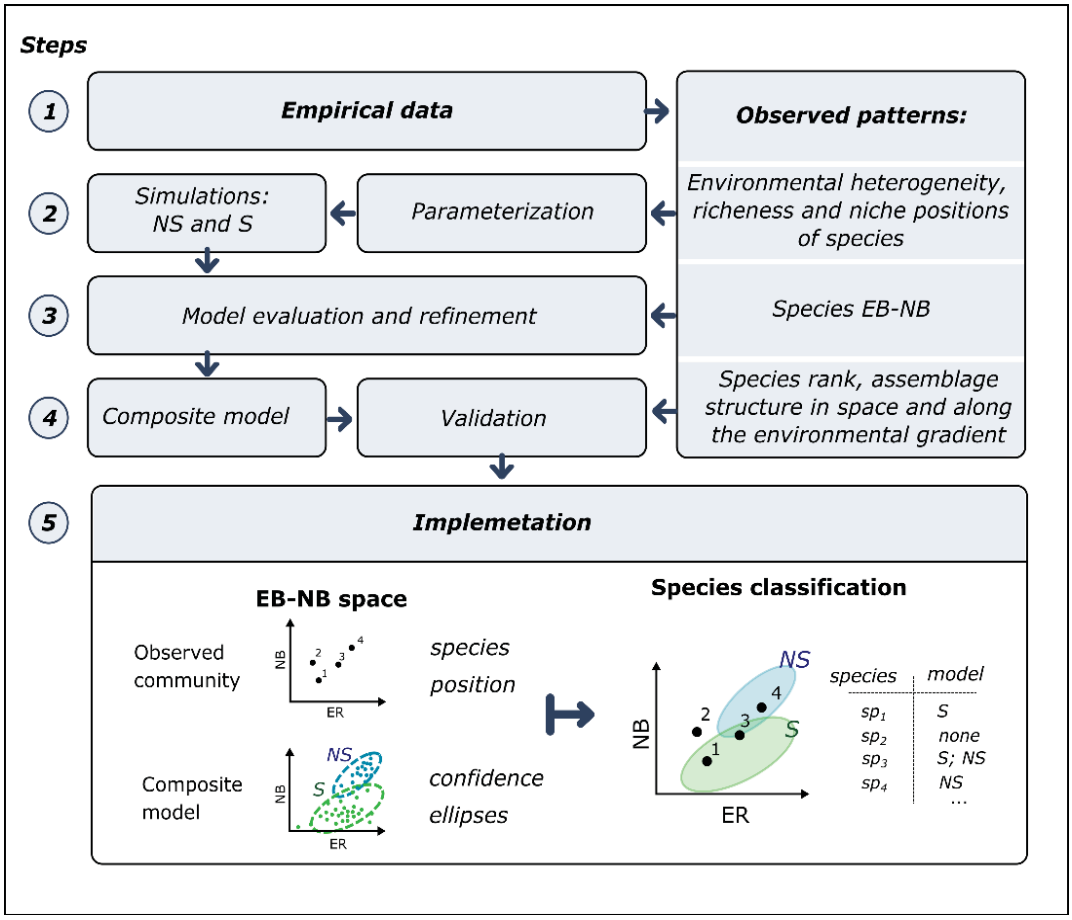


Fig. 2 – Workflow of the proposed analytical framework. (1) Empirical data, (2) simulations, (3) model evaluation and refinement, (4) validation and (5) implementation. NS- non-selection model, S- selection model.

3.4.1 The observed data

In step 1, the information from the empirical community and its environment were combined. As an empirical example, we used a nematode data set comprising an abundance of 194 species from Araçá Bay (southeastern coast of Brazil). This area exhibits a well-defined environmental gradient (Corte et al. 2017), which is crucial for the purposes of this study. The environmental data set was composed of 9 environmental variables: chlorophyll *a* ($mg.m^{-2}$), bathymetry (meters), percentage of total organic carbon, percentage of coarse sands (as the sum of pebbles, very coarse,

coarse, and medium grains), percentage of fine sand, percentage of very fine sand, mean grain size, and sorting coefficient. Because the analytical framework compares empirical communities with those generated by spatially explicit simulation models, it requires that both environmental and species data be structured in referenced grid sites. Details of the sampling and laboratory procedures are available in Appendix S1.

3.4.2 Simulations

In step 2, communities under different degrees of dispersal and environmental filtering were simulated using the parameters obtained in step 1. We performed simulations using the R v.3.3.2 software (R Development Core team 2008) and the "*neutral.vp"* package (Smith and Lundholm 2010). It consists of a spatially explicit, individual-based simulation model (Bell 2003, 2005) in which communities' dynamics are governed by key processes of birth, death, dispersal, and immigration. The spatial template is composed of M x M cells, all occupied by K individuals belonging to S species. Each cell starts with equal numbers of individuals from each species and carries capacity K. In each time step, individuals are born and die with probabilities *b* and *d*, respectively. Each marginal cell receives a single immigrant with probability *m*. With probability 1-*u,* each newborn individual randomly moves to one of the eight adjacent cells and continues moving until the probability of dispersal is not met. Any cell that exceeds K has individuals randomly removed until the local population is reduced to K. Environmental filtering is implemented by assigning each cell to an environmental value (EV) and each species to a specific optimum (SO). The closer EV and SO values are, the higher birth rates and lower death rates are. The strength of environmental filtering is inversely related to the parameter *sel* (the lower the *sel*, the stronger the filtering). The equations used to calculate *b* and *d* are described in Smith and Lundholm 2010.

In non-selection simulations, we settled both EV and SO values to 1 (ecological equivalence assumption). In selection simulations, we defined the set of EV and SO using both empirical environmental gradient and observed niche positions of species. For that, we first interpolated the measured environmental variables to a 60 x 60 size lattice (package "akima," Akima and Gebhardt 2016). Then, we ran a PCA and used the first PC scores as the EV units for all simulations (PCA results are available in

Appendix S2). According to the total number of observed species encountered in the bay, S was parameterized at 194. Lastly, we used the niche position of observed species (obtained from OMI analysis) as SO units for all simulations.

After preliminary trials, we restricted our simulations to the immigration rate of 0.005, the base birth rate of 0.505, and base death rate of 0.5. These values were taken from previous studies (Bell 2003, 2005, Smith and Lundholm 2010). We subsequently performed a total of 20 simulations varying selection strength (*sel*) and dispersal (*u*) (Table 1). K was settled using the maximum (3,435) abundances per site observed over the whole data set. We performed simulations without (non-selection simulations) and with the influence of the parameter *sel* (selection simulations). In selection simulations, *sel* varied from 100 to 900 and *u* from 0.001 to 0.4 (Appendix S2). Each simulation ran for 10 000 time steps.

3.4.3 Model evaluation and refinement

In step 3, the parameter space of simulations was reduced using measurements of environmental boundary and niche breadth. A redundancy analysis (RDA, was performed to evaluate the influence of parameters *sel* and *u* on the simulations. The final number of species (S_f) and the mean and variance of EBs and NBs were used as explanatory variables for the RDA.

A simulated two-dimensional EB-NB space comprising the scatters of EBs and NBs of all simulated species was generated and; the 99% confidence ellipses around EBs and NBs from each simulation were calculated ("*car*" R package, Fox and Weisberg 2011). Then, these confidence ellipses were superimposed with the EB-NB space from the observed species. The set of simulations that maximizes the correspondence between confidence ellipses and the observed EB-NB space was assembled into a composite model.

3.4.4 Validation

In step 4, additional parameters of the community were generated with the composite model. For comparison purposes, only the sites of the composite model which coordinates were closest to the empirical geo-locations were considered

(package “RANN”, Arya et al. 2017). This approach mimics the empirical sampling procedure. Three different, complementary assemblage patterns were examined: species-rank abundance distribution; the structure of the assemblages in space; and the structure of assemblages along the environmental gradient.

Rank abundance distributions (RADs) of log-transformed abundances were used to determine if the observed and simulated assemblages exhibit the same patterns of commonness and rarity. Abundances from both simulated and observed data were normalized prior to analysis to reduce the influence of dominant species. A total of 999 rank abundance distributions were generated from the simulated species by randomly sampling the observed number of species from the simulated species assemblage. Then, we calculated the 95% upper and lower confidence limits for each simulated specie rank, equivalent to the 97.5 and 2.5 percentiles, respectively. These confidence limits were compared to the RAD constructed for the observed data.

Mantel correlograms were used to compare the spatial structure of the simulated and observed assemblages (package “ecodist”, Goslee and Urban 2007). The Bray-Curtis dissimilarity matrix of each assemblage was modelled as a function of the geographic distance between samples, using 14 distance classes. The number of distance classes was calculated according to Sturge’ rule (Legendre and Legendre 1998). Ninety-nine percent confidence envelopes of the correlograms were estimated by 999 bootstrap resampling. The confidence limits of the simulated assemblage were defined using a resampling level of 15% (based on the proportion of observed to simulated number of species).

Lastly, we compared the multivariate structure of empirical and simulated assemblages along the environmental gradient. A distance-based redundancy analysis (dbRDA) was used to obtain a standard and comparable predictor value (“vegan” package, Oksanen et al. 2017), expressed as a proportion of the total variation in the response data (R^2). We performed dbRDAs on the empirical and simulated assemblages using their respective environmental tables as constrained explanatory variables. Statistical significance of the dbRDA models was tested using a permutational ANOVA-like test (function anova.cca; package). dbRDA models were compared using permutational tests of multivariate dispersion of dbRDA scores (“betadisper” function in “vegan” package). This method produces an independent dissimilarity value for each sample and has been proposed to express variation in community structure among groups (Anderson 2006).

3.4.5 Implementation

In the last step of this framework, the observed species were classified along the selection/non-selection continuum using the two-dimensional EB-NB space of the composite model. We further analyzed these data by dividing the empirical assemblage into three groups of log-normally distributed species: resident, semi-resident, and rare (Magurran and Henderson 2003). In order to identify these groups, we used piecewise regression following Gray et al. 2005. The analysis' script is available in the supplementary material S3.

3.5 RESULTS

3.5.1 EB-NB sensitivity

Based on the present data-set of marine nematodes, sensitivity analysis indicated that the measures of EB and NB were s robust (i.e. within the confidence interval of 95%) from 70 samples onwards (Fig. 3)

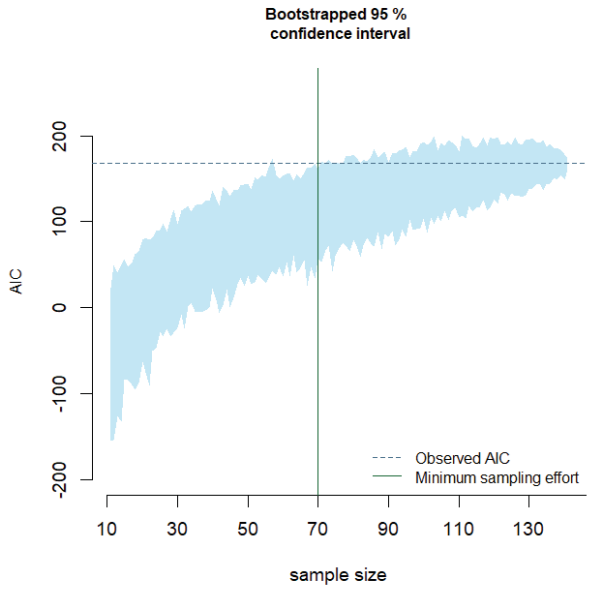


Fig. 3. Sensitivity analysis to variation of the sampling effort. The shaded area represents the bootstrapped 95 % confidence intervals of the AIC values.

3.5.2 Simulations, model evaluation, and refinement

The RDA analyses showed that both dispersal and selection were important parameters in determining the EBs and NBs patterns of simulated species (Fig. 4a). The parameter *sel* was important in distinguishing two “families” of simulations: simulations 1 to 4 (non-selection simulations) and 5 to 20 (selection simulations). Dispersal, on the other hand, explained differences within the families of simulations. Based on these results, we reduced the number of simulations to be included in the simulated two-dimensional EB-NB space. Among the four levels of selection simulations (sel = 100, 300, 600 and 900), we chose the level in which simulations maximized the correspondence between observed and simulated EB-NB vales (Appendix S2: Table S1).

A total of 8 simulations were selected to be included in the composite model (simulations 1 to 4 and 17 to 20; Fig. 4b). In all these simulations, dispersion influenced both the EBs’ mean and variance; the higher the dispersion rates, the larger and less variable the species’ EBs. The non-selection models (1 to 4) showed the highest values of NBs. Among the non-selection simulations, dispersion did not affect the overall NBs means, whereas the NB variance increased towards low dispersal models. On the other hand, the selection-based simulations (17 to 20) showed lower NB values. NB mean and variance decreased with decreasing dispersion among the selection-based simulations. All selection simulations showed some degree of overlap between each other. The selection simulations 19 and 20 were partially overlapped with non-selection models.

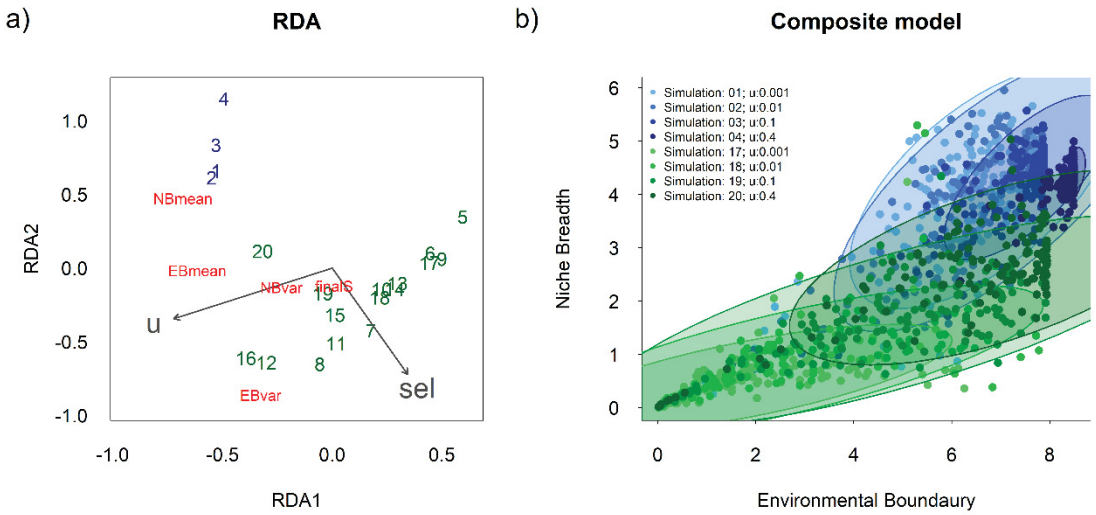


Fig. 4 – RDA and the composite model. (a) Ordination plots from the redundancy analysis of models. Blue and green numbers indicate the non-selection and selection models, respectively; simulation parameters are represented by vectors; vector length indicates the relative weight of a given variable, and direction indicates the correlation of that variable with each axis: sel- selection; u – dispersal; Response variables are given in red; FinalS- the final number of species; EB- environmental boundary, NB- niche breadth; var- variance (b) EB and NB space from the composite model. Each colored area on the graph represents the 99% confidence ellipse of one community.

3.5.3 Validation

The patterns of the field assemblage and that sampled from the composite model were concordant for all the parameters investigated. The rank-abundance distribution of nematodes followed a hollow shape, with small number of very abundant species and many rare species. The bootstrapped confidence limits of the simulated rank-abundance encompassed 177 ranks of the observed abundance distribution, with only low ranks (i.e., extremely rare species) being placed outside the confidence region (Fig. 5a).

Both multivariate analysis showed congruent assemblage patterns between model simulations and field observations. The Mantel correlogram of the nematodes showed significant decrease in similarity (or conversely, an increase in dissimilarity) with increasing geographic distance (Fig. 5b). All the distance classes exhibited significant spatial autocorrelation. At the largest spatial distance, the stations tended to zero, towards a no correlation. The correlogram obtained from the simulated assemblage displayed similar shape, and its confidence limits encompassed all the

empirical observations. The distance-based redundancy analyses calculated from empirical and simulated assemblages produced similar R^2 values, both significant at 0.1% level (Fig 5c, d; R^2_{obs}=0.19 and R^2_{sim}=0.26). Tests of multivariate dispersion of dbRDA scores showed no significant difference between observed and simulated RDA models ($F_{obs, sim}$=2.55, $p_{obs, sim}$=0.11).

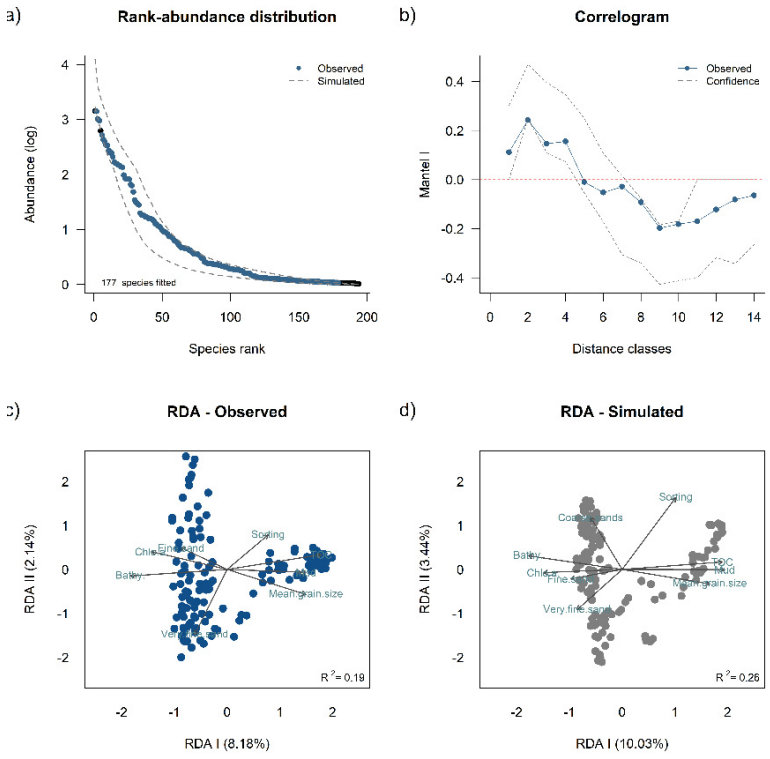


Fig. 5 – The three patterns evaluated from the observed (nematodes) and simulated (sampled from the composite model) assemblages. (a) The empirical rank-abundance distribution together with the 95% confidence interval (dashed lines) of the simulated distribution. (b) Mantel correlogram of the empirical assemblage (blue circles); positive values of Mantel I indicate positive autocorrelation of the assemblage composition for a given distance class; dashed lines indicate the 95% prediction interval of the simulated assemblage. (c) Redundancy-analysis of the observed and (d) simulated assemblages.

3.5.4 Implementation

About 96% of 194 nematode species analyzed were included within at least one ellipse (Fig. 5). Simulations 2 and 19 were the most inclusive ones, respectively fitting 34% and 79% of all empirical species (Table 1). Together, ellipses from the non-selection and selection simulations included 34% and 85% of the species, respectively.

A total of 23% percent of the species were placed in the overlapping zone between the two families of simulations.

About 97% of 194 nematode species analyzed were included within at least one ellipse (Fig. 6). Simulations 18 and 19 were the most inclusive ones, respectively fitting 56% and 77% of all empirical species (Table 1). Together, ellipses from the non-selection and selection simulations included 30% and 87% of the species, respectively. A total of 20% percent of the species were placed in the overlapping zone between the two families of simulations.

Both families of models included vacant, semi-resident, and resident species. The non-selection models accounted together for 4%, 16%, and 109% of the species classified respectively as rare, semi-resident, and resident. The selection models accounted together for 44% of rare, 29% of semi-resident, and 13% of resident species. All nematode species placed outside the ellipses were classified either as rare or intermediate.

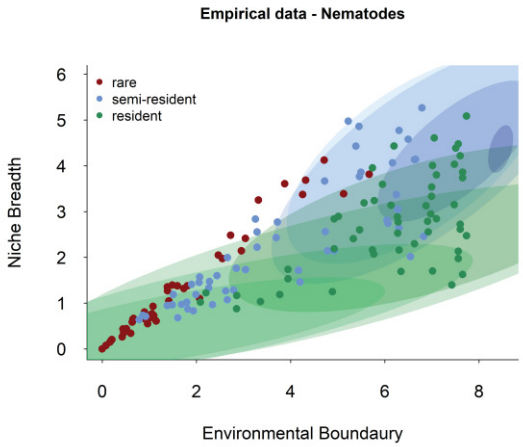


Fig. 6 – Empirical evaluation. Environmental boundary and niche breadth of all nematode species (classified by frequencies) observed at Araçá Bay and confidence ellipses produced by the composite model. The colors of ellipses are according to Fig. 4b.

Table 1. Summary of results for each model. S_{obs}: percentage of nematode species included in each 99% confidence ellipse.

	S_{obs}
Model 1	30%
Model 2	34%
Model 3	11%
Model 4	0%
Model 17	48%
Model 18	62%
Model 19	79%
Model 20	32%
Non-selection models (1,2,3,4)	34%
Selection models (17,18,19,20)	85%
Overlap between model families	23%
Outside species	4%

3.6 DISCUSSION

Our framework provides a complementary and alternative approach to identify the strength of environmental filtering and dispersal processes of empirical communities. Differently from previous approaches, we proposed a link between the empirical data and data from spatially explicit models to examine how the relative importance of these processes explains the distribution of each species in a community. Our methodology also includes a sensitivity analysis in order to infer the robustness of NB-EB based on sampling effort. This tool is particularly important when dealing with reduced data sets. The proposed approach was able to match three independent patterns of the community and therefore provide a good indication that the strengths of the underlying process can be inferred by comparing empirical and simulated patterns. It is well recognized that when a model can reproduce multiple patterns, each describing a different feature of the system, the risk that the model would be wrong is lower than when relying only on one pattern (Grimm 2005).

In this study, non-selection and selection models produced very distinct EB and NB patterns and differences within models were determined by changing the parameter that controls dispersal limitation. As expected from the theoretical background (Hirzel and Le Lay 2008), selection models show narrower EB and NB

values when compared to non-selection models, and both measures increase gradually as dispersal increases. The present study demonstrates that such inverse relationship is a result of dispersion counteracting the effects of the environmental filtering. Once environmental filtering is implemented, it has a weak influence on EB–NB patterns. It is also clear from the composite model that both non-selection and selection models have significant contributions in explaining the distribution of the empirical species. These findings support the idea that both environmental filtering and dispersal limitation are supplementary rather than mutually exclusive assembly processes (Vellend et al. 2014).

Our simulations show that dispersal plays a leading role in species' EB–NB patterns, creating a gradient within selection and non-selection models. Overall, the species richness decreases as the strength of dispersal limitation increases. This happens because high dispersion rates increase the chance of a species getting out of the simulated grid. Among the selection models, dispersion increases the overall mean and variance of both EB and NB. This occurs because low dispersion allows species to persist in locally favorable sites, forming environmentally defined patches, and therefore, showing narrow NB values (Heino 2005). In this scenario, range boundary populations are likely to be smaller, more variable, and less likely to receive immigrants from the main patch (Channell and Lomolino 2000). In contrast, at high dispersal rates, species can encounter a broader array of favorable conditions, expanding the NB values of each species. In the absence of environmental filtering, species are distributed solely by dispersion, filling the entire range of environmental conditions, which naturally increases both NB and EB. In this non-selection scenario, EB and NB converge to high and less variable values.

It is well recognized that the success of a simulation depends on how close the input parameters are from the empirical data (Jackson et al. 2000). The present framework allows the approximation between empirical and simulated communities by considering: (1) the total number of empirical species; (2) a lattice template that accounted PC1 as a proxy of the total variation in the habitat's heterogeneity and; (3) an empirically estimated environmental optima for each species along PC1 (niche position). Both environmental heterogeneity and species optima affect the establishment of species across landscapes by increasing or decreasing their probability of success, thus exerting an important effect on the location of species along the continuum between non-selection and selection processes (Bar-Massada et al.

2014). Future applications of our framework should attempt to empirically constrain additional parameters such as interspecific dispersal rates, species interactions and individual variation. As such, the implementation of these variables on the simulations could provide a more realistic picture. The present framework treats conspecific individuals as ecologically equivalent, which allows to locate the exact position of each species along the non-selection/selection continuum. The shortcoming is that species are considered as the average of its parts. Yet, developing a model and simulating a theoretical community representing hundreds of species with individual-level dynamics is at the very limit of the computational simulation methods used in this study (Codling and Dumbrell 2012).

The current framework suggested that 85% of nematode species have distributions constrained by selection processes. The importance of environmental factors on the distribution of marine nematode species is well documented (Gallucci et al. 2009, Vieira and Fonseca 2013, Fonseca et al. 2014) although very little attention has been given to determine the niche dimension of nematode species. The analytical framework also suggests that eleven percent of nematode species are exclusively distributed according to non-selection processes. Different from terrestrial environments, the sediment matrix in shallow coastal environments is very dynamic at the horizontal and vertical dimensions. Meiobenthic organisms are very susceptible to passive dispersion (Boeckner et al. 2009). These results suggest that, at the scale of the studied bay, the distribution of these species does not reflect the associated environmental conditions; however, it reflects the capability of species to inhabit contrasting environmental conditions at comparable densities.

Another critical issue is the scale of observation and the ecosystem considered. A species structured by non-selection processes at this scale could be niche-structured at larger scales. Similarly, it is very likely that a species structured by non-selection processes at this bay may show a selection-oriented pattern in another marine ecosystem. It is well accepted that selection processes are more easily detectable at coarser spatial scales, while non-selection dominates at finer scales (Chase 2014). It has been also suggested that when a sampling scale encompasses fewer individuals, the relative contribution of non-selection processes to the overall structure of the community increases (Chase 2014). However, the current framework shows that non-selection models include all frequency groups of species, with rare species displaying the highest deviation from the predicted by the composite model.

This is principally because of the high degree of noise in individual species patterns, causing difficulty in representing the among species relationships in a low-dimensional ordination (Clarke 1993). Because the likelihood of finding the real distribution of a species increases as we sample more individuals, misleading interpretation is less likely to occur with the resident species (Brown 1984). Accordingly, the simulated rank abundance distribution did not predict extremely rare species. In this regard, our multivariate analyses provided a more robust diagnostic of the composite model by analyzing the multivariate structure of the assemblages based on both geographical and environmental distances (Legendre and Legendre 1998).

Finally, we suggest a combined interpretation of the results provided by this framework and the output provided by the PCA and OMI analysis, particularly regarding the niche position of niche-structured species (Appendix S2: Table S1). Species may differ in their niche position along the PCA, indicating that they are correlated with a distinct set of environmental variables. As such, the conclusions will be restricted to the aspects measured in the study area. Once the underlying environmental gradient is properly quantified, the importance of species-specific niches for the community becomes more apparent. The inaccurate environmental assessment may cause unexplained variation in species' distributions and may provide misleading support for neutrality (Gilbert and Lechowicz 2004). Therefore, we recommend an accurate investigation of the most relevant environmental variables to have a precise interpretation.

In conclusion, our framework offers an alternative way to estimate the stochastic component of a community by considering the distributions of individual species across a multivariate space. Placing empirical niche measurements in the context of model-generated expectations enables a much deeper understanding of community assembly processes and how they vary from species to species.

3.7 ACKNOWLEDGMENTS

This work was supported by the São Paulo Research Foundation (FAPESP) under Project 2011/50317-5. We are grateful to Dr. Antonia Cecília Amaral Zacagnini who offered all the support needed for the development of this study. Special thanks are due to the laboratory of Marine Ecotoxicology and Microphytobenthos coordinated by Prof. Dr. Eduinetty Ceci Pereira Moreira de Sousa and to the laboratory of Marine

Microbiology coordinated by Prof. Dr. Ana Julia Fernandes Cardoso de Oliveira. We are also grateful to the Marine Biology Centre (CEBIMar) for making their facilities available during our stay at São Sebastião-SP. D.C.V is sponsored by the Coordination for the Improvement of Higher Education Personnel (CAPES).

3.8 APPENDIX S1 – SAMPLING AND LABORATORY PROCEDURES

Study area and sampling

The observational data set used here covers four sampling campaigns conducted at Araçá Bay, southeastern Brazil, over a year between 2012 and 2013 (October 2012, February 2013, May 2013, and September 2013). Samples were taken during the low tide from thirty-seven geo-referenced sites arranged on an irregular sampling grid (Fig. 2). At each sampling site, sediment samples were collected simultaneously for the investigation of nematode assemblages, microphytobenthos, and granulometry. For the nematode analysis, sediment was taken using a PVC corer 2.5 cm in diameter and 5 cm deep. Samples were immediately fixed in 4% formaldehyde. For the microphytobenthic analysis, five replicates of the top 1 cm sediment layer were taken using a corer 2 cm in diameter and stored in dark bottles. These samples were kept on ice and stored at -20 °C. Additional samples of sediment were taken for the granulometric and organic content analyses.

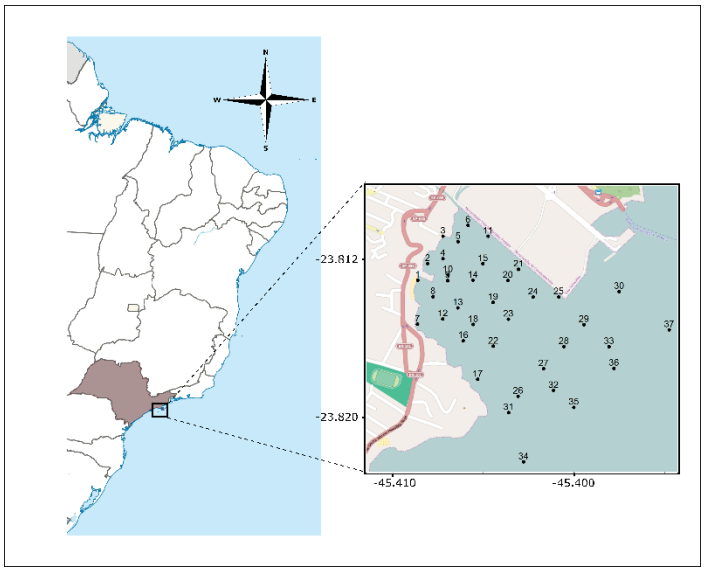


Fig. S1 Araçá Bay, Brazil.

Sample processing.

Nematode samples were washed through a 45-μm sieve and extracted by flotation with a solution of 1.18 g.cm^{-3} density colloidal silica (LUDOX TM-50). These

samples were stained with Rose Bengal; about 100 nematodes were randomly picked from each sample, identified at the genus level, and separated into morphospecies.

The microphytobenthic analysis was performed from the estimated chlorophyll-*a* and phaeopigments biomass. These pigments were extracted from each sample with 10 mL 100% acetone and 0.07 g of $MgCO_3$ for 24 h in the dark at 4 °C. Absorbance was read at 750, 665, and 430 nm in a spectrophotometer, before and after acidification with 1N HCl according to Plante-Cuny 1978. The calculations to obtain the chlorophyll-*a* and phaeopigments contents were performed using the Plante-Cuny equations (Plante-Cuny 1978).

The sediment granulometric analysis was carried out using the routine sieving and pipetting techniques described by Suguio (1973); sediment parameters were obtained using the SysGran software, version 3.0 (Camargo 2006) in accordance with the classifications of Folk and Ward (1957).

3.9 APPENDIX S2 – SUPPLEMENTARY RESULTS

The PCA analysis on the environmental parameters showed that PC1 and PC2 accounted for 45.67% and 22.9% of the total variation present, respectively. On the first principal component, the highest positive values were associated mainly with the percentages of mud content and total organic carbon whereas the highest negative values were due to chlorophyll a and bathymetry. In the second component, the percentage of very fine sand was associated with high positive values whereas the percentage of coarse sands was the main factor responsible for the most negative values.

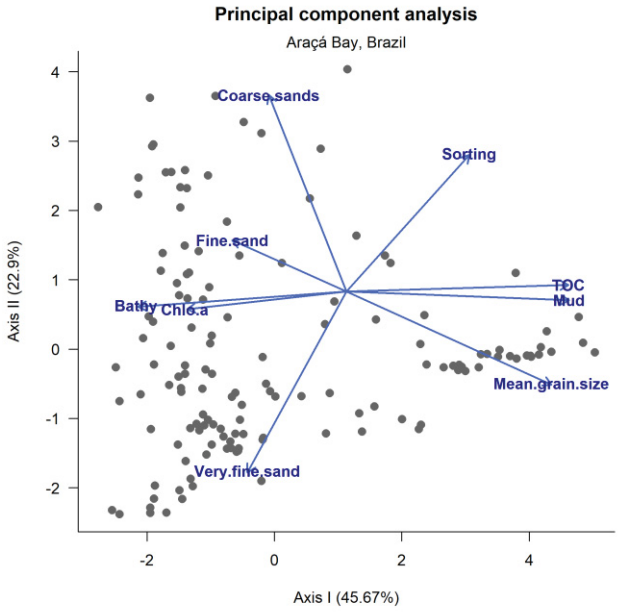


Fig. S2. Principal component analysis

Table S1. Simulations parameters and the exploratory variables used in the RDA analysis. Sel- selection (Inversely proportional to strength); u- dispersal; K- carrying capacity. Exploratory variables: Sf- the final number of species, EB- environmental boundary, NB- niche breadth; var- variance. Sobs: number of nematode species included in each 99% confidence ellipses. The simulations included into the composite model are shown in bold.

Simulation	sel	u	m	S_f	EBmean	NBmean	EBvar	NBvar	S_{obs}
1	0	0.001	0.0001	194	6.37	3.94	1.60	1.08	58
2	0	0.010	0.0001	194	6.54	3.87	1.71	1.16	65
3	0	0.100	0.0001	193	7.24	4.10	1.13	0.69	21
4	0	0.400	0.0001	192	8.38	4.30	0.09	0.10	0
5	100	0.001	0.0001	38	1.67	0.38	1.23	0.12	59
6	100	0.010	0.0001	27	2.44	0.51	2.80	0.17	62

Simulation	sel	u	m	S_f	EBmean	NBmean	EBvar	NBvar	S_{obs}
7	100	0.100	0.0001	43	3.04	0.83	5.83	0.46	121
8	100	0.400	0.0001	42	3.87	1.27	7.44	0.66	149
9	300	0.001	0.0001	62	1.93	0.43	2.76	0.17	57
10	300	0.010	0.0001	63	2.70	0.80	3.71	0.82	105
11	300	0.100	0.0001	65	3.76	1.15	6.39	0.64	130
12	300	0.400	0.0001	73	5.25	1.81	7.51	0.94	159
13	600	0.001	0.0001	103	2.57	0.74	3.41	0.61	93
14	600	0.010	0.0001	106	2.63	0.83	3.85	0.47	114
15	600	0.100	0.0001	110	4.28	1.37	5.43	0.49	137
16	600	0.400	0.0001	110	5.59	2.10	7.32	1.10	164
17	900	0.001	0.0001	140	2.13	0.63	2.63	0.27	93
18	900	0.010	0.0001	138	2.91	0.95	4.04	0.59	121
19	900	0.100	0.0001	143	4.76	1.61	4.60	0.56	153
20	900	0.400	0.0001	129	6.41	2.52	3.84	0.73	63

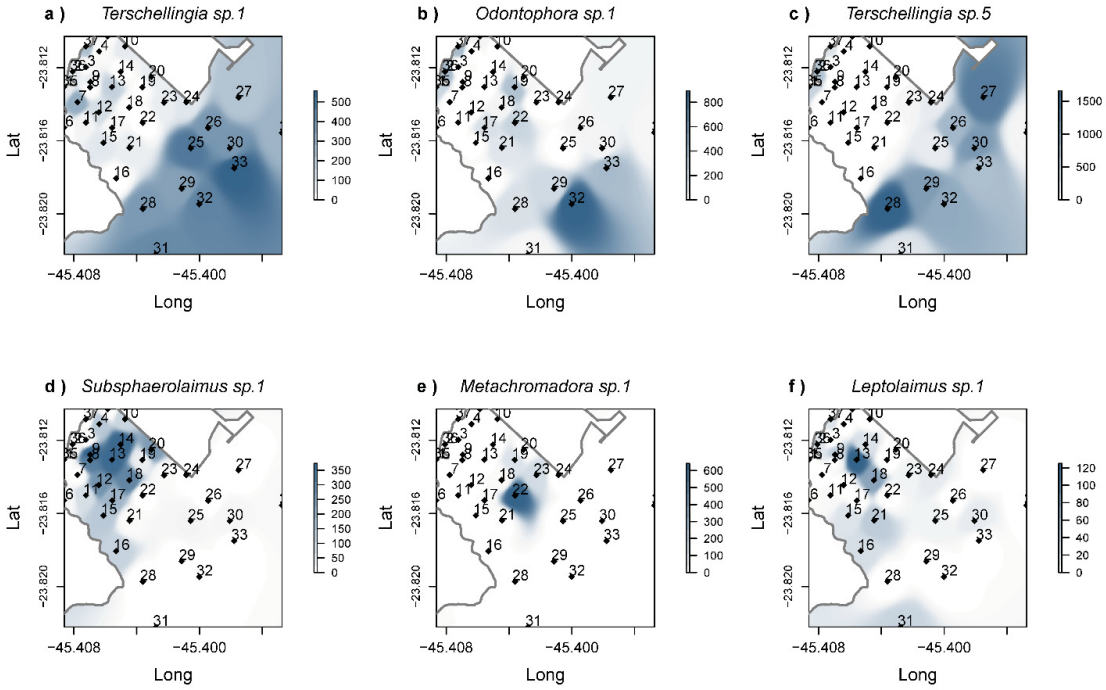


Fig. S3 –**Abundance maps**. Example of different types of distribution of nematode species classified along the models. (a, b) species from non-selection models including two levels of dispersal (u = 0.01 and 0.001, respectively); (c, d, e) species from selection models including three levels of dispersal (u = 0.1, 0.01 and 0.001, respectively); and (f) species placed outside the models.

Table S2. Niche Position (NP), environmental boundary (EB), and niche breadth (NB) of nematodes across the first PC axis; and the respective models in which species were included.

	Abund	Freq.group	NP	EB	NB	Model
Acanthonchus sp.1	654.11	semi-resident	-0.69	2.34	0.96	17, 18, 19
Actinonema sp.1	30.47	rare	0.80	4.32	3.69	1, 2

	Abund	Freq.group	NP	EB	NB	Model
Aegialoalaimus sp.1	7.75	rare	1.38	0.00	0.00	17, 18, 19
Anoplostoma sp.1	1183.26	resident	-0.87	6.31	2.16	2, 18, 19, 20
Aponema sp.1	3264.22	resident	2.96	7.74	2.47	19, 20
Araeolaimus sp.1	1.05	rare	-1.52	0.00	0.00	17, 18, 19
Ascolaimus sp.1	35.96	rare	3.08	2.96	2.14	19
Atrochromadora sp.1	6.08	rare	-1.31	1.81	1.38	18, 19
Axonolaimus sp.1	11.03	rare	-0.97	0.00	0.00	17, 18, 19
Bathylaimus sp.1	40.12	semi-resident	-1.31	0.78	0.64	17, 18, 19
Bolbolaimus sp.1	12.88	rare	-1.37	0.17	0.15	17, 18, 19
Calyptronema sp.1	138.31	semi-resident	-1.08	2.26	1.33	17, 18, 19
Campylaimus sp.1	160.72	semi-resident	3.02	6.05	2.81	1, 2, 19, 20
Catanema sp.1	50.15	rare	-0.68	0.00	0.00	17, 18, 19
Cephalanticoma sp.1	484.00	resident	-0.80	2.20	1.22	17, 18, 19
Cervonema sp.1	72.45	rare	-0.87	4.26	3.37	1, 2
Cheironchus sp.1	103.52	semi-resident	1.11	5.46	4.86	1, 2
Choanolaimus sp.1	359.56	resident	-0.53	5.47	2.60	1, 2, 19, 20
Chromadora sp.1	848.14	resident	-0.94	2.86	1.16	17, 18, 19
Chromadorella sp.1	66.77	semi-resident	1.12	1.39	0.95	17, 18, 19
Chromadoridae sp.1	17.64	rare	-1.23	0.00	0.00	17, 18, 19
Chromadorina sp.1	241.65	semi-resident	-0.87	3.29	2.22	19, 20
Cienfuegia sp.1	8.56	rare	-1.37	0.21	0.20	17, 18, 19
Cobbia sp.1	5451.65	resident	2.14	7.65	3.74	1, 2, 3, 20
Comesa sp.4	5.79	rare	3.76	0.46	0.35	17, 18, 19
Comesoma sp.1	5812.87	resident	-1.02	7.56	1.97	18, 19, 20
Comesoma sp.2	165.97	semi-resident	-1.95	1.82	0.87	17, 18, 19
Comesomatidae sp.1	11.55	rare	-0.78	1.38	1.35	18, 19
Cyartonema sp.1	1096.63	resident	1.55	7.56	4.48	1, 2, 3
Daptonema sp.5	3726.44	resident	-1.22	6.35	1.69	18, 19, 20
Daptonema sp.6	311.97	semi-resident	3.02	6.66	4.14	1, 2, 3
Daptonema sp.7	701.03	resident	0.48	5.74	3.96	1, 2
Dasynemella sp.2	3.86	rare	-1.04	0.00	0.00	17, 18, 19
Dasynemoides sp.1	145.24	semi-resident	1.40	6.31	4.77	1, 2
Deontolaimus sp.1	29.81	rare	3.31	1.49	1.39	18, 19
Desmodora sp.2	39.86	semi-resident	0.99	5.22	4.98	none
Desmolaimus sp.1	300.16	semi-resident	-1.07	4.17	1.72	2, 18, 19, 20
Desmoscolex sp.1	9.88	rare	4.16	0.00	0.00	17, 18, 19
Dichromadora sp.1	19.27	rare	-2.02	1.42	1.05	17, 18, 19
Diplolaimella sp.1	4.25	rare	-0.21	1.60	1.38	18, 19
Diplolaimelloides sp.1	20.87	rare	2.71	1.73	1.32	18, 19
Diplopeltula sp.1	56.46	semi-resident	4.43	1.51	1.18	17, 18, 19
Dolichosomatum sp.1	6.00	rare	1.79	0.00	0.00	17, 18, 19
Doliolaimus sp.1	37.14	rare	-1.40	0.19	0.18	17, 18, 19
Dorylaimidae sp.1	49.16	rare	1.16	1.09	0.68	17, 18, 19
Dorylaimopsis sp.1	6163.87	resident	-0.32	7.10	2.84	1, 2, 19, 20
Eleutherolaimus sp.1	9.91	rare	3.43	0.00	0.00	17, 18, 19

	Abund	Freq.group	NP	EB	NB	Model
Elzalia sp.1	351.18	semi-resident	2.52	5.56	3.19	1, 2, 20
Elzalia sp.2	1021.22	resident	2.62	4.93	2.84	1, 2, 19, 20
Epacanthion sp.1	10.07	rare	1.47	0.00	0.00	17, 18, 19
Eubostrichus sp.1	29.86	rare	1.14	3.88	3.61	none
Eurystomina sp.3	29.20	rare	-1.00	1.14	0.61	17, 18, 19
Gomphionema sp.1	2847.00	resident	-0.72	4.89	1.25	17, 18, 19, 20
Graphonema sp.1	7.31	rare	1.28	0.00	0.00	17, 18, 19
Halalaimus sp.1	1240.06	resident	-0.73	7.01	4.01	1, 2, 3, 20
Halichoanolaimus sp.1	450.90	resident	-0.90	2.09	1.02	17, 18, 19
Hypodontolaimus sp.1	335.19	semi-resident	-2.03	1.49	0.96	17, 18, 19
Kraspedonema sp.1	5.22	rare	-1.29	0.00	0.00	17, 18, 19
Laimella sp.1	84.45	semi-resident	1.38	6.50	4.58	1, 2
Leptogastrella sp.1	1.05	rare	-1.41	0.00	0.00	17, 18, 19
Leptolaimoides sp.1	103.94	semi-resident	-0.69	3.05	1.73	18, 19, 20
Leptolaimus sp.1	673.42	resident	0.11	5.78	3.24	1, 2, 20
Linhomoeus sp.1	13.01	rare	-1.12	1.39	1.27	18, 19
Linhystera sp.1	41.83	semi-resident	3.21	2.66	1.99	19
Linhystera sp.2	105.54	semi-resident	2.35	5.49	3.86	1, 2
Longicyatholaimus sp.1	9.00	rare	-0.85	0.00	0.00	17, 18, 19
Marisalbinema sp.1	58.00	rare	2.56	0.64	0.58	17, 18, 19
Marylynnia sp.1	130.34	semi-resident	-0.26	6.25	3.37	1, 2, 3, 20
Metachromadora sp.1	1260.56	semi-resident	-1.94	1.93	0.83	17, 18, 19
Metachromadora sp.4	20.63	rare	-1.45	0.08	0.07	17, 18, 19
Metacyatholaimus sp.1	1121.82	resident	3.05	4.93	2.18	1, 2, 18, 19, 20
Metadasynemoides sp.1	153.01	semi-resident	-1.57	1.80	1.02	17, 18, 19
Metalinhomoeus sp.1	1213.13	resident	-0.38	6.99	3.33	1, 2, 3, 19, 20
Metalinhomoeus sp.2	844.77	resident	0.58	5.95	3.59	1, 2, 3, 20
Metalinhomoeus sp.8	26.86	rare	4.81	0.00	0.00	17, 18, 19
Metaparoncholaimus sp.1	4.78	rare	-2.08	0.00	0.00	17, 18, 19
Metoncholaimus sp.1	576.94	semi-resident	-1.13	6.83	2.46	2, 19, 20
Microlaimidae sp.1	4.74	rare	3.68	0.00	0.00	17, 18, 19
Microlaimus sp.1	49.87	semi-resident	-0.35	3.25	2.84	none
Microlaimus sp.4	227.22	resident	0.71	6.20	4.43	1, 2
Microlaimus sp.6	147.13	semi-resident	-1.80	0.92	0.71	17, 18, 19
Microlaimus sp.8	500.77	semi-resident	-1.61	1.38	0.95	17, 18, 19
Molgolaimus sp.1	913.06	resident	2.50	7.60	2.61	19, 20
Monhystrella sp.1	63.07	semi-resident	-0.89	3.69	2.43	2, 19, 20
Morfotipo sp.1	13.07	rare	2.36	0.00	0.00	17, 18, 19
Morfotipo sp.2	8.93	rare	-1.06	0.00	0.00	17, 18, 19
Morfotipo sp.3	10.96	rare	1.28	0.00	0.00	17, 18, 19
Morfotipo sp.4	60.00	semi-resident	-1.59	1.61	0.68	17, 18, 19
Morfotipo sp.5	17.86	rare	-1.06	0.00	0.00	17, 18, 19
Nannolaimus sp.1	36.15	rare	2.08	0.67	0.66	17, 18, 19
Nudora sp.1	12.36	rare	-1.97	0.61	0.34	17, 18, 19
Odontanticoma sp.1	28.13	rare	-0.66	0.99	0.67	17, 18, 19

	Abund	Freq.group	NP	EB	NB	Model
Odontophora sp.1	6135.22	resident	2.14	7.74	5.09	1, 2, 3
Omicronema sp.1	7.35	rare	-1.41	0.00	0.00	17, 18, 19
Oncholaimellus sp.1	1676.12	resident	-1.10	7.56	2.13	18, 19, 20
Oncholaimellus sp.2	398.22	semi-resident	-1.65	6.54	2.01	18, 19, 20
Oncholaimus sp.1	1120.92	resident	-1.18	3.95	1.53	18, 19, 20
Oxystomina sp.1	351.04	resident	-0.95	3.95	1.73	2, 18, 19, 20
Paracanthonchus sp.1	2502.80	resident	-1.32	5.77	2.07	1, 2, 18, 19, 20
Paracanthonchus sp.2	12.33	rare	-1.13	0.44	0.43	17, 18, 19
Parachromadorita sp.1	25.12	rare	-1.41	1.08	0.93	17, 18, 19
Paracyatholaimoides sp.1	233.85	semi-resident	-1.94	4.93	2.15	1, 2, 18, 19, 20
Paralinhomoeus sp.1	210.45	semi-resident	-0.55	5.46	3.76	1, 2
Paralongicyatholaimus sp.1	484.88	semi-resident	-1.33	0.86	0.73	17, 18, 19
Paramesonchium sp.1	118.21	semi-resident	-1.14	2.44	1.60	18, 19
Parametoncholaimus sp.1	77.58	rare	-1.26	0.00	0.00	17, 18, 19
Paramicrolaimus sp.1	6.61	rare	0.84	0.00	0.00	17, 18, 19
Paramonhystera sp.1	4422.23	resident	-0.72	7.51	4.39	1, 2, 3
Paramonhystera sp.2	3930.98	resident	2.88	7.48	3.15	1, 2, 19, 20
Paramonhystera sp.3	1066.55	semi-resident	-1.85	4.20	1.46	2, 17, 18, 19, 20
Parasphaerolaimus sp.1	16.49	rare	2.83	3.32	3.25	none
Parasphaerolaimus sp.2	80.56	rare	1.65	5.13	3.39	1, 2, 20
Parodontophora sp.1	393.83	resident	2.71	6.88	3.13	1, 2, 3, 19, 20
Phanoderma sp.2	23.84	rare	-0.53	2.47	2.05	19
Polygastrophora sp.1	251.69	semi-resident	-0.93	2.79	1.28	17, 18, 19, 20
Polygastrophora sp.2	150.04	semi-resident	-1.28	1.99	1.19	17, 18, 19
Polysigma sp.1	8.00	rare	-0.72	0.00	0.00	17, 18, 19
Pomponema sp.1	343.40	semi-resident	-1.76	2.62	1.27	17, 18, 19
Pomponema sp.2	161.24	rare	-2.54	2.07	1.10	17, 18, 19
Prochromadora sp.1	15.33	rare	-0.97	0.00	0.00	17, 18, 19
Prochromadorella sp.1	47.02	semi-resident	-0.09	3.29	2.55	none
Promonhystera sp.1	13.07	rare	2.36	0.00	0.00	17, 18, 19
Pseudochromadora sp.1	11.88	rare	2.35	0.00	0.00	17, 18, 19
Pseudosteineria sp.1	1388.64	resident	-2.11	5.74	2.16	1, 2, 18, 19, 20
Ptycholaimellus sp.1	2607.66	resident	-0.31	6.29	2.77	1, 2, 19, 20
Ptycholaimellus sp.2	193.23	semi-resident	-1.43	2.06	1.45	18, 19
Rhabditidae sp.2	7.31	rare	1.28	0.00	0.00	17, 18, 19
Rhabditidae sp.3	6.61	rare	-0.57	0.90	0.66	17, 18, 19
Rhabditidae sp.4	8.37	rare	1.14	1.09	0.73	17, 18, 19
Rhabditidae sp.5	1.04	rare	-1.08	0.00	0.00	17, 18, 19
Rhabditidae sp.6	1.04	rare	-1.08	0.00	0.00	17, 18, 19
Rhabditidae sp.7	3.84	rare	-0.47	0.90	0.80	17, 18, 19
Rhabditidae sp.8	10.19	rare	-0.20	0.43	0.26	17, 18, 19
Rhabditidae sp.9	2.77	rare	-0.72	0.00	0.00	17, 18, 19
Rhabdocoma sp.1	50.15	rare	-0.68	0.00	0.00	17, 18, 19
Sabatieria sp.1	5101.63	resident	3.09	6.63	2.30	2, 19, 20
Sabatieria sp.1*a*	2484.39	resident	-1.54	5.02	2.89	1, 2, 19, 20

	Abund	Freq.group	NP	EB	NB	Model
Sabatieria sp.2	3455.44	resident	-0.82	7.65	1.62	18, 19, 20
Sabatieria sp.2*c*	305.46	semi-resident	2.44	6.28	3.05	1, 2, 3, 19, 20
Sabatieria sp.4	795.92	semi-resident	-0.38	6.79	5.26	1, 2
Sabatieria sp.5	1752.16	resident	2.85	6.28	2.53	1, 2, 19, 20
Sabatieria sp.6	63.69	semi-resident	2.63	5.39	4.43	1, 2
Sabatieria sp.7	66.83	rare	3.08	3.04	2.41	none
Sabatieria sp.8	315.09	semi-resident	1.74	6.30	2.65	1, 2, 19, 20
Scaptrella sp.1	9.89	rare	-2.06	0.00	0.00	17, 18, 19
Setosabatieria sp.1	1988.89	resident	1.29	7.10	3.80	1, 2, 3, 20
Siphonolaimus sp.1	37.12	rare	0.78	5.67	3.81	1, 2
Southerniella sp.1	13.02	rare	-1.27	2.73	2.48	none
Sphaerolaimus sp.1	586.55	resident	1.12	7.05	4.61	1, 2, 3
Sphaerolaimus sp.2	362.24	resident	2.27	5.33	2.41	1, 2, 19, 20
Sphaerolaimus sp.3	174.19	semi-resident	2.02	6.63	4.13	1, 2, 3
Sphaerolaimus sp.4	14.88	rare	0.00	0.08	0.07	17, 18, 19
Sphaerotheristus sp.1	515.17	resident	2.97	6.13	3.14	1, 2, 3, 19, 20
Spilophorella sp.1	367.59	semi-resident	-1.10	2.66	1.07	17, 18, 19
Squanema sp.1	2304.74	resident	-2.09	3.36	1.03	17, 18, 19, 20
Steineria sp.1	455.15	resident	1.31	7.48	4.03	1, 2, 3, 20
Steineria sp.2	36.02	semi-resident	0.14	3.72	2.77	20
Stygodesmodora sp.1	197.87	semi-resident	2.05	4.74	2.56	1, 2, 19, 20
Stylotheristus sp.1	515.50	semi-resident	-0.80	6.07	2.74	1, 2, 19, 20
Stylotheristus sp.3	4688.59	resident	3.14	6.92	2.55	1, 2, 19, 20
Stylotheristus sp.4	515.14	resident	-0.52	6.27	2.88	1, 2, 3, 19, 20
Subsphaerolaimus sp.1	4183.21	resident	-0.90	3.77	1.18	17, 18, 19, 20
Synonchiella sp.1	52.10	rare	-1.94	0.51	0.44	17, 18, 19
Synonchium sp.1	8.41	rare	-1.41	0.00	0.00	17, 18, 19
Syringolaimus sp.2	3.65	rare	1.28	0.00	0.00	17, 18, 19
Terschellingia sp.1	6724.84	resident	1.67	7.65	3.86	1, 2, 3, 20
Terschellingia sp.2	3025.88	semi-resident	1.40	4.79	2.14	1, 2, 18, 19, 20
Terschellingia sp.3	2394.14	resident	2.19	7.00	3.54	1, 2, 3, 20
Terschellingia sp.5	13827.75	resident	2.70	7.60	2.72	19, 20
Terschellingia sp.6	203.64	semi-resident	-0.35	4.72	3.67	1, 2
Thalassoalaimus sp.1	150.53	semi-resident	0.00	2.85	1.76	18, 19, 20
Thalassomonhystera sp.1	146.04	semi-resident	-0.77	1.90	1.41	18, 19
Theristus sp.10	1.45	rare	-1.67	0.00	0.00	17, 18, 19
Theristus sp.10C	20.35	rare	-1.67	0.00	0.00	17, 18, 19
Theristus sp.8	103.38	semi-resident	-1.46	2.28	1.47	18, 19
Theristus acribus	391.09	resident	-1.54	2.85	0.87	17, 18, 19
Theristus sp.F2	13.38	rare	-1.50	0.00	0.00	17, 18, 19
Theristus flevensis	75.16	semi-resident	-1.52	1.69	0.97	17, 18, 19
Theristus longissimecaudatus	2076.44	resident	-0.93	6.91	2.79	1, 2, 19, 20
Trefusia sp.1	307.69	semi-resident	-1.31	2.07	1.57	18, 19
Trefusia sp.2	110.08	rare	-0.77	0.96	0.55	17, 18, 19
Tricoma sp.1	11.84	rare	-0.13	0.00	0.00	17, 18, 19

	Abund	Freq.group	NP	EB	NB	Model
Tricotheristus sp.1	2475.12	resident	0.12	7.60	4.22	1, 2, 3, 20
Trileptium sp.1	3.65	rare	1.28	0.00	0.00	17, 18, 19
Tripyloides sp.1	34.07	rare	1.12	4.71	4.12	1, 2
Trissonchulus sp.1	3043.30	resident	-1.12	7.42	1.39	19, 20
Trochamus sp.1	64.07	semi-resident	1.90	6.17	4.07	1, 2
Valvaelaimus sp.1	27.64	rare	1.98	0.65	0.64	17, 18, 19
Vasostoma sp.1	2063.27	resident	2.23	6.98	2.95	1, 2, 3, 19, 20
Viscosia sp.1	3341.28	resident	-0.98	7.01	1.70	18, 19, 20
Wieseria sp.1	39.19	rare	-0.72	1.06	0.77	17, 18, 19
Xyalidae sp.1	45.39	rare	2.25	2.55	1.97	19
Xyalidae sp.2	20.39	rare	2.37	0.00	0.00	17, 18, 19
Xyalidae sp.3	20.03	rare	-1.39	0.20	0.20	17, 18, 19

3.10 APPENDIX S3 VIGNETTE

`**require**(vegan)

```
require(segmented)
require(car)
require(sp)
require(ade4)
require(neutral.vp) ## Smith & Lundholm, 2010
require(RANN)
```

Loading the main functions

EBNB

The Function EBNB calculates EB and NB on abundance and variables matrices. It was based on the OMI analysis (Dolédec, 2000) and uses the functions “niche” and dudi.pca {ade4}.

Arguments: - abund: abundance table (samples as rows x species as columns) - envi: environmental table (samples as rows x variable as columns) - PC: Principal Component which analysis will be performed (1,2)

```
EBNB<-function(abund, envi, PC=1)
{
  require(ade4)
  EBNB.table<-data.frame(matrix(NA, ncol(abund),3 ))
  colnames(EBNB.table)<-c("EB","NB","NP")
  pca<-dudi.pca(envi, scale=T, scannf=F)
  omi<-niche(pca,abund, scannf=FALSE)
  omi_scores<-omi$ls[,PC]
  df <- sweep(abund, 2, apply(abund, 2, sum), "/")
  pa<-abund
  pa[pa>0]<-1
  for(i in 1:ncol(abund))
  {
    NP <- sum(df[, i] * omi_scores)
```

```
    sd <- sqrt(sum(df[, i] * (omi_scores - NP)^2))
    NB<-abs(diff(c((NP+sd),(NP-sd) )))
    EB<-abs(diff(range(omi_scores[pa[,i]>0])))
    EBNB.table[i,1:3]<-c(EB,NB,NP)}
  rownames(EBNB.table)<-colnames(abund)
  return(EBNB.table)
}
```

expand_grid

The function expand_grid interpolates the environmental table to a MxM grid size using the function “interp” {akima} The function returns either the interpolated environmental table (coord.return=F) or the interpolated coordinates (coord.return=T). In both tables, samples are rows (number of rows=M*M) and variables are columns.

Arguments: - envi: environmental table (samples as rows x variable as columns) - coords: A two-column dataframe containing the sample coordinates - M: The desired number of rows and columns in the grid

```
expand_grid<-function(envi, coords, M, coord.return=F) {
  require(akima)
  new.data<-data.frame(matrix(NA,M*M, ncol(envi)))
  colnames(new.data)<-colnames(envi)
  for (ij in 1:ncol(envi))
  {    var<-envi[,ij]
  s=interp(coords[,1],coords[,2],var,extrap=T,linear=F,duplicate="mean",nx = M,
ny = M)
  var1<-as.vector(s$z) ### serÃ¡ preenchido no setido cima.Baixo, (por coluna)
  new.data[,ij]<-var1
  if(coord.return==T)
  {return(expand.grid(long.dec=s$x,lat.dec=s$y))}}
  return(new.data)
}
```

simu_loop

The function simu_loop generates multiple simulations using non-selection and selection processes. The function was based on the function neut.simulate {neutral.vp} (Smith & Lundholm, 2010) and returns a list of objects of class (sim) {neutral.vp}. The final number of objects will be equal to: length(sel.vector) * length(u.vector) *length(u.vector)

Arguments: - S: Number of species - M: Number of rows and columns in the grid - m: Immigration rate - b: Base death rate - d: Base birth rate - K: Carrying capacity for each grid cell - sel.vector: A vector containing the Selection strengths - u.vector: A vector containing the migration rates - envi_simu: Environmental table created from expand_grid function - cycles: Number of cycles for each simulation. Each cycle is equivalent to 500 timesteps

```
simu_loop<-function(S,M,m,b,K, d,  sel.vector, u.vector, envi_simu,  cycles, SO
=NULL)

{options(scipen=999)

  EV<-matrix(dudi.pca(envi_simu, scale=T, nf=10, scannf=F)$li[,1], M,M)

  if(is.null(SO))

  { SO<-seq(min(EV),max(EV), length.out=S)}

  require(neutral.vp)

  for(i in 1:length(u.vector))

  {

    cat(paste("Simulation: ",i,"/",length(u.vector)*length(sel.vector)+ length(
u.vector)

              , sep=""))

    cat(paste("\n Non-selecton simulation; u: ",u.vector[i],"\n"))

    assign(paste('NS.simu_u',u.vector[i], sep=""),  neut.simulate(M = M, K = K,
S = S, cycles=cycles,time=500,  u=u.vector[i], m=m, b=b,d=d))

  }

  n.simu<-length(u.vector)+1

  for(j in 1:length(sel.vector))

  {  for(i in 1:length(u.vector))

  { cat(paste("Simulation: ",n.simu,"/",length(u.vector)*length(sel.vector)+len
gth(u.vector)

              , sep=""))

    cat(paste("\n Selection simulation; sel:",sel.vector[j]),";","disp:",u.vect
or[i], "\n")

    assign(paste('S.simu_sel', sel.vector[j],"_u" ,u.vector[i], sep=""),
```

```
            neut.simulate(M=M, K=K, S=S, cycles=cycles, time=500,  u=u.vector[i]
, fitness=SO,
                          habitat=EV, sel=sel.vector[j], m=m, b=b,d=d))
    n.simu<-n.simu+1
  }}
  rm(b,d,M, K, S,  u.vector, m, EV, SO,  sel.vector, cycles, i, j, n.simu)
  final_list<-as.list(environment(), all=T)
  cat("\n ------------- End of simulations -----------")
  cat("\n Non-Selection simulations : NS.simu")
  cat("\n Selection simulations : S.simu")
  return(rev(final_list))
}
```

Build_CompModel

The function Build_CompModel performs the EBNB analysis on the simulation outputs (list of sim objects) using the environmental table created from the "expand_grid function". From each sim object, the function extracts the final species abundance table and performs the EBNB analysis. The function returns a EBNB_summary object (class EBNB_summary) which comprises: (1) A dataframe containing a summary table (see Appendix S2:TableS1), (2) A list of dataframes containing the species EBNB of each simulation. (3) A list of dataframes containing the final species abundance from each simulation.

Arguments - sim.list: list of "sim" objects created from simu_loop function - envi_simu: Environmental table created from expand_grid function - PC: Principal Component which analysis will be performed (1,2)

```
Build_CompModel<-function(sim.list, envi_simu, PC=1) ##
{
  n.models<-1:(length(sim.list))
  SUMMARY<-data.frame(matrix(NA,length(sim.list),10,
                             dimnames=list(1:length(sim.list),  c("S",'sel','u',
"m","K","finalS","EBmean","NBmean","EBvar","NBvar"))))
  ABUND_list <- sapply(paste("abund_","model", 1:length(sim.list), sep=""),func
tion(x) NULL)
  EBNB_list<-sapply(paste("EBNB_","model", 1:length(sim.list), sep=""),function
(x) NULL)
```

```
    finalS<-function (sim){species<- census(sim) ## remove any extincts species
from the model abundance tables
  S.sum<-apply(species, 2, sum)
  if(any(S.sum==0)){species<-data.frame(species[,-which(S.sum==0)])}
  return(species)}
  for(i in 1:length(sim.list))
  {
    ABUND_list[[i]]<-data.frame(finalS(sim.list[[i]]))
    ABUND_list[[i]]<-  ABUND_list[[i]][,apply(decostand(ABUND_list[[i]],"pa"),2,
sum)>1]
    EBNB_list[[i]]<-data.frame(EBNB(ABUND_list[[i]], envi_simu))
    SUMMARY[i,1]<-sim.list[[i]][[1]]$S
    SUMMARY[i,2]<-sim.list[[i]][[1]]$sel
    SUMMARY[i,3]<-sim.list[[i]][[1]]$u
    SUMMARY[i,4]<-sim.list[[i]][[1]]$m
    SUMMARY[i,5]<-sim.list[[i]][[1]]$K
    SUMMARY[i,6]<-ncol(ABUND_list[[i]])
    SUMMARY[i,7:8]<-apply(EBNB_list[[i]], 2, mean)[1:2]
    SUMMARY[i,9:10]<-apply(EBNB_list[[i]], 2, var)[1:2]
    message(paste("\n loop", i,"-",length(sim.list)))  }
  SUMMARY$sel[SUMMARY$sel==999]<-0
  order_models<-function(x){x[ order(x[,2], x[,3], decreasing=F), ]} ## functio
n to order by selection and dispersal
  SUMMARY<-order_models(SUMMARY)
  ord.models<-as.numeric(rownames(SUMMARY))
  ABUND_list<-ABUND_list[ord.models]
  EBNB_list<-EBNB_list[ord.models]
  rownames(SUMMARY)<-1:nrow(SUMMARY)
  names(ABUND_list)<-1:nrow(SUMMARY)
  names(EBNB_list)<-1:nrow(SUMMARY)
  FINAL_list<-list(SUMMARY=SUMMARY,EBNB_list=EBNB_list, ABUND_list=ABUND_list)
  cat("Number of models:", length(sim.list), "\n")
  cat("\n Available objects")
  cat("\n One data frame:", names(FINAL_list)[1])
  cat("\n Two lists:      ")
  cat(names(FINAL_list)[2:3], sep="; ")
  cat ("\n Attention: new names assigned to the models: Numeric, sorted by sele
ction and dispersal.")
  class(FINAL_list)<-'CompModel'
```

```
  return(FINAL_list)}
 ## Dissimilarity profile table
```

EBNB_space

The function EBNB_space displays the two-dimensional array of EB and NB model using the ‘CompModel’ object. The function returns a list with the confidence ellipses calculated from the EBNB space of each simulation.

Arguments: - CompModel: CompModel object - chosen.simu: Simulations to be plotted (use the simulation numbers) - conf.level: draw elliptical contours at this confidence level(function dataEllipse {car}) - legend: display the legend of the plot

```
EBNB_space<-function(CompModel, chosen.simu=1:nrow(CompModel[[1]]), conf.level=
0.99, legend=F) ## plot the composite model
{
  require("car")
  SUMMARY<-CompModel$SUMMARY[chosen.simu,]
  NS.simu.colors<-colorRampPalette(c("darkblue", "steelblue1"))(length(levels(a
s.factor(SUMMARY$u))))
  S.simu.colors<-colorRampPalette(c("darkgreen", "springgreen2"))(length(levels
(as.factor(SUMMARY$u))))
  color.vector<-SUMMARY$u
  model_colors<-c(rep(NS.simu.colors,table(color.vector[SUMMARY$sel==0]) ),
                  rep(S.simu.colors, length(levels(as.factor(SUMMARY$sel)))))
  SUMMARY<-CompModel[[1]][chosen.simu,]
  EBNB_list<-CompModel[[2]][chosen.simu]
  xrange<-range(do.call("rbind", EBNB_list)[,1])
  yrange<-range(do.call("rbind", EBNB_list)[,2])
  par(mar=c(6,6,5,4), cex=1.4, xpd=F)
  plot(EBNB_list[[1]][,1],EBNB_list[[1]][,2], xlab="Environmental Boundaury",
       ylab="Niche Breadth", type="n", main="Simulated data", bty="l",
       xlim=xrange, ylim=yrange, las=1, tck=-.01, lwd=2)
  order_models<-function(x){x[ order(x[,2], x[,3], decreasing=F), ]}
  SUMMARY<-order_models(SUMMARY)
  conf<-list()
  for(i in length(EBNB_list):1)
  { tryCatch({
```

```
    conf[[i]]<-dataEllipse(data.frame(EBNB_list[i])[,1],  data.frame(EBNB_list[i
])[,2], col=model_colors[i], levels=c( conf.level), add=T,

                           plot.points=F, robust=T, lwd=1.8, fill=T, lty=1 , fi
ll.alpha=.2, center.pch=F, fig=c(2,3,4,5))

    points(data.frame(EBNB_list[i])[,1],data.frame(EBNB_list[i])[,2], pch=16, c
ol=model_colors[i], cex=1.2)

  }, error=function(e){})}

  u.parameters<-NULL

  sel.parameters<-NULL

  for(i in 1:length(EBNB_list))

  { sel.parameters[i]<-SUMMARY$sel[i]

  u.parameters[i]<-SUMMARY$u[i] }

  partial.legend<-paste("Model: ",  sprintf("%02d", as.numeric(rownames(SUMMARY
))),"  - sel:" ,sel.parameters, " - u:",u.parameters, sep="")

  if(legend==T)

  {legend("topleft", pch=c(16), col=model_colors, legend=c(partial.legend), bty
="n", cex=.8)}

   return(conf)

}
```

EBNB_ANALYSIS

The function EBNB_ANALYSIS performs the empirical evaluation using a given composite model (CompModel object)..The function returns a list of results which comprises: (1) full.table: The overhall results of the EBNB analysis (2) binatry table: indicating the species encopassed by each simulation (3) small table: the number of species encopassed by the ellipses from (a) non-selection and (b) selection simulations.

Arguments: - abund: Abundance table (samples as rows x species as columns) - envi: Environmental table (samples as rows x variable as columns) - PC: Principal Component which analysis will be performed (1,2) - CompModel: CompModel object - chosen.simu: Simulations chosen for the composite model (use the simulation numbers in CompModel$SUMMARY) - conf.level: Confidence level of the elliptical contours - freq.list: freq.list object - point.sp: The species to be indicated in the plot (use the column numbers of “abund” table) - title: title of the plot

```
EBNB_ANALYSIS<-function(abund,envi,PC=1,  CompModel,chosen.simu=1:nrow(CompModel
[[1]]),
```

```
                          conf.level=0.99, freq.list=NULL, point.sp=NULL, title="
Nematodes")

{

  require("car")

  require(sp)

  if(anyNA(chosen.simu)){ chosen.simu<-1:nrow(CompModel$SUMMARY)}

  EBNB_obs<-EBNB(abund, envi, PC)

  EBNB_list<-CompModel$EBNB_list[chosen.simu]

  xrange<-range(do.call("rbind", EBNB_list)[,1])

  yrange<-range(do.call("rbind", EBNB_list)[,2])

  full.table<-data.frame(matrix(NA,nrow(EBNB_obs), length(chosen.simu), dimname
s=list(rownames(EBNB_obs),paste("Sim",chosen.simu, sep="_"))))

  conf.points<-list()

  for(i in 1:length(chosen.simu))

  { conf<-dataEllipse(EBNB_list[[i]][,1], EBNB_list[[i]][,2],draw=F, level=(con
f.level), robust=T)

  conf.points[[i]]<-conf

  full.table[,i]<-point.in.polygon(EBNB_obs[,1], EBNB_obs[,2], conf[,1], conf[,
2])}

  binary<-full.table

  EBNB_class<-list()

  Abund_class<-list()

  for(i in 1:length(chosen.simu))

  {EBNB_class[[i]]<-EBNB_obs[full.table[,i]==1,]

  Abund_class[[i]]<-abund[,full.table[,i]==1] }

  Table.temp<-data.frame(matrix(NA,nrow(full.table), 1, dimnames=list(c(rowname
s(full.table)),c('Sim'))))

  for(i in 1:ncol(full.table)){full.table[which(full.table[,i]==1),i]<-gsub("Si
m_", "", colnames(full.table)[i])}

  for(i in 1:nrow(full.table)) { Table.temp[i,1]<-paste(full.table[i,which(full
.table[i,]>0)], collapse=", ")}

  Table.temp[which(Table.temp[,1]==""),1]<-"none"

  NPs<-EBNB_obs[3]

  Table.S2<-data.frame(Abund=round(apply(abund, 2, sum),2), Freq.group=NA,NP=NP
s[,1],EBNB_obs[,1:2],Table.temp )

  rownames(Table.S2)<-rownames(Table.temp)

  for(i in 1:length(freq.list))

  { Table.S2$Freq.group[which(rownames(Table.S2)%in%freq.list[[i]])]<-names(fre
q.list)[i]}

  Table.S2<-Table.S2[order(rownames(Table.S2), decreasing=F ),]

  Table.S2<-Table.S2[order(Table.S2$Abund, Table.S2$Freq.group,decreasing=T ),]
```

```
  NS<-which(CompModel$SUMMARY$sel[chosen.simu]==0)

  S<-which(CompModel$SUMMARY$sel[chosen.simu]!=0)

  neu<-apply(data.frame(binary[,NS]),1,sum)

  nic<-apply(data.frame(binary[,S]),1,sum)

  neu.nic<-data.frame(neu,nic)

  neu.nic[neu.nic>0]<-1

  neu.excl<-length(rownames(neu.nic)[neu.nic$neu==1&neu.nic$nic==0])

  nic.excl<-length(rownames(neu.nic)[neu.nic$nic==1&neu.nic$neu==0])

  t.excl<-length(rownames(neu.nic)[neu.nic$nic==1&neu.nic$neu==1])

  out.excl<-length(rownames(neu.nic)[neu.nic$nic==0&neu.nic$neu==0])

  small.table<-data.frame(No_species=rbind(Non_selection=neu.excl,Selection=nic
.excl,Both=t.excl,Outside=out.excl))

  final.results<-list(full.table=Table.S2,  binary=binary, small.table=small.ta
ble)

  x11(8,7)

  par(mar=c(6,6,5,2), cex=1.4, xpd=F)

  plot(EBNB_list[[1]][,1],EBNB_list[[1]][,2], xlab="Environmental Boundaury",

       ylab="Niche Breadth", type="n", main=paste("Empirical data -", title), b
ty="l",

       xlim=xrange, ylim=yrange, las=1, tck=-.01, lwd=2, cex.main=.9)

  NS.simu.colors<-colorRampPalette(c("darkblue",  "steelblue1"))(length(levels(a
s.factor(CompModel$SUMMARY$u))))

  S.simu.colors<-colorRampPalette(c("darkgreen",  "springgreen2"))(length(levels
(as.factor(CompModel$SUMMARY$u))))

  color.vector<-CompModel$SUMMARY$u

  model_colors<-c(rep(NS.simu.colors,table(color.vector[CompModel$SUMMARY$sel==
0]) ),

                  rep(S.simu.colors,  length(levels(as.factor(CompModel$SUMMARY$
sel)))))

  for(i in 1:length(chosen.simu))

  {  tryCatch({dataEllipse(data.frame(EBNB_list[i])[,1], data.frame(EBNB_list[i]
)[,2], col=model_colors[i], levels=c( conf.level), add=T,plot.points=F, robust=
T, lwd=1.8, fill=T, lty=1 , fill.alpha=.2, center.pch=F, fig=c(2,3,4,5))}, erro
r=function(e){})}

  points(EBNB_obs[,1],EBNB_obs[,2], pch=16, col='black', cex=1)

  if(is.null(freq.list)!=T)

  {  freq.colors<-colorRampPalette(c('darkred', 'cornflowerblue','seagreen'))(le
ngth(freq.list))

  sp_colors<-freq.list

  for(i in 1:length(freq.list)) {sp_colors[[i]]<-rep(freq.colors[i],length(freq
.list[[i]]) )}

  sp_colors<-rapply(sp_colors, c)

  sp_names<-rapply(freq.list, c)
```

```
  names(sp_names)<-NULL

  names(sp_colors)<-NULL

  data.colors<-data.frame(sp_names, sp_colors)

  sp_colors<- as.character(data.colors[with(data.colors, order(sp_names)), ][,2
])

  for(i in 1:length(names(freq.list)))

  { points(EBNB_obs[which(rownames(EBNB_obs)%in%as.vector(freq.list[[i]])),1],

           EBNB_obs[which(rownames(EBNB_obs)%in%freq.list[[i]]),2], pch=16, col
=freq.colors[i], cex=1)}

  if(is.null(point.sp)!=T)

  { points(EBNB_obs[point.sp,1],EBNB_obs[point.sp,2], pch=15, col='white', cex=
1.6)

    points(EBNB_obs[point.sp,1],EBNB_obs[point.sp,2], pch=16, col=sp_colors[poi
nt.sp], cex=1)}

  legend("topleft",inset = 0,pch=16, legend=names(freq.list),

         col=freq.colors, cex=.8, bty="n")} else{ points(EBNB_obs[,1],EBNB_obs[
,2], pch=16, col="black", cex=1)}

  if(is.null(point.sp)!=T)

  { points(EBNB_obs[point.sp,1],EBNB_obs[point.sp,2], pch=22, col='black', cex=
1.6)

    for(i in 1:length(point.sp)){ text(EBNB_obs[point.sp[i],1]+0.3,EBNB_obs[poi
nt.sp[i],2]+0.3,labels=letters[i], cex=.9, font=2)}}

  cat("\n",round(length(which(apply(binary, 1, sum)==0))/nrow(EBNB_obs)*100,2),
'% of species were not included in any simulation ')

  return(final.results)}
```

virtual.sampling

The function “virtual.sampling” returns the samples which the simulated coordinates are closest to the empirical locations.

Argurments: - coords: the empirical site coordinates - coord_simu: Simulated coordinates table created from expand_grid function

```
virtual.sampling<-function(coords, coord_simu)

{

  require(RANN)

  new.samples<-NULL

  new.coord<-data.frame(sample=matrix(NA,nrow(coords),2))

  sample.names<-1:nrow(coord_simu)
```

```
  for(i in 1:nrow(coords))
  { pt<-nn2(data = coord_simu, query = coords[i,], k = 1)$nn.idx
        new.samples[i]<-sample.names[pt]
  sample.names<-sample.names[-pt]
        new.coord[i,]<-coord_simu[pt,]
  coord_simu<-coord_simu[-pt,]}
  return(new.samples)
}
```

RAD_compare

The function “RAD_compare” generates a plot combining the observed rank abundance distribution (RAD) and the confidece limits from the simulated RAD. The function returns upper and lower confidence limits for each simulated specie rank, equivalent to the 97.5 and 2.5 percentiles, respectively.

Arguments: - obs: observed community table - sim: simulated community table - n.sim: number of bootstraps

```
RAD_compare<-function(obs,sim, n.sim)
{
  obs<-obs[,apply(obs,2,sum)>0]
  obs<-decostand(obs, "normalize")
  obs<-log(sort(apply(obs, 2,sum),decreasing=T)+1)
  plot(obs, type="n",
       main="", las=1, xlab="Species rank", ylab="Abundance", cex=1.2, bty="l")
  points(obs, col="black", pch=16)
  S<-length(obs)
  boot.result<-matrix(NA,n.sim,length(obs))
  probs<-apply(sim,2,sum)/sum(apply(sim,2,sum))
  sim<-decostand(sim,"normalize")
  sim<-sim[,apply(decostand(sim,"pa"),2,sum)>0]
  for( i in 1:n.sim)
  {
    new.sample<-sample(1:ncol(sim),S)
```

```
    sim.new<-sim[,new.sample]
    sim.new<-decostand(sim.new,"normalize")
    boot.result[i, ]<-log(sort(apply(sim.new, 2,sum),decreasing=T)+1)
    message("\n loop", i)
  }
  conf<-apply(boot.result,2,quantile,probs=c(0.005,0.995,0.5))
  up<-sort(conf[2,], decreasing = T)
  dow<-sort(conf[1,], decreasing = T)
  mean<-sort(conf[3,], decreasing = T)
  points(obs)
  x<-1:S
  points(x[obs>=dow&obs<=up],obs[which(obs>=dow&obs<=up)], col="steelblue4", pch=16)
  lines(up, lty=2, col="red")
  lines(dow, col="red",lty=2)
  legend("topright",legend=c(paste(length(obs[which(obs>=dow&obs<=up)]),"  species fitted")), bty="n")
  result<-data.frame(x,obs,up,dow, mean)
  result$sig<-FALSE
  result$sig[result$x[result$obs>=result$dow&result$obs<=result$up]]<-TRUE
  return(result)}
```

Mantel_compare

The function “Mantel_compare” generates a plot combining the (a) the mantel correlogram obtained from the observed data and; (b) confidence limits of the mantel correlogram obtained from the simulated data. The confidence limits are constructed using a resampling level based on the proportion of observed and simulated number of spcies (ncol(obs)/ncol(sim)).

Arguments: - coords: observed site coordinates - obs: observed community table - sim: simulated community table

```
Mantel_compare<-function(coords,obs,sim)
{
  require(ecodist)
  level=ncol(obs)/ncol(sim)
  sampleloc.edist <- vegdist(coords,"euclidean")
```

```
  obs.mantel<-vegdist(obs, "bray")
  sim.mantel<-vegdist(sim, "bray")
  obs <-mgram(obs.mantel, sampleloc.edist,mrank=T,nboot=999)
```

```
  n.class<- nrow(obs$mgram)
  sim <-mgram(sim.mantel, sampleloc.edist, nclass =n.class,mrank=T,  pboot=level,nboot=999)
  x<-1:n.class
  obs<-obs$mgram[,3]
  up<-sim$mgram[,5]
  dow<-sim$mgram[,6]



  plot(x,obs, ylim=c(min(up), max(dow)), type="o", col="steelblue4",
       pch=16, xlim=c(0,n.class), las=1,bty="l", ylab="Mantel I", xlab="Distance classes", main="Correlogram")


  points(x[obs>=dow&obs<=up],obs[which(obs>=dow&obs<=up)], col="gray60", pch=16)


  lines(up, col="gray40", lty=2)
  lines(dow, col="gray40", lty=2)
  return(list(obs=obs,sim=sim))



}
```

RDAs_compare

The function “RDAs_compare” performs a redundancy analysis (RDA) on both observed and simulated communities. It also performs a permutational tests of multivariate dispersion. The function returns a list containing the (a) RDA results of both observed and simulated data and; (b) the anova table from the multivariate dispersion test.

Arguments: - obs: observed community table - envi: observed environmental table - sim: simulated community table - envi_simu: simulated environmental table

```
RDAs_compare<-function(obs,envi,sim,envi_simu)
{ sim<-decostand(sim,"normalize")
  obs<-decostand(obs,"normalize")
  cat("\n 1. RDA")
  RDA.obs<-dbrda ( obs~ ., envi,  distance = 'euclidean', sqrt.dist=T)
  RDA.sim<-dbrda (sim~ ., envi_simu, distance = 'euclidean',sqrt.dist=T)
  par(mfrow=c(1,2))
  plot(RDA.obs)
  plot(RDA.sim)
  RDAs<-list(obs=RDA.obs,sim=RDA.sim)
  cat("\n 3. Full betadisper")
  dat1a<-scores(RDA.obs, display="sites")
  dat1b<-scores(RDA.sim, display='sites')
  fac<-c(rep(1,nrow(dat1a)),rep(2,nrow(dat1b)))
  dist11<-vegdist(rbind(dat1a,dat1b),"euclidean")
  betadisp<-permutest(betadisper(dist11,fac), permutations=999)
  resulta<-list(RDAs=RDAs,betadisp=betadisp)
  return(resulta)
}
```

freq_groups

The function freq_groups performs a piecewise regression to define species frequency groups. It returns a list of species classified in each frequency group (freq.list object).

Args: - abund: abundance table (samples as rows x species as columns) - breaks: break-point(s) in the segments. If is NULL, the break-points are automatically detected. - logscale: logical, indicating if data shold be transformed in log scale - plot.seg: logical, indicating if the graph should be displayed.

```
freq_groups<-function(abund, breaks=NULL,logscale=T,plot.seg=F)
{ repeat
{ suppressPackageStartupMessages(library(segmented))
  freq<-na.omit(abund)
  freq[freq>0]<-1
  freq<-apply(freq, 2, sum)
```

```
  freq.tab<-data.frame(table(freq))

  freq.tab[,1]<-as.numeric(freq.tab[,1])

  x<-freq.tab[,1]

  y<-freq.tab[,2]

  if(logscale==T) {y<-log(freq.tab[,2])}

  lin.mod<-lm(y~x)

  if(is.null(breaks)){breaks<-median (x)} else {breaks<-breaks}

  repeat {tryCatch(segmented.mod<-segmented(lin.mod, seg.Z = ~x, psi=breaks), e
rror = function(e) {})

    if(exists('segmented.mod')==T) break}

  est.breaks<-NULL

  if (nrow(segmented.mod$psi)>=2)

  { for(i in 1:nrow(segmented.mod$psi))

  {  est.breaks[i]<-segmented.mod$psi[i,2]

  }} else {est.breaks<-segmented.mod$psi[1,1]}

  if(plot.seg==T) {plot(x, y, main=" Piecewise regression", xlab="Number of Obs
ervations", ylab="Number of species", bty="l", pch=16)

    plot(segmented.mod, add=T)}

  ngroups<-length(est.breaks)+1

  freq.groups<- vector("list",ngroups)

  est.breaks<-c(0,est.breaks,max(x)*10)

  is.between <- function(x, a, b) { x>= a & x<= b}

  for( i in 1:ngroups)

  { freq.groups[[i]]<-names(which(is.between(freq, est.breaks[i], round(est.bre
aks[i+1]))==T))}

  names(freq.groups)<-paste("Freq",1:ngroups, sep="")

  if(sum(sapply(freq.groups, length))==ncol(abund)) break}

  class(freq.groups)<-"freq.list"

  return(freq.groups)

}
```

compartments

The function “compartments” classifies the observed species into four components" (a) non-selection“, (b)”selection“, (c)” Both" and (d)"outside. The function returns either a list with the species included into each component (return.sp=T), or a data frame with the number of species included into each component (return.sp=F).

Arguments: - obs: observed community table - EBNB_result: object created from the function EBNB_ANALYSIS - NS: vector informing which of the simulations are non-selection based - S: vector informing which of the simulations are selection based

```
compartments<-function(obs,EBNB_result, NS=c(1:4), S=(5:8))
{ probs<-EBNB_result$binary
  neu<-apply(data.frame(probs[,NS]),1,sum)
  nic<-apply(data.frame(probs[,S]),1,sum)
  neu.nic<-data.frame(neu,nic)
  neu.nic[neu.nic>0]<-1
  neu.obs<-obs[,rownames(neu.nic)[neu.nic$neu==1]]
  nic.obs<-obs[,rownames(neu.nic)[neu.nic$nic==1]]
  neu.excl<-obs[,rownames(neu.nic)[neu.nic$neu==1&neu.nic$nic==0]]
  nic.excl<-obs[,rownames(neu.nic)[neu.nic$nic==1&neu.nic$neu==0]]
  t.excl<-obs[,rownames(neu.nic)[neu.nic$nic==1&neu.nic$neu==1]]
  out.excl<-obs[,rownames(neu.nic)[neu.nic$nic==0&neu.nic$neu==0]]
  return(list(NS=neu.excl,S=nic.excl,B=t.excl,O=out.excl))
}
```

Analysis

Loading the datasets

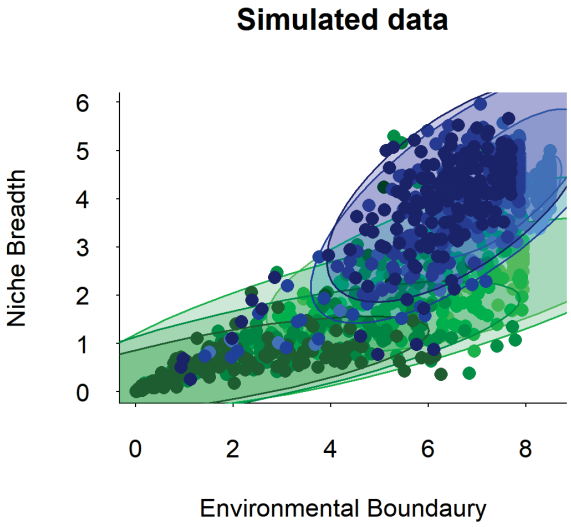
```
load("envi") ## environmental table
load("coords") ## coordinates table
load("abundF") ## species abundance table
```

Simulating the environment

```
envi_simu<-expand_grid(envi, coords, 60)
coord_simu<-expand_grid(envi, coords, 60, coord.return = T)
sim.list<-simu_loop(S=194,M=60,m=0.0001,b=0.505, d=0.5, K=1037, sel.vector=c(10
0,300,600,900), u.vector=c(0.001,0.01,0.1,0.4), envi_simu, cycles=20) # take se
veral minutes...
```

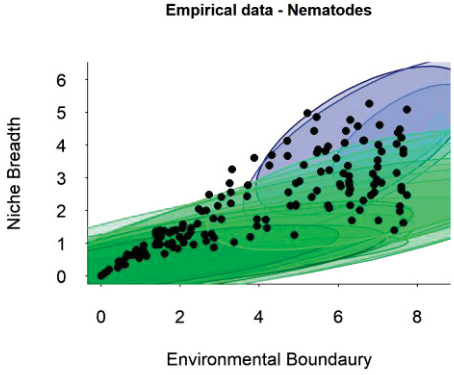
Composite model

```
CompModel<-Build_CompModel(sim.list, envi_simu, PC=1)

EBNB_conf<-EBNB_space(CompModel, chosen.simu =select )
```



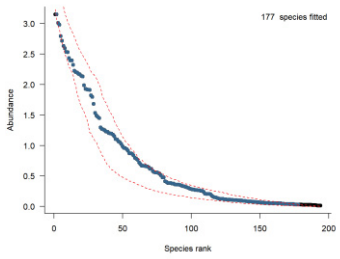
EBNB analysis using all simulations

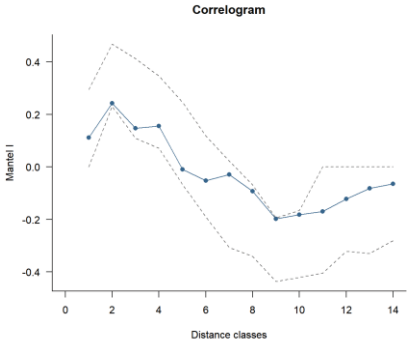
```
obs.full_EBNB<-EBNB_ANALYSIS(abundF,envi,PC=1, CompModel) #Number of observed s
pecies encopassed by each simulation

##

##  2.58 % of species were not included in any simulation

predicted.EBNB<-apply(obs.full_EBNB$binary,2,sum) #simulations selected to form
the final composite model
```

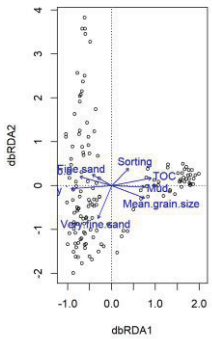


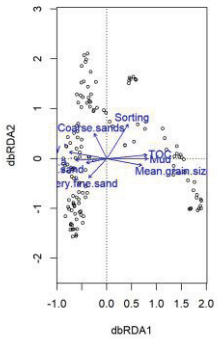
Validation

```
ns<-virtual.sampling(coords, coord_simu) # Sampling coordinates

sim.commu<-do.call("cbind",CompModel$ABUND_list[select])[ns,] ### sampling the s
imulated communitie

rad<-RAD_compare(abundF,sim.commu,n.sim=999) # Rank abundance distribution - co
mparasion

mantel<-Mantel_compare(coords,abundF,sim.commu) # Mantel correlogram - comparas
ion

RDA.result<-RDAs_compare(abundF,envi,sim.commu,envi_simu[ns,]) # Redundancy ana
lysis - comparasion

##

##  1. RDA

##

##  3. Full betadisper
```









Implementation

```
freq.list<-freq_groups(abundF, breaks=c(10,23))

names(freq.list)<-c("rare", "semi-resident","resident")

obs_EBNB<-EBNB_ANALYSIS(abundF,envi,1, CompModel, chosen.simu = select)

##

##  3.61 % of species were not included in any simulation

head(obs_EBNB$full.table)

##                     Abund Freq.group         NP       EB       NB

## Terschellingia_5 13827.75         NA  2.6966725 7.603886 2.721466

## Terschellingia_1  6724.84         NA  1.6699159 7.654898 3.858321

## Dorylaimopsis_1   6163.87         NA -0.3224786 7.097548 2.841815

## Odontophora_1     6135.22         NA  2.1397916 7.737426 5.086396
```

```
## Comesoma_1        5812.87         NA -1.0207512 7.562702 1.971469
## Cobbia_1          5451.65         NA  2.1428186 7.654898 3.738814
##                            Sim
## Terschellingia_5       19, 20
## Terschellingia_1  1, 2, 3, 20
## Dorylaimopsis_1  1, 2, 19, 20
## Odontophora_1         1, 2, 3
## Comesoma_1         18, 19, 20
## Cobbia_1          1, 2, 3, 20
head(obs_EBNB$small.table)
##               No_species
## Non_selection         21
## Selection            122
## Both                  44
## Outside                7
```

4 THE RELATIVE CONTRIBUTION OF NON-SELECTION AND SELECTION PROCESSES IN MARINE BENTHIC ASSEMBLAGES

Danilo Cândido Vieira[1,2]; Guilherme Corte[,2]; Hélio Checon Nascimento, [,2] Antônia Cecília Amaral Zacagnini[,2]; Gustavo Fonseca[3]

[1]Centro de Estudos do Mar - Universidade Federal do Paraná - Caixa Postal 50.002, 83255-000, Pontal do Paraná, PR, Brazil.

[2] Universidade Estadual de Campinas – Cidade Universitária “Zeferino Vaz”,– CEP 13083-970, Campinas, SP, Brazil.

[3] Universidade Federal de São Paulo – Av. Dona Ana Costa, 95 – CEP 11060-001, Santos, SP, Brazil.

Keywords: macrofauna, meiofauna, simulation, niche breadth, environmental boundary, environmental filtering, dispersal

Corresponding author: vieiradc@yahoo.com.br

phone: +55 16 993 626 902

4.1 RESUMO

O ambiente bentônico marinho é composto por uma grande diversidade de organismos, que variam em tamanho e histórias de vida. Tal complexidade é resultado da interação entre os processos baseados em seleção (nicho) e não-seleção (neutralidade). Entretanto, ainda é grande o desafio para quantificar a influência relativa destes processes em grupos de organismos que contrastam em tamanho e estratégia de vida. Através simulações computacionais, este estudo investigou a borda ambiental (EB) e a amplitude de nicho (NB) de espécies de nematoides e macrofauna, a partir de um desenho amostral espaço-temporal de uma baía costeira. Nossa hipótese é que a omnipresente meiofauna cria assembleias estruturadas principalmente por processos de não-seleção, enquanto que a dispersão ativa da macrofauna resulta em espécies mais ambientalmente restritas. Nós comparamos os EB-NBs da macrofauna e nematoides com as expectativas de comunidades simuladas a partir de diferentes pressupostos de equivalência de espécies, dispersão e filtragem ambiental. Nossas simulações explicaram 96% e 53% das assembleias de nematoides e macrofauna, respectivamente. Ambas as espécies de macrofauna e nematoides foram principalmente explicadas por processos de seleção (85% e 52%, respectivamente). Embora nosso estudo rejeitou a hipótese de que os nematoides são estruturados principalmente por processos de não-seleção, cerca de 34% das espécies de nematoides e 7% das espécies de macrofauna foram explicado por processos de não-seleção. Estes resultados destacam a importância, ainda que negligenciada, da dispersão, colonização e dinâmica estocástica de extinção local na determinação da distribuição das comunidades bentônicas ao longo do *continuum* hipotético de seleção/não-seleção.

4.2 ABSTRACT

The benthic environment is teeming with an enormous diversity of organisms varying in size and life-histories. Such complexity is a result of the interplay between non-selection and selection processes. Nevertheless, there is still a challenge in community ecology to quantify their relative influence among different species. Using a simulation-based approach, we investigated the environmental boundary (EB) and niche breadth (NB) of macrofauna and free-living nematodes species across a spatiotemporal sampling design at a coastal bay with a clear environmental gradient. We hypothesized that the ubiquitous of meiofauna creates assemblages that are driven mainly by non-selection processes, whereas the active dispersal of macrofauna results in species more environmentally-restricted. We compared the empirical EB-NBs of macrofauna and nematodes with generated-expectations of simulated communities using different assumptions of species equivalence, dispersal and, environmental filtering. Our simulations explained 96% and 53 % of nematodes and macrofauna assemblages, respectively. Both macrofauna and nematodes species were mainly explained by selection processes (85% and 52%, respectively.) Although we rejected the hypothesis that nematodes are mainly driven by non-selection processes, about 34% of nematodes species and 7% of macrofauna species were explained by non-selection processes. These results highlight the importance yet overlooked role of dispersal, colonization and local extinction stochastic dynamics in determining the location of benthic communities along the hypothesized non-selection - selection continuum.

4.3 INTRODUCTION

The benthic environment comprises the most widespread habitats on Earth and houses an enormous diversity of organisms varying in size and life-history (Snelgrove 1997, Rundell and Leander 2010). Most of the current knowledge about the structure of benthic communities is grounded on selection-processes and assumes the matching between the environment and species' niche requirements as the primary structuring (Gray 2002, Josefson 2016, Corte et al. 2017). However, it has been increasingly recognized that non-selection processes, such as dispersal limitation and ecological drift, plays an additional role (Thompson and Towsend 2006, Vinet and Zhedanov 2010). Furthermore, the intensity of both non-selection and selection mechanisms can vary between small and large-bodied species and; between passive and active dispersers. Larger taxa have higher dispersal limitation and lower plasticity in their niche requirements compared to the smaller ones (Shurin et al. 2009, Farjalla et al. 2012). Active dispersers, in turn, have lower dispersal limitation than passive ones because they are independent from vectors and can actively select for suitable habitats (Heino 2013, Rádková et al. 2014, Heino et al. 2015). Because larger benthic taxa usually hold for active dispersal, much uncertainty still exists about the importance of environmental filtering and dispersal between benthic groups with varying size and dispersal capabilities.

Meiofauna and Macrofauna are two contrasting size groups of organisms widely studied by benthic ecologists. The macrofauna is composed of organisms retained on a sieve of 0.5 mm mesh size, and the meiofauna are those which pass through a sieve of 0.5 mm mesh size and are retained on one of 0.063 mm (Warwick et al. 2006, Giere 2009). They are ecologically distinct from each other: all meiofaunal organisms are in-situ breeders and lack planktonic larval stages, thus having limited dispersal capabilities (Eleftheriou and McIntyre 2007). Besides, meiofauna have typically faster generation times than macrofauna and occur at significantly higher densities (Coull 1999, Giere 2009). In contrast, dispersal limitation in macrofauna is often seen as unimportant (McClain and Rex 2015). The planktonic larval phases of many macrofaunal species allow for long-distance dispersal and enable species to track environmental conditions before settlement (Pilditch et al. 2015).

The present study investigated the relative importance of dispersal and environmental filtering on both macrofauna and nematodes using two species-level

metrics: environmental boundary (EB) and niche breadth (NB). Briefly, EB is directly related to species dispersal and describes the entire environmental range occupied by the species, whereas NB reflects the variance of resources used by a species along an environmental gradient (MacArthur 1972, Swihart et al. 2006, Hirzel and Le Lay 2008). Given the differences in lifestyles and dispersal strategies between macrofauna and meiofauna, we hypothesized that the ubiquitous of meiofauna creates assemblages that are unlimited by its environment (i.e., species with show a large EB and NB). Macrofauna species, however, would be more environmentally-restricted (i.e., species with narrow EB and NB). If accepted, these hypotheses will support that nematodes are mainly driven by non-selection processes, while macrofauna by selective processes.

4.4 METHODOLOGY

4.4.1 The study area and sampling design

Four sampling campaigns were conducted at the Araçá Bay, southeastern Brazil (Fig. 1), over a year between 2012 and 2013 (October 2012, February 2013, May 2013, September 2013). The area is environmentally heterogeneous and includes different sedimentary textures, mangroves and, rocky shores (Amaral et al. 2016). The bay has a gentle slope reaching a maximum depth of 10 meters next to the open sea and tides vary between 2.06 m and -0.04 m (Gubitoso et al. 2008).

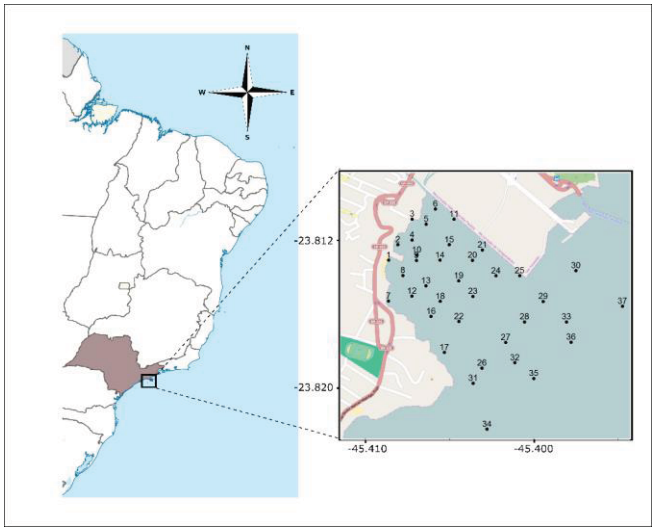


Fig. 1. Study Area. Sampling design at Araçá Bay, Brazil.

In each campaign, the sediment of 37 sites arranged in an irregular grid (Fig. 1) were taken for analysis of meiofauna, macrofauna, total organic carbon, microphytobenthos, and granulometry. All samples were taken during low tide. Samples were collected manually at the intertidal and shallow sublittoral (< 3 m deep), and with a multi-corer on deeper sites. For nematodes, one sample was collected using a PVC corer 2.5 cm in diameter and 5 cm deep. For macrofauna, four samples were collected using a corer 10 cm in diameter and 20 cm deep. Additional sediment samples were taken to evaluate the environmental gradient. Microphytobenthos samples were obtained from five replicates of the top 1 cm of the sediment using a corer 2 cm in diameter. These samples were conditioned in dark bottles and stored at -20°C. One additional sample of sediment was taken for granulometric and organic content analysis using a corer 3 cm in diameter and 20 cm deep.

4.4.2 Sampling processing

Macrofauna samples were stored in plastic bags and posteriorly sieved with a 0.3 mm mesh. The organisms retained were sorted in taxonomic groups and fixed in 70% ethanol. All individuals were identified to the species level.

Samples for meiofauna evaluation were obtained using a corer of 2.5 cm in diameter and 5 cm depth. Meiofauna samples were fixed in 4% formalin immediately after sampling. In the laboratory, samples were washed through a 45 µm mesh sieve, and organisms were extracted by flotation with Ludox TM 50 (specific density 1.18) (Heip et al. 1985). All organisms were counted under a stereomicroscope. A total of 100 nematodes were randomly picked from each sample, transferred to a glycerol solution, and later mounted on permanent slides (Heip et al. 1985). Nematodes were identified to genus level and separated into morphological species.

Microphytobenthic analyses were performed from the estimated biomass of chlorophyll-a and phaeopigments. These pigments were extracted from each sample with 10mL 100% acetone together with 0.07g MgCO3 for 24h in the dark at 4° C. Absorbance was read at optical densities of 750, 665 and 430nm in a spectrophotometer, before and after acidification with 1N HCl, according to Plante-Cuny (1978). The calculation to obtain the content of chlorophyll-a and phaeopigments were performed by using the equations of Plante-Cuny (1978). Total organic carbon

was evaluated using a modified Walkley-Black titration method, described by Gaudette and Flight (1974).

4.4.3 Data analysis

All analyses were performed using R v.3.3.2 (R Development Core team 2008). Principal component analysis (PCA) was used to determine the major environmental gradient in the bay. A total of 9 environmental variables were ordinated using normalized Euclidean: chlorophyll *a*, bathymetry (meters), % total organic carbon, % coarse sands (as the sum of pebbles, very coarse, coarse and medium sands), % fine sands, very fine sands, mean grain size and sorting coefficient.

Following this, we investigated the patterns of commonness and rarity in nematodes and macrofauna. For each assemblage, we performed a piecewise regression procedure (Gray et al. 2005) where species are separated into frequency groups of log-normally distributed species (Ugland and Gray 1982, Magurran and Henderson 2003). Thereafter, measurements of niche position (NP), niche breadth (NB) and environmental boundary (EB) were obtained for all species using the R function “EBNB’ provided in Chapter 3). In this function, measurements of NP, NB, and EB were obtained respectively from the mean, variance and 95% of the interquartile range of the distribution of each species along the first PC score.

Lastly, we applied a model-data integration framework (Chapter 1), characterized by two key steps: (a) simulate communities under non-selection and selection models using parameters of the empirical community and its environment and (b) compare the empirical measurements of EB and NB with those obtained from the simulations.

4.4.4 Simulations

The purpose of the simulations was to obtain a composite model, whereby the strengths of non-selection and selections processes on empirical communities can be inferred. Hence, the composite model comprises two families of simulations: (a) the non-selection simulations, assuming individuals as ecologically equivalents and, (b) selection simulations, assuming individuals to have different probabilities of

establishment according to the environment. The ability of the composite model to predict observed patterns was recently discussed in Chapter 3.

Simulations were performed using a spatially explicit, individual-based model ruled by processes of birth, death, dispersal, and immigration (Bell 2000, 2003, 2005). The spatial template for the simulations comprised 60 x 60 habitat cells, all occupied by K individuals belonging to S species. The simulation proceeds as discrete death-birth cycles in the site: individuals are born and die with probabilities *b* and *d*, respectively, and randomly disperses to one of the adjacent habitats with probability *u*. Each marginal habitat receives a single new immigrant with probability *m*. Any habitat that exceeds K has individuals randomly removed until the local population is reduced to K. Environmental filtering is implemented by assigning each habitat to an environmental value (EV) and each species to a specific optimum (SO). Demographic rates of each habitat cell habitat were modeled as a function of the difference between EV and SO values: The closer EO and SO values are, the higher birth rates and lower death rates are (Smith and Lundholm, 2010).

In non-selection simulations, both EV and SO values are settled to 1 (ecological equivalence assumption). In selection simulations, EV and SO values are defined using the environmental gradient observed at Araçá Bay. To this end, we interpolated the measured environmental variables to a 60 x 60 size lattice (package "akima," Akima and Gebhardt 2016). Subsequently, we ran a PCA and used the first PC scores as the EV units for the selection-based simulations. Lastly, we used the niche position of observed species (obtained from OMI analysis) as SO units for all simulations.

Birth, death and immigration rates were settled according to previous studies: $b = 0.505$, $d = 0.5$, $m = 0.005$ (Bell 2000, 2003, 2005). Non-selection and selection simulations were performed using four degrees of dispersal: very low ($u = 0.001$), low ($u = 0.01$), medium ($u = 0.1$) and high ($u = 0.4$). Two different rounds of simulations were carried out for nematodes and macrofauna. In each round, parameters S and K were settled in accordance with empirical measurements from each assemblage. S was settled using the total number of species found during the campaigns (157 and 194 for macrofauna and nematodes, respectively; Appendix A1). K was settled using the maximum abundances per site observed over the whole data set (796 and 3,435 for macrofauna and nematodes, respectively (Appendix A1, 2018). Simulations from

each round were assembled into distinct composite models. Each simulation ran for 10000-time steps.

4.5 RESULTS

4.5.1 Habitat heterogeneity at Araçá Bay

The PCA analysis on the environmental parameters showed that PC1 and PC2 accounted respectively for 45.67% and 22.9% of the total variation present. Based on the PCA scores (Fig. 2), the distribution of the sites along axis 1 was mainly driven by differences in mud percentages, total organic carbon, the concentration of chlorophyll a and bathymetry. Deeper stations were characterized by higher percentage of sand and chlorophyll a, while shallower stations showed a larger quantity of organic matter and mud. The dominant gradients detected along the second axis of the PCA were the percentage of very fine and coarse sands.

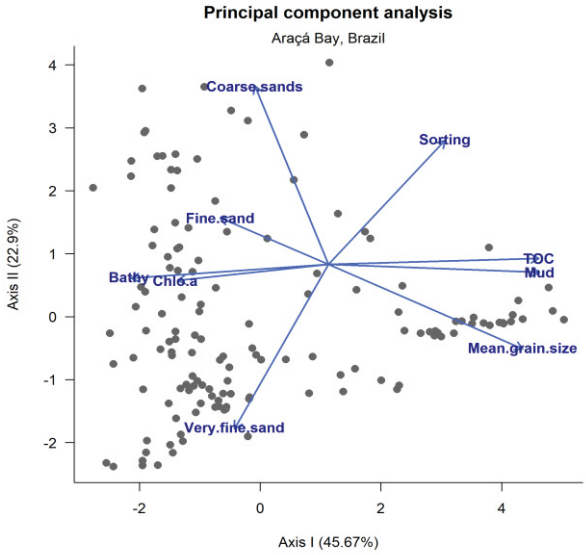


Fig. 2. Principal component analysis. PCA of the environmental variables evaluated at Araçá Bay, Brazil

4.5.2 Nematodes and macrofauna at Araçá Bay.

We recorded a total of 352 taxa (194 nematodes and 157 macrofauna species). The 4 most abundant nematodes species altogether accounted for 22.2% of

the total number of individuals (*Terschellingia* sp.5- 9.3%, *Terschellingia* sp.1- 4.5%, *Dorylaimopsis* sp.1 – 4.1% and *Odontophora* sp.1 – 4.1%) (Table 1). For macrofauna, the 4 most abundant species represented 77.7% of the total number of individuals (*Monokalliapseudes schubarti* – 45.7%, *Capitella nonatoi* – 25.8%, *Armandia hosfeldi* - *3.2%* and *Laeonereis acuta* - 2.9% (Table 1).

Table 1. EBNB results of the 4 most abundant species of nematodes and macrofauna. Number of individuals (Nºindv), frequency group (Freq.group), Niche Position (NP), environmental boundary (EB), and niche breadth (NB) of nematodes across the first PC axis; and the respective simulations in which species were included. Non-selection simulations: 1 and 2; Selection simulations: 3, 4, 5 and 6. Species are sorted by the total number of individuals recorded.

Benthic group	Specie	Nº indiv.	Freq.group	NP	EB	NB	Model
Nematodes	*Terschellingia* sp.5	19614	Resident	2.95	8.37	2.80	3, 4
	Terschellingia sp.1	9539	Resident	1.81	7.73	4.16	1, 2, 3
	Dorylaimopsis sp.1	8743	Resident	0.45	7.58	3.08	1, 2, 3, 4
	Odontophora sp.1	8702	Resident	2.31	8.37	5.53	1, 2
Macrofauna	*Monokalliapseudes schubartii*	5156	Resident	1.05	3.39	1.47	4, 5
	Capitella nonatoi	2912	Resident	1.87	3.48	1.34	4, 5
	Armandia hossfeldi	367	Resident	1.53	3.04	1.62	4
	Laeonereis acuta	330	Resident	1.54	2.38	1.00	4, 5

Three frequency groups were distinguished in nematodes (piecewise R2 = 0.74, Fig. 3a): rare (species occurring in 5 or fewer replicates), semi-resident (species that occurred in 6-30 samples) and resident (species that occurred in 31 or more samples). The nematode assemblage was characterized by 16% of abundant, 34% of semi-resident and 50% of rare species. For macrofauna, piecewise regression procedure separated into two groups: resident and rare species with the breakpoint at 11 sites (piecewise R2 = 0.87, Fig. 3b). Resident species comprised 12% of all macrofaunal species.

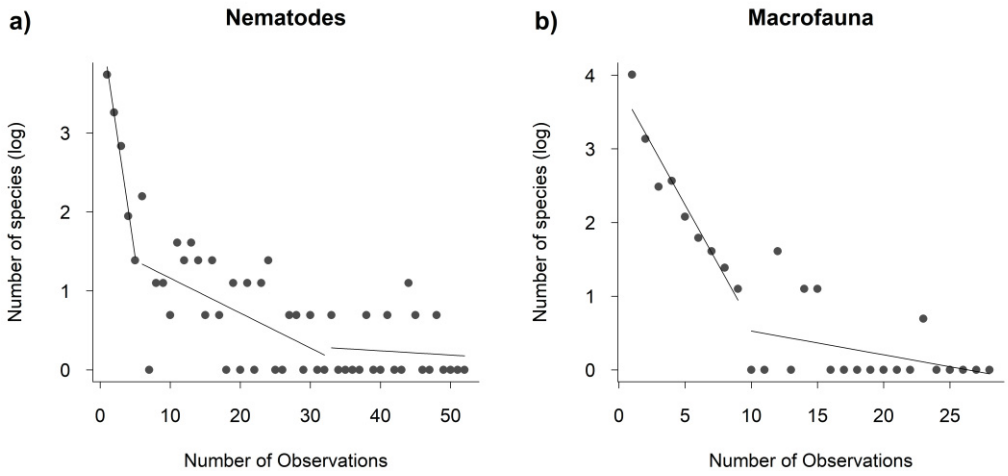


Fig. 3. Piecewise regression. Two and three frequency groups were distinguished in macrofauna (a) and nematodes (b) data, respectively.

4.5.3 EB-NB of Nematodes

The composite model explained a total of 96% of all nematodes species (Fig. 4a, b). Non-selection and selection simulations encompassed 34% and 85% of the empirical species, respectively. A total of 23% of the species were placed in the transition between the two families of simulations. Selection and non-selection simulations included rare, semi-resident and resident species. The non-selection simulations, together, accounted for 3%, 12% and 19% species classified respectively as rare, semi-resident and resident. Selection simulations, together, accounted for 42% of rare, 22% of semi-resident and 25% resident. All observations outside the ellipses were classified either as rare or as semi-resident species.

The position of the four most abundant nematodes species along the composite model varied according to the NB values. *Odontophora* sp.1 showed the largest value of NB (5.53) and was exclusively included within the non-selection ellipses. On the other hand, *Terschellingia* sp.5 showed the lowest NB (2.8), and exclusively included within the selection ellipses. *Dorylaimopsis* sp.1 and *Terschellingia* sp.1 were placed in the overlapping zone between non-selection and selection ellipses.

4.5.4 EB-NB of Macrofauna

A total of 53% of all macrofauna species were included within at least one ellipse (Fig. 4c, d). Together, ellipses from the non-selection simulations encompassed 7% of the empirical macrofauna species, whereas ellipses from selection simulations encompassed 52% of the species. Five percent of species were placed in the transition between the two families of simulations. Both families of simulations included rare and resident species. All the 73-species placed outside the ellipses were classified as rare. The non-selection models, together, accounted for 4% of rare and 3% of resident species. Selection models, together, accounted for 39% of rare and 13% of resident species. The four most abundant macrofaunal species showed similar values of EB and NB and were exclusively encompassed by selection simulations.

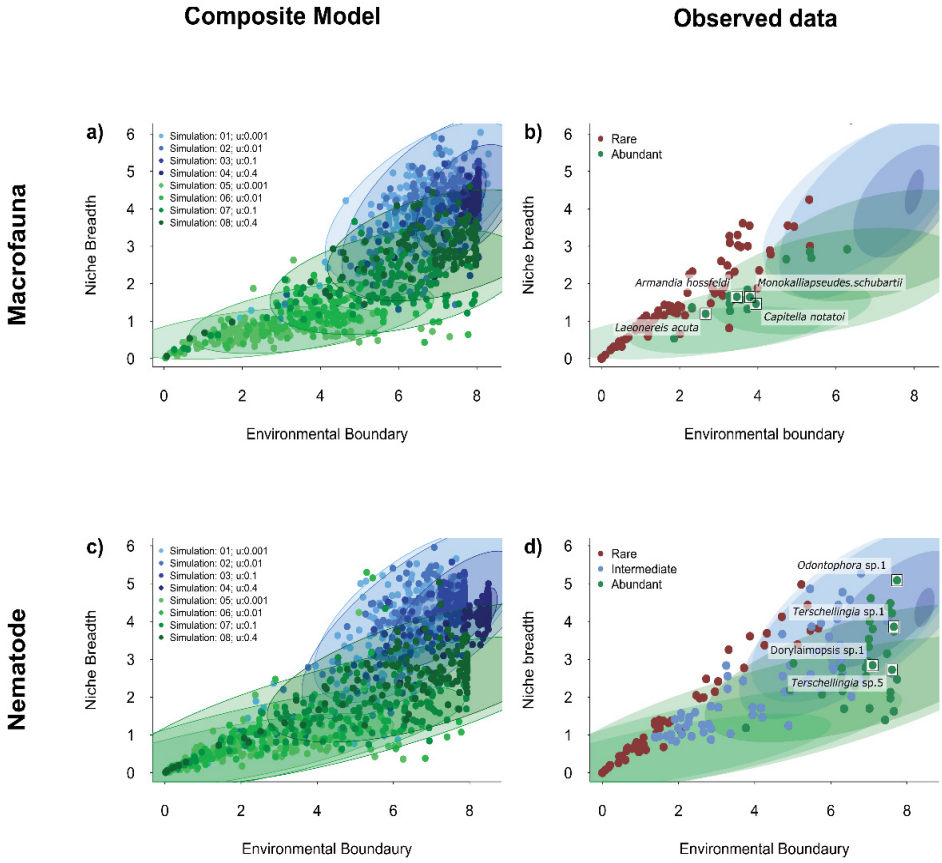


Figure 4. EB-NB array of nematodes and macrofauna. (a, c) EB-NB array from the composite models. (c, d) species position along the composite models (classified by frequencies). Each colored area of the graph represents 99% confidence ellipse of one simulated community

4.6 DISCUSSION

The present study corroborates that both non-selection and selection mechanisms play a role in benthic communities, with varying intensities depending on the group of organisms under study. The selection processes were the dominant structuring mechanism for both macrofaunal and nematode assemblages (Gray 2002, Josefson 2016). This indicates that environmental variables measured in this study (i.e. bathymetry, granulometric parameters, organic matter, and concentration of chlorophyll *a* are the major structuring factors in sorting species spatial distribution. Although our results rejected the hypothesis that nematodes are mainly driven by non-selection processes, about 34% of nematodes species were explained by non-selection processes, highlighting non-selection processes as additional and important drivers of nematodes assemblages.

The non-selection component of nematode assemblages in the Araçá Bay was characterized by species with large environmental boundary and niche breadth. Such component in nematodes may be a result of the interaction between huge populations sizes and stochastic environmental processes. Nematodes are the most abundant (densities between 105–108 individuals m^2) and species-rich meiofaunal group in marine sediments (Heip et al. 1985), which implies in a broad range of responses to the environmental conditions, ranging from high sensitivity to high resistance (Wilson and Kakouli-Duarte 2013). In addition, in the last decades, it has been demonstrated that meiofaunal organisms can be regularly found in the water column (Palmer 1988, Grüner et al. 2011) and colonize large defaunated areas (Sherman and Coull 1980). At the scale of the Bay, the distribution of approximately one-quarter of all species is not determined by the environment, but rather reflects the ability of the species to disperse and inhabit any part of the Bay. It because dispersal in nematodes is mostly passive (Palmer 1988), and even when active dispersal occurs (Ullberg and Ólafsson 2003, Schratzberger et al. 2004), it may act to facilitate active emergence from the sediment, thus increasing the chance of passive dispersal (Palmer 1988).

As expected, selection processes are the prime determinant of macrofauna communities. The predominant role of the environmental filtering on macrofauna has previously been highlighted by Labrune *et al.* (2007, 2008). Surprisingly, 47% of the species did not fit any of our simulations. All these species were classified as rare. The higher amount of rare species could be an indication of a stronger environmental

filtering. In theory, while resident species require a strong match between their niche requirements and local environmental conditions to maintain their populations (Supp and Ernest 2014), rare (or transient) species can be defined as those dispersed from other optimal habitats that cannot increase populations because they are continuously susceptible to the environmental filtering (i.e. source-sink dynamics) (Henderson and Magurran 2010, Supp et al. 2015). It is important to note that, rare occurrences could be also a consequence of inappropriate sampling strategy. Low abundance species are seldom caught in sufficient numbers with corers and, a large number of samples may be required to detect their optimum distribution (Rumohr 1999). Yet, instead of increasing the number of corer samples, trawls and dredges (which cover up to several hundreds of square meters) may also be used as an alternative to estimate densities of low-abundant macrofauna species (Eleftheriou and McIntyre 2007).

The current framework permitted to access the position of each species along the selection/non-selection gradient and make inferences about the main processes shaping their spatial distribution. For instance, four of the 4 most abundant macrofauna species (*Monokalliapseudes schubartii*, *Capitella nonatoi*, *Armandia hosfeldi,* and *Laeonereis acuta*) showed low values of EB and NB and were all included in the selection simulations of medium and low dispersal. The polychaetes *C. nonatoi*, *A. hosfeldi* and *L. acuta* are all described as active dispersers and the importance of the environmental filtering shaping their populations has been previously reported (Rouse and Pleijel. 2001, Corte et al. 2017). Actively dispersing taxa are relatively better at maintaining their position, as they are not as dependent on hydrodynamic conditions and can actively select for suitable habitats. Thus, if a species can actively choose its habitat, species are more likely to maintain sorting along environmental gradients according to their niche requirements (Vanschoenwinkel et al. 2007, Heino 2013, Rodil et al. 2017). The crustacean *M. schubartii*, in turn, is described as having limited locomotion capacity and absence of larval phases (Bemvenutti and E. & Netto 1998). However, this tube dwelling organism needs fine sedimentary fractions for constructing their tubes, and thus, the strong sedimentary gradient of the bay is likely to be its main constraining agent.

Yet, the 4 most abundant species of nematodes (*Terschellingia* sp.5, *Terschellingia* sp.1, *Dorylaimopsis* sp.1, and *Odontophora* sp.1) showed distinct positions along the hypothesized non-selection - selection continuum. Species of these genera are frequently reported as tolerant to a wide range of environmental conditions

and are typically found in enriched muddy sediments (Armenteros et al. 2010, Moens et al. 2013, Wilson and Kakouli-Duarte 2013). In our study, these genera showed similar values of EB, but different values of NB, suggesting similar dispersal potential but differential environmental filtering. *Terschellingia* sp.1 and *Terschellingia* sp.5 were both included by the selection models of high dispersal. However, their distinct values of NBs and, consequently, their distinct positions at the composite model are indicative of their different degrees of spatial distribution and most probably of specialization.

A major challenge facing community ecology research is to understand how changes at the species level scale up to impact the community and ecosystem properties. Our study allows identifying the set of species which are strongly constrained by environmental factors and suggest that the relative contribution of environmental and dispersal factors may differ in contrasting groups of species within a benthic community. This study highlights the importance yet overlooked role of dispersal, colonization and local extinction stochastic dynamics in determining the location of benthic communities along the hypothesized non-selection - selection continuum, adding novelty to the understanding of benthic communities. A step forward would be to improve the sampling assessments by using different scales and thus generate revised expectations of species environmental boundary and niche breadth for both nematodes and macrofauna.

4.7 APPENDIX S1 – SUPPLEMENTARY RESULTS

Table S1. Summary of the empirical evaluation for each benthic group. Percentage of empirical species included in each 99% confidence elipses

	Macrofauna	Nematodes
	S_{obs}	S_{obs}
Simulation 1	6%	30%
Simulation 2	5%	34%
Simulation 3	0%	11%
Simulation 4	4%	0%
Simulation 5	26%	48%
Simulation 6	33%	62%
Simulation 7	20%	79%
Simulation 8	4%	32%
Non-selection: 1,2,3,4	7%	34%
Selection simulations 5, 6, 7, 8	52%	85%
Overlap between families	5%	23%
Outside species	47%	4%

Table S2.2. Nematodes. Number of individuals (Nºindv), frequency classification (Freq.group), Niche Position (NP), environmental boundary (EB), and niche breadth (NB) of nematodes across the first PC axis; and the respective simulations in which species were included. Non-selection simulations: 1 and 2; Selection simulations: 3,4,5 and 6. Species are sorted by the total number of individuals recorded.

	Abund	Freq.group	NP	EB	NB	Model
Terschellingia sp.5	13828	resident	2.70	7.60	2.72	7, 8
Terschellingia sp.1	6725	resident	1.67	7.65	3.86	1, 2, 3, 8
Dorylaimopsis sp.1	6164	resident	-0.32	7.10	2.84	1, 2, 7, 8
Odontophora sp.1	6135	resident	2.14	7.74	5.09	1, 2, 3
Comesoma sp.1	5813	resident	-1.02	7.56	1.97	6, 7, 8
Cobbia sp.1	5452	resident	2.14	7.65	3.74	1, 2, 3, 8
Sabatieria sp.1	5102	resident	3.09	6.63	2.30	2, 7, 8
Stylotheristus sp.3	4689	resident	3.14	6.92	2.55	1, 2, 7, 8
Paramonhystera sp.1	4422	resident	-0.72	7.51	4.39	1, 2, 3

Subsphaerolaimus sp.1	4183	resident	-0.90	3.77	1.18	5, 6, 7, 8
Paramonhystera sp.2	3931	resident	2.88	7.48	3.15	1, 2, 7, 8
Daptonema sp.5	3726	resident	-1.22	6.35	1.69	6, 7, 8
Sabatieria sp.2	3455	resident	-0.82	7.65	1.62	6, 7, 8
Viscosia sp.1	3341	resident	-0.98	7.01	1.70	6, 7, 8
Aponema sp.1	3264	resident	2.96	7.74	2.47	7, 8
Trissonchulus sp.1	3043	resident	-1.12	7.42	1.39	7, 8
Terschellingia sp.2	3026	semi-resident	1.40	4.79	2.14	1, 2, 6, 7, 8
Gomphionema sp.1	2847	semi-resident	-0.72	4.89	1.25	5, 6, 7, 8
Ptycholaimellus sp.1	2608	resident	-0.31	6.29	2.77	1, 2, 7, 8
Paracanthonchus sp.1	2503	resident	-1.32	5.77	2.07	1, 2, 6, 7, 8
Sabatieria sp.1a	2484	resident	-1.54	5.02	2.89	1, 2, 7, 8
Tricotheristus sp.1	2475	resident	0.12	7.60	4.22	1, 2, 3, 8
Terschellingia sp.3	2394	resident	2.19	7.00	3.54	1, 2, 3, 8
Squanema sp.1	2305	semi-resident	-2.09	3.36	1.03	5, 6, 7, 8
Theristus longissimecaudatus	2076	resident	-0.93	6.91	2.79	1, 2, 7, 8
Vasostoma sp.1	2063	resident	2.23	6.98	2.95	1, 2, 3, 7, 8
Setosabatieria sp.1	1989	resident	1.29	7.10	3.80	1, 2, 3, 8
Sabatieria sp.5	1752	semi-resident	2.85	6.28	2.53	1, 2, 7, 8
Oncholaimellus sp.1	1676	resident	-1.10	7.56	2.13	6, 7, 8
Pseudosteineria sp.1	1389	semi-resident	-2.11	5.74	2.16	1, 2, 6, 7, 8
Metachromadora sp.1	1261	semi-resident	-1.94	1.93	0.83	5, 6, 7
Halalaimus sp.1	1240	resident	-0.73	7.01	4.01	1, 2, 3, 8
Metalinhomoeus sp.1	1213	resident	-0.38	6.99	3.33	1, 2, 3, 7, 8
Anoplostoma sp.1	1183	semi-resident	-0.87	6.31	2.16	2, 6, 7, 8
Metacyatholaimus sp.1	1122	semi-resident	3.05	4.93	2.18	1, 2, 6, 7, 8
Oncholaimus sp.1	1121	semi-resident	-1.18	3.95	1.53	6, 7, 8
Cyartonema sp.1	1097	resident	1.55	7.56	4.48	1, 2, 3
Paramonhystera sp.3	1067	semi-resident	-1.85	4.20	1.46	2, 5, 6, 7, 8
Elzalia sp.2	1021	semi-resident	2.62	4.93	2.84	1, 2, 7, 8
Molgolaimus sp.1	913	semi-resident	2.50	7.60	2.61	7, 8
Chromadora sp.1	848	semi-resident	-0.94	2.86	1.16	5, 6, 7
Metalinhomoeus sp.2	845	semi-resident	0.58	5.95	3.59	1, 2, 3, 8

Sabatieria sp.4	796	semi-resident	-0.38	6.79	5.26	1, 2
Daptonema sp.7	701	semi-resident	0.48	5.74	3.96	1, 2
Leptolaimus sp.1	673	semi-resident	0.11	5.78	3.24	1, 2, 8
Acanthonchus sp.1	654	semi-resident	-0.69	2.34	0.96	5, 6, 7
Sphaerolaimus sp.1	587	resident	1.12	7.05	4.61	1, 2, 3
Metoncholaimus sp.1	577	semi-resident	-1.13	6.83	2.46	2, 7, 8
Stylotheristus sp.1	516	semi-resident	-0.80	6.07	2.74	1, 2, 7, 8
Sphaerotheristus sp.1	515	semi-resident	2.97	6.13	3.14	1, 2, 3, 7, 8
Stylotheristus sp.4	515	semi-resident	-0.52	6.27	2.88	1, 2, 3, 7, 8
Microlaimus sp.8	501	semi-resident	-1.61	1.38	0.95	5, 6, 7
Paralongicyatholaimus sp.1	485	rare	-1.33	0.86	0.73	5, 6, 7
Cephalanticoma sp.1	484	semi-resident	-0.80	2.20	1.22	5, 6, 7
Steineria sp.1	455	semi-resident	1.31	7.48	4.03	1, 2, 3, 8
Halichoanolaimus sp.1	451	semi-resident	-0.90	2.09	1.02	5, 6, 7
Oncholaimellus sp.2	398	semi-resident	-1.65	6.54	2.01	6, 7, 8
Parodontophora sp.1	394	semi-resident	2.71	6.88	3.13	1, 2, 3, 7, 8
Theristus acribus	391	semi-resident	-1.54	2.85	0.87	5, 6, 7
Spilophorella sp.1	368	semi-resident	-1.10	2.66	1.07	5, 6, 7
Sphaerolaimus sp.2	362	semi-resident	2.27	5.33	2.41	1, 2, 7, 8
Choanolaimus sp.1	360	semi-resident	-0.53	5.47	2.60	1, 2, 7, 8
Elzalia sp.1	351	semi-resident	2.52	5.56	3.19	1, 2, 8
Oxystomina sp.1	351	semi-resident	-0.95	3.95	1.73	2, 6, 7, 8
Pomponema sp.1	343	semi-resident	-1.76	2.62	1.27	5, 6, 7
Hypodontolaimus sp.1	335	semi-resident	-2.03	1.49	0.96	5, 6, 7
Sabatieria sp.8	315	semi-resident	1.74	6.30	2.65	1, 2, 7, 8
Daptonema sp.6	312	semi-resident	3.02	6.66	4.14	1, 2, 3
Trefusia sp.1	308	rare	-1.31	2.07	1.57	6, 7
Sabatieria sp.2c	305	semi-resident	2.44	6.28	3.05	1, 2, 3, 7, 8
Desmolaimus sp.1	300	semi-resident	-1.07	4.17	1.72	2, 6, 7, 8
Polygastrophora sp.1	252	semi-resident	-0.93	2.79	1.28	5, 6, 7, 8
Chromadorina sp.1	242	semi-resident	-0.87	3.29	2.22	7, 8
Paracyatholaimoides sp.1	234	rare	-1.94	4.93	2.15	1, 2, 6, 7, 8
Microlaimus sp.4	227	semi-resident	0.71	6.20	4.43	1, 2

Paralinhomoeus sp.1	210	rare	-0.55	5.46	3.76	1, 2
Terschellingia sp.6	204	semi-resident	-0.35	4.72	3.67	1, 2
Stygodesmodora sp.1	198	semi-resident	2.05	4.74	2.56	1, 2, 7, 8
Ptycholaimellus sp.2	193	rare	-1.43	2.06	1.45	6, 7
Sphaerolaimus sp.3	174	semi-resident	2.02	6.63	4.13	1, 2, 3
Comesoma sp.2	166	semi-resident	-1.95	1.82	0.87	5, 6, 7
Pomponema sp.2	161	rare	-2.54	2.07	1.10	5, 6, 7
Campylaimus sp.1	161	semi-resident	3.02	6.05	2.81	1, 2, 7, 8
Metadasynemoides sp.1	153	semi-resident	-1.57	1.80	1.02	5, 6, 7
Thalassoalaimus sp.1	151	semi-resident	0.00	2.85	1.76	6, 7, 8
Polygastrophora sp.2	150	semi-resident	-1.28	1.99	1.19	5, 6, 7
Microlaimus sp.6	147	rare	-1.80	0.92	0.71	5, 6, 7
Thalassomonhystera sp.1	146	rare	-0.77	1.90	1.41	6, 7
Dasynemoides sp.1	145	semi-resident	1.40	6.31	4.77	1, 2
Calyptronema sp.1	138	semi-resident	-1.08	2.26	1.33	5, 6, 7
Marylynnia sp.1	130	semi-resident	-0.26	6.25	3.37	1, 2, 3, 8
Paramesonchium sp.1	118	semi-resident	-1.14	2.44	1.60	6, 7
Trefusia sp.2	110	rare	-0.77	0.96	0.55	5, 6, 7
Linhystera sp.2	106	semi-resident	2.35	5.49	3.86	1, 2
Leptolaimoides sp.1	104	semi-resident	-0.69	3.05	1.73	6, 7, 8
Cheironchus sp.1	104	rare	1.11	5.46	4.86	1, 2
Theristus sp.8	103	semi-resident	-1.46	2.28	1.47	6, 7
Laimella sp.1	84	semi-resident	1.38	6.50	4.58	1, 2
Parasphaerolaimus sp.2	81	rare	1.65	5.13	3.39	1, 2, 8
Parametoncholaimus sp.1	78	rare	-1.26	0.00	0.00	5, 6, 7
Theristus flevensis	75	semi-resident	-1.52	1.69	0.97	5, 6, 7
Cervonema sp.1	72	rare	-0.87	4.26	3.37	1, 2
Sabatieria sp.7	67	rare	3.08	3.04	2.41	none
Chromadorella sp.1	67	rare	1.12	1.39	0.95	5, 6, 7
Trochamus sp.1	64	rare	1.90	6.17	4.07	1, 2
Sabatieria sp.6	64	rare	2.63	5.39	4.43	1, 2
Monhystrella sp.1	63	rare	-0.89	3.69	2.43	2, 7, 8
Morfotipo sp.4	60	rare	-1.59	1.61	0.68	5, 6, 7

Marisalbinema sp.1	58	rare	2.56	0.64	0.58	5, 6, 7
Diplopeltula sp.1	56	rare	4.43	1.51	1.18	5, 6, 7
Synonchiella sp.1	52	rare	-1.94	0.51	0.44	5, 6, 7
Catanema sp.1	50	rare	-0.68	0.00	0.00	5, 6, 7
Rhabdocoma sp.1	50	rare	-0.68	0.00	0.00	5, 6, 7
Microlaimus sp.1	50	rare	-0.35	3.25	2.84	none
Dorylaimidae sp.1	49	rare	1.16	1.09	0.68	5, 6, 7
Prochromadorella sp.1	47	rare	-0.09	3.29	2.55	none
Xyalidae sp.1	45	rare	2.25	2.55	1.97	7
Linhystera sp.1	42	rare	3.21	2.66	1.99	7
Bathylaimus sp.1	40	rare	-1.31	0.78	0.64	5, 6, 7
Desmodora sp.2	40	rare	0.99	5.22	4.98	none
Wieseria sp.1	39	rare	-0.72	1.06	0.77	5, 6, 7
Doliolaimus sp.1	37	rare	-1.40	0.19	0.18	5, 6, 7
Siphonolaimus sp.1	37	rare	0.78	5.67	3.81	1, 2
Nannolaimus sp.1	36	rare	2.08	0.67	0.66	5, 6, 7
Steineria sp.2	36	rare	0.14	3.72	2.77	8
Ascolaimus sp.1	36	rare	3.08	2.96	2.14	7
Tripyloides sp.1	34	rare	1.12	4.71	4.12	1, 2
Actinonema sp.1	30	rare	0.80	4.32	3.69	1, 2
Eubostrichus sp.1	30	rare	1.14	3.88	3.61	none
Deontolaimus sp.1	30	rare	3.31	1.49	1.39	6, 7
Eurystomina sp.3	29	rare	-1.00	1.14	0.61	5, 6, 7
Odontanticoma sp.1	28	rare	-0.66	0.99	0.67	5, 6, 7
Valvaelaimus sp.1	28	rare	1.98	0.65	0.64	5, 6, 7
Metalinhomoeus sp.8	27	rare	4.81	0.00	0.00	5, 6, 7
Parachromadorita sp.1	25	rare	-1.41	1.08	0.93	5, 6, 7
Phanoderma sp.2	24	rare	-0.53	2.47	2.05	7
Diplolaimelloides sp.1	21	rare	2.71	1.73	1.32	6, 7
Metachromadora sp.4	21	rare	-1.45	0.08	0.07	5, 6, 7
Xyalidae sp.2	20	rare	2.37	0.00	0.00	5, 6, 7
Theristus sp.10C	20	rare	-1.67	0.00	0.00	5, 6, 7
Xyalidae sp.3	20	rare	-1.39	0.20	0.20	5, 6, 7

Dichromadora sp.1	19	rare	-2.02	1.42	1.05	5, 6, 7
Morfotipo sp.5	18	rare	-1.06	0.00	0.00	5, 6, 7
Chromadoridae sp.1	18	rare	-1.23	0.00	0.00	5, 6, 7
Parasphaerolaimus sp.1	16	rare	2.83	3.32	3.25	none
Prochromadora sp.1	15	rare	-0.97	0.00	0.00	5, 6, 7
Sphaerolaimus sp.4	15	rare	0.00	0.08	0.07	5, 6, 7
Theristus sp.F2	13	rare	-1.50	0.00	0.00	5, 6, 7
Morfotipo sp.1	13	rare	2.36	0.00	0.00	5, 6, 7
Promonhystera sp.1	13	rare	2.36	0.00	0.00	5, 6, 7
Southerniella sp.1	13	rare	-1.27	2.73	2.48	none
Linhomoeus sp.1	13	rare	-1.12	1.39	1.27	6, 7
Bolbolaimus sp.1	13	rare	-1.37	0.17	0.15	5, 6, 7
Nudora sp.1	12	rare	-1.97	0.61	0.34	5, 6, 7
Paracanthonchus sp.2	12	rare	-1.13	0.44	0.43	5, 6, 7
Pseudochromadora sp.1	12	rare	2.35	0.00	0.00	5, 6, 7
Tricoma sp.1	12	rare	-0.13	0.00	0.00	5, 6, 7
Comesomatidae sp.1	12	rare	-0.78	1.38	1.35	6, 7
Axonolaimus sp.1	11	rare	-0.97	0.00	0.00	5, 6, 7
Morfotipo sp.3	11	rare	1.28	0.00	0.00	5, 6, 7
Rhabditidae sp.8	10	rare	-0.20	0.43	0.26	5, 6, 7
Epacanthion sp.1	10	rare	1.47	0.00	0.00	5, 6, 7
Eleutherolaimus sp.1	10	rare	3.43	0.00	0.00	5, 6, 7
Scaptrella sp.1	10	rare	-2.06	0.00	0.00	5, 6, 7
Desmoscolex sp.1	10	rare	4.16	0.00	0.00	5, 6, 7
Longicyatholaimus sp.1	9	rare	-0.85	0.00	0.00	5, 6, 7
Morfotipo sp.2	9	rare	-1.06	0.00	0.00	5, 6, 7
Cienfuegia sp.1	9	rare	-1.37	0.21	0.20	5, 6, 7
Synonchium sp.1	8	rare	-1.41	0.00	0.00	5, 6, 7
Rhabditidae sp.4	8	rare	1.14	1.09	0.73	5, 6, 7
Polysigma sp.1	8	rare	-0.72	0.00	0.00	5, 6, 7
Aegialoalaimus sp.1	8	rare	1.38	0.00	0.00	5, 6, 7
Omicronema sp.1	7	rare	-1.41	0.00	0.00	5, 6, 7
Graphonema sp.1	7	rare	1.28	0.00	0.00	5, 6, 7

Rhabditidae sp.2	7	rare	1.28	0.00	0.00	5, 6, 7
Paramicrolaimus sp.1	7	rare	0.84	0.00	0.00	5, 6, 7
Rhabditidae sp.3	7	rare	-0.57	0.90	0.66	5, 6, 7
Atrochromadora sp.1	6	rare	-1.31	1.81	1.38	6, 7
Dolichosomatum sp.1	6	rare	1.79	0.00	0.00	5, 6, 7
Comesa sp.4	6	rare	3.76	0.46	0.35	5, 6, 7
Kraspedonema sp.1	5	rare	-1.29	0.00	0.00	5, 6, 7
Metaparoncholaimus sp.1	5	rare	-2.08	0.00	0.00	5, 6, 7
Microlaimidae sp.1	5	rare	3.68	0.00	0.00	5, 6, 7
Diplolaimella sp.1	4	rare	-0.21	1.60	1.38	6, 7
Dasynemella sp.2	4	rare	-1.04	0.00	0.00	5, 6, 7
Rhabditidae sp.7	4	rare	-0.47	0.90	0.80	5, 6, 7
Syringolaimus sp.2	4	rare	1.28	0.00	0.00	5, 6, 7
Trileptium sp.1	4	rare	1.28	0.00	0.00	5, 6, 7
Rhabditidae sp.9	3	rare	-0.72	0.00	0.00	5, 6, 7
Theristus sp.10	1	rare	-1.67	0.00	0.00	5, 6, 7
Araeolaimus sp.1	1	rare	-1.52	0.00	0.00	5, 6, 7
Leptogastrella sp.1	1	rare	-1.41	0.00	0.00	5, 6, 7
Rhabditidae sp.5	1	rare	-1.08	0.00	0.00	5, 6, 7
Rhabditidae sp.6	1	rare	-1.08	0.00	0.00	5, 6, 7

Table S3. Macrofauna. Number of individuals (Nºindv), frequency classification (Freq.group), Niche Position (NP), environmental boundary (EB), and niche breadth (NB) of nematodes across the first PC axis; and the respective simulations in which species were included. Non-selection simulations: 1 and 2; Selection simulations: 3,4,5 and 6. Species are sorted by the total number of individuals recorded. The respective table for nematodes was presented in Appendix S2 from Chapter 3.

Specie	Abund	Freq.group	NP	EB	NB	Simulation
Monokalliapseudes schubartii	5156	Resident	0.88	3.80	1.64	6, 7
Capitella nonatoi	2912	Resident	1.96	3.95	1.46	6, 7
Armandia hossfeldi	367	Resident	1.29	3.46	1.65	6, 7

Specie	Abund	Freq.group	NP	EB	NB	Simulation
Laeonereis acuta	330	Resident	1.71	2.66	1.20	5, 6
Oligochaeta	238	Resident	2.03	1.86	0.54	5, 6
Scoloplos leodamas sp. A	227	Resident	0.78	3.73	1.59	6, 7
Haploscoloplos sp.	226	Resident	0.79	3.73	1.84	6, 7
Capitella sp. F	180	Rare	1.58	3.27	0.81	5, 6
Heteromastus sp. A	168	Resident	1.21	3.51	1.66	6, 7
Scolelepis squamata	123	Rare	1.40	2.01	0.66	5, 6
Olivella minuta	118	Resident	1.01	3.73	1.32	5, 6, 7
Aricidea fragilies	94	Resident	0.27	3.29	1.27	5, 6, 7
Haploscoloplos sp. B	65	Resident	0.94	3.28	1.48	6, 7
Poecilochaetus australis	65	Rare	-1.43	4.33	2.79	1, 7, 8
Isolda pulchella	57	Rare	1.17	2.90	1.75	6, 7
Mediomastus sp. A	55	Resident	0.66	6.14	2.02	2, 7, 8
Hermundura tricuspis	53	Resident	0.53	3.57	1.57	6, 7
Anomalocardia brasiliana	44	Resident	1.61	3.27	1.67	6, 7
Armandia agilis	41	Resident	0.78	2.32	1.37	6
Corbula sp.	39	Resident	-0.18	6.28	2.92	1, 2, 7, 8
Glycinde multidens	36	Resident	0.64	3.73	1.81	6, 7
Marphysa sebastiana	30	Resident	0.94	5.36	2.69	1, 2, 7, 8
Carycorbula caribaea	28	Resident	-0.72	5.33	2.86	1, 2, 7, 8
Protankyra benedeni	27	Rare	-2.57	3.10	1.68	6, 7
Edwardsia sp.	25	Rare	0.47	1.91	1.24	6
Poecilochaetus perequensis	25	Rare	0.21	2.32	1.34	6
Haploscoloplos sp. A	21	Rare	0.14	2.14	1.16	5, 6
Neanthes bruaca	20	Resident	-2.32	4.73	2.65	1, 2, 7, 8
Scoletoma tetraura	18	Rare	0.73	3.99	1.88	6, 7
Capitella sp. G	15	Rare	2.44	1.61	1.43	none
Magelona papillicornis	15	Rare	1.04	3.33	1.56	6, 7
Magelona variolamellata	14	Rare	0.20	2.04	1.37	6
Cerithium atratum	13	Rare	0.39	1.21	0.70	5, 6
Microphiopholis subtilis	13	Rare	-2.62	4.03	2.35	7
Syllis cf cornuta	11	Rare	0.50	1.19	0.59	5, 6

Specie	Abund	Freq.group	NP	EB	NB	Simulation
Thysanocardia catarinae	11	Rare	-0.29	0.52	0.32	5
Laonice cirrata	10	Rare	-1.71	3.40	2.32	7
Magelona posterolongata	10	Rare	0.49	1.20	0.90	5, 6
Ophiothela danae	10	Rare	-0.40	0.00	0.00	none
Parapriono spiopinnata	10	Rare	-0.16	4.76	3.56	1
Tellina sp.	10	Rare	1.02	1.18	0.79	5, 6
Aphelochaeta marioni	9	Rare	0.24	5.33	3.00	1, 2, 7, 8
Clymenella dalesi	9	Rare	0.74	1.21	0.76	5, 6
Naineris bicornis	9	Rare	0.28	3.14	1.91	7
Pinnixa sayana	9	Rare	-1.85	4.31	2.89	1, 7
Tellina lineata	9	Rare	1.72	2.79	1.48	6, 7
Naineris setosa	8	Rare	0.90	3.05	1.78	6, 7
Nucula semiornata	8	Rare	0.55	1.93	1.21	5, 6
Upogebia brasiliensis	8	Rare	0.19	1.77	1.40	6
Bulla striata	7	Rare	0.59	1.85	1.24	6
Alpheus sp.	6	Rare	-2.17	1.36	1.05	5, 6
Capitella sp. E	6	Rare	-0.30	0.24	0.24	5
Felaniella candeana	6	Rare	0.48	1.75	1.28	6
Goniada maculata	6	Rare	-1.41	3.50	3.02	none
Laonice sp. B	6	Rare	-1.70	2.85	1.86	7
Owenia fusiformis	6	Rare	0.82	0.53	0.48	5
Rashgua lobatus	6	Rare	-0.39	3.27	2.23	7
Sigambra tentaculata	6	Rare	-1.61	5.31	4.25	1, 2
Tagelus divisus	6	Rare	0.73	2.80	1.96	none
Aricidea wassi	5	Rare	0.44	0.00	0.00	none
Automate sp.	5	Rare	-0.74	2.20	1.76	none
Gymnonereis crosslandi	5	Rare	-0.13	0.46	0.45	5
Microphiopholis atra	5	Rare	-0.03	1.07	0.79	5, 6
Mooreonuphis lineata	5	Rare	0.34	1.60	1.16	5, 6
Strigella camaria	5	Rare	0.22	4.93	3.53	1, 2
Terebellides anguicomus	5	Rare	-0.12	1.48	1.18	5, 6
Abra sp. 1	4	Rare	0.64	0.36	0.33	5

Specie	Abund	Freq.group	NP	EB	NB	Simulation
Bulla ocidentalis	4	Rare	0.13	0.87	0.76	5, 6
Eunice sp. 1	4	Rare	-0.27	0.52	0.45	5
Eunoe serrata	4	Rare	-2.05	3.48	3.30	none
Grubeulepis cf geayi	4	Rare	0.04	0.98	0.78	5, 6
Laonice branchiata	4	Rare	-0.40	0.00	0.00	none
Notomastus hemipodus	4	Rare	1.06	3.22	2.49	none
Ophellina sp.	4	Rare	1.10	0.37	0.32	5
Poecilochaetus sp.	4	Rare	-1.71	3.57	2.98	none
Sigambra grubei	4	Rare	0.48	1.96	1.42	6
Sternaspis capillata	4	Rare	-1.49	1.32	1.02	5, 6
Sthenelais cf limicola	4	Rare	-2.38	3.74	3.01	none
Strigella producta	4	Rare	0.94	1.81	1.28	6
Timarete filigera	4	Rare	0.59	1.36	1.11	5, 6
Diopatra aciculata	3	Rare	-0.38	0.39	0.32	5
Goniada littorea	3	Rare	0.27	0.63	0.52	5
Macoma uruguaiensis	3	Rare	-0.31	3.28	3.09	none
Magelona nonatoi	3	Rare	-2.24	3.06	2.60	none
Nassarius vibex	3	Rare	1.61	1.45	1.28	6
Paradentalium disparile	3	Rare	0.34	0.00	0.00	none
Pinnixa sp.	3	Rare	-3.19	0.70	0.58	5, 6
Polyonyx gibbesi	3	Rare	-1.60	3.79	3.55	none
Prinospio multibranchiata	3	Rare	-2.51	0.00	0.00	none
Protoaricia sp. A	3	Rare	-0.54	1.72	1.42	none
Sthenelais sp. C	3	Rare	-1.14	1.46	1.19	6
Albunea paretti	2	Rare	-0.77	3.27	3.27	none
Ambidexter symmetricus	2	Rare	-0.73	0.86	0.86	5, 6
Aphelochaeta sp.	2	Rare	-2.25	1.00	1.00	5, 6
Apseudes sp.	2	Rare	0.43	0.44	0.44	5
Automate cf evermanni	2	Rare	-2.52	0.00	0.00	none
Chaetopterus sp.	2	Rare	-0.29	0.23	0.23	5
Chione cancellata	2	Rare	1.54	0.00	0.00	none
Ctena pectinella	2	Rare	-1.39	2.25	2.25	none

Specie	Abund	Freq.group	NP	EB	NB	Simulation
Diplodonta punctata	2	Rare	0.30	0.45	0.45	5
Eunoe tuerkayi	2	Rare	-2.14	0.10	0.10	5
Hemipolis elongata	2	Rare	-1.33	2.33	2.33	none
Magelona sp. B	2	Rare	0.06	0.00	0.00	none
Ophiactis lymani	2	Rare	-0.40	0.00	0.00	none
Phyllodoce cf arenae	2	Rare	-0.41	6.38	6.38	none
Prinospio streenstrupi	2	Rare	-2.11	3.61	3.61	none
Scyphoproctus sp. A	2	Rare	-2.67	0.32	0.32	5
Sipunculus nudus	2	Rare	0.24	1.09	1.09	5, 6
Tellina gibber	2	Rare	1.93	0.03	0.03	5
Tellina martinicensis	2	Rare	-0.29	0.23	0.23	5
Tivella mactroides	2	Rare	0.30	1.15	1.15	none
Acantholobulus schmitti	1	Rare	1.39	0.00	0.00	none
Ampelisca sp.	1	Rare	-0.40	0.00	0.00	none
Ampelisciphotis sp.	1	Rare	0.52	0.00	0.00	none
Amphiodia pulchella	1	Rare	-0.17	0.00	0.00	none
Amphipholis squamata	1	Rare	2.78	0.00	0.00	none
Amphiura kinberg	1	Rare	-0.17	0.00	0.00	none
Anachis obesa	1	Rare	-1.79	0.00	0.00	none
Automate cf rugosa	1	Rare	-2.50	0.00	0.00	none
Callinectes cf danae	1	Rare	-1.12	0.00	0.00	none
Capitella sp. D	1	Rare	-0.24	0.00	0.00	none
Cirriformia sp.	1	Rare	1.00	0.00	0.00	none
Cirriformia tentaculata	1	Rare	2.93	0.00	0.00	none
Clymenella brasiliensis	1	Rare	0.91	0.00	0.00	none
Clymenella sp.	1	Rare	-0.40	0.00	0.00	none
Cooperela atlantica	1	Rare	0.13	0.00	0.00	none
Crassinella sp.	1	Rare	-0.09	0.00	0.00	none
Cylichna discus	1	Rare	-3.88	0.00	0.00	none
Dorvillea sp.	1	Rare	1.27	0.00	0.00	none
Exogone breviantennata	1	Rare	0.13	0.00	0.00	none
Harmothoe sp. A	1	Rare	-0.17	0.00	0.00	none

Specie	Abund	Freq.group	NP	EB	NB	Simulation
Hemipodia simplex	1	Rare	0.73	0.00	0.00	none
Idunella nana	1	Rare	-0.27	0.00	0.00	none
Juliacorbula acquivalves	1	Rare	-2.52	0.00	0.00	none
Maera sp.1	1	Rare	-3.88	0.00	0.00	none
Magelona sp. A	1	Rare	-2.51	0.00	0.00	none
Malmgreniella sp. A	1	Rare	0.12	0.00	0.00	none
Ogyrides alphaerostris	1	Rare	-2.82	0.00	0.00	none
Ophellina alata	1	Rare	0.98	0.00	0.00	none
Persephona crinita	1	Rare	0.81	0.00	0.00	none
Pholoe sp. B	1	Rare	-3.51	0.00	0.00	none
Phoxocephalopis sp.	1	Rare	0.60	0.00	0.00	none
Pitar fulminatus	1	Rare	0.08	0.00	0.00	none
Polydora websteri	1	Rare	1.23	0.00	0.00	none
Prinospio dayi	1	Rare	0.34	0.00	0.00	none
Prinospio dayi 1	1	Rare	1.46	0.00	0.00	none
Prinospio malmgreni	1	Rare	1.00	0.00	0.00	none
Processa sp.	1	Rare	-0.30	0.00	0.00	none
Protocirrinereis sp.	1	Rare	-3.60	0.00	0.00	none
Sicyonia sp.	1	Rare	-1.12	0.00	0.00	none
Sipunculos sp.	1	Rare	-0.12	0.00	0.00	none
Solen tehuelchus	1	Rare	1.23	0.00	0.00	none
sp.iophanes duplex	1	Rare	-0.38	0.00	0.00	none
Tagelus plebeius	1	Rare	-0.09	0.00	0.00	none
Tellina trinitatis	1	Rare	0.71	0.00	0.00	none
Trachycardium muricatum	1	Rare	0.34	0.00	0.00	none
Upogebia sp.	1	Rare	0.91	0.00	0.00	none

5 INTEGRATING SIMULATIONS AND EXPERIMENTATION TO PREDICT COMMUNITY RESPONSES TO DISTURBANCE AND IMMIGRATION REGIMES.

Danilo Cândido Vieira[1,2]; Gustavo Fonseca[2]

[1]Centro de Estudos do Mar - Universidade Federal do Paraná - Caixa Postal 50.002, 83255-000, Pontal do Paraná, PR, Brazil.

[2] Universidade Federal de São Paulo – Av. Dona Ana Costa, 95 – CEP 11060-001, Santos, SP, Brazil.

Corresponding author: vieiradc@yahoo.com.br

phone: +55 16 993 626 902

Key-words: simulation, microcosms, nematodes, pattern-oriented modeling, immigration, disturbance

5.1 RESUMO

Este artigo explora a habilidade de modelos-baseados em indivíduos em prever respostas de nematoides marinhos de vida livre a eventos de perturbação e imigração. Técnicas de parametrização inversa foram utilizadas para modelar a hipótese de que perturbações associadas a baixas taxas de imigração diminuem a abundância e a diversidade, enquanto perturbações associadas à alta imigração resultam em uma assembleia mais abundante, diversa e uniformemente distribuída. Experimentos foram conduzidos em microcosmos, onde cada réplica correspondeu a uma unidade experimental. As simulações incluíram cinco pressupostos de equivalência de espécies: ao longo de (1) todas as espécies; (2) um mesmo grupo trófico; (3) um mesmo grupo de grupos-traços; (4) um mesmo grupo de abundância e; (5) não equivalente como um todo. Contrariamente às expectativas, os resultados não mostraram diferenças entre o controle e o tratamento de baixa imigração, enquanto que altos níveis de imigração diminuíram significativamente tanto a abundância quanto a riqueza das espécies, enquanto aumentaram a equitabilidade. Simulações sugeriram equivalência ecológica entre espécies de nematoides cujas características de história de vida exibem similares adaptações para colonização e persistência no sistema. Sob esse pressuposto, as simulações dos tratamentos previram os efeitos da perturbação associados aos baixos níveis de imigração, mas falharam na previsão dos efeitos do alto nível de imigração. No entanto, após a recalibração da capacidade suporte (i.e. o parâmetro mais influente no espaço de parâmetros), as simulações foram capazes de prever ambos os tratamentos de microcosmos. Juntos, nossos resultados sugerem que as assembleias de nematoides permanecem próximas da capacidade suporte dos microcosmos. Em nossas simulações, este parâmetro pode ser interpretado como um fator que afeta o equilíbrio das assembleias. O aumento dos eventos de imigração pode alterar este estado de equilíbrio, aumentando a assimetria da competição.

5.2 ABSTRACT

The present article tests the ability of discrete-event simulations to predict the response of free-living marine nematodes to events of disturbance and immigration. We used inverse parameterization techniques to model the hypothesis that disturbance associated with low immigration rates decreases abundance and diversity, whereas disturbance associated with high immigration result in a more abundant, diverse and evenly distributed assemblage. Experiments were conducted in microcosms, where each replicate corresponds to an experimental unit. Simulations included five assumptions of species equivalence: across (1) all species; (2) a same feeding type; (3) a same trait group; (4) a same abundance group and; (5) non-equivalent at all. Contrary to expectations, the results showed no differences between control and low immigration treatment, whereas high levels of immigration decreased significantly both the abundance and richness of species, while increasing evenness. Simulations suggested ecological equivalence among nematode species whose life history traits exhibit similar adaptations for colonizing and persisting in the system. Under this assumption, simulations mimicking the treatments predicted the effects of disturbance associated with low immigration levels but failed in predicting the effects of high immigration level. However, after recalibrating the carrying capacity parameter (the most influential factor on the parameter space), simulations were able to predict both microcosms treatments. Together, our results suggested that nematodes assemblages remain near the carrying capacity of the microcosms. In our simulations, this parameter can be interpreted as a factor that affects the equilibrium of the assemblages. Increased immigration events may alter this equilibrium state by increasing the asymmetry of competition.

5.3 INTRODUCTION

Community ecologists probably face one of the most challenging systems to understand. The dynamic and amount of response and predictable variables are so high that precludes any simplistic analytical approach (Van Dyke Parunak et al. 1998). Controlled experiments are probably the most used approach to confront ecological theories. However, its contribution is usually conceptual and its integration with simulations modeling remains a challenge (Morgan 2005, Grimm and Railsback 2012). Discrete-event simulation and live experimentation represent two extreme techniques: simulation presents a controlled, repeatable idealized system, whereas live experimentation achieves realism, but surrender repeatability and the ability to monitor internal dynamics.

At one hand, pattern-oriented modeling (POM, *sensu* Grimm 2002) together with individual-based model (IBMs) provide a robust framework for designing empirical research that directly supports theory and model development (Grimm et al. 2017). For example, while POM uses patterns observed at multiple levels in the system to reach sufficiency in parameterization (Grimm 2002, Grimm and Railsback 2012), IBMs are especially useful for describing processes where many individuals can develop and interact deterministically and stochastically (Railsback 2008, Grimm et al. 2017). Deterministic processes impose species identity as a key element in determining ecological responses (selection-based models, Leibold 1995, Chase and Leibold. 2003), whereas stochastic events of birth, death, immigration, and extinction influence assemblage patterns independent of the species identity (neutral-based models, Hubbell 2001, Etienne and Alonso 2006).

On the other hand, laboratory experimental ecosystems (microcosms) increase the understanding of natural processes by simplifying the complexities of the natural environment (Fraser and Keddy 1997). Microcosms, in particular, allow isolating community assembly processes that could not be as readily studied in outdoor experimental ecosystems (Gallucci et al. 2015, Santos et al. 2018) because direct cause-and-effect relations are often confounded and difficult to isolate (La Point and Perry 1989). The importance of microcosm studies stems from their simplicity and manipulability, which allow the development of precise cause-and-effect relationships (Giesy and Allred 1985).

The present article tests the ability of deterministic and stochastic models to provide realistic predictions for experimentally manipulated communities. For this purpose, we designed a microcosm experiment in which communities were exposed to distinct disturbance/immigration regimes. Disturbance and immigration are key factors influencing the structure and composition of natural communities (Netto Paulo da Cunha 1994, Adler 2011). Disturbance is the major cause of local extinction and consequently, a potent deterministic driver of community assembly at local scale (Souza 1984, Netto Paulo da Cunha 1994). Immigration acts in the opposite direction, being responsible for the local supply of propagules, determining the recovery rate of a community after a perturbation and preventing the local extinction of species (Fukumori et al. 2015).The experiment tested the hypothesis that disturbance associated with low immigration rates decreases abundance and diversity, whereas disturbance associated with high immigration result in a more abundant, diverse and evenly distributed community.

The study was performed on free-living marine nematodes. Free-living marine nematodes are excellent candidates for testing modelling predictions: (1) they are characterized by a high local species richness and abundance (Heip et al. 1985, Coull 1999) at a scale which is amenable to controlled and adequately replicated laboratory microcosms experiments (tens of species and hundreds to thousands of individuals in patches as small as 10 cm^2); (2) entire nematode assemblages are easily manipulated under laboratory conditions (Balsamo et al. 2012, Gingold et al. 2013, Gallucci et al. 2015) and; (3) they exhibit a relatively rapid population dynamics (Heip et al. 1985), which allows relevant data on population dynamics to be collected over fairly short timescales.

5.4 METHODOLOGY

5.4.1 Microcosm experiment

On January 2017, surface sediments were collected intertidally at low tide in Araçá bay, São Sebastião, Brazil. The sediments were defaunated by successive freezing and thwarting (Schratzberger et al. 2000). Three days later, surface sediments were collected at the same location to obtain the fauna. Microcosms consisted of 500-

ml glass beakers filled with 100g of defaunated sediment and 100g of sediment containing fauna. All beckers were topped up with seawater and carefully homogenized. The microcosms were constantly aerated and covered by parafilm to prevent evaporation and salinity increase. The beckers were randomly assigned on the bench and maintained under dark conditions and 21◦C. Part of the sediment matrix containing the fauna ("source" sediment) was maintained in a plastic box (80 x 40 x 30cm) and held underwater in running filtered seawater during the whole experiment. At the end of the experiment, three random samples were taken from these sediments using a corer of 3 cm height and 5-deep.

The microcosm set-up consisted of eleven units, each considered as an independent experimental unit. To characterize the initial fauna of the microcosms, two units were sampled after 3 days of incubation (Ctl_0). The remaining of the microcosms ran for 27 days and included one control group and two treatments, each with three replicates. The control group did not include any disturbance or immigration event. The other two treatments included a low and high number of immigration events (every 6 and 2 days, respectively). Both treatments included disturbance events every four days.

Each immigration event consisted of adding to the microcosms about 100 specimens of nematodes extracted from the "source" sediment. Live extraction of species was performed through decantation technique preceded by anesthetization of the fauna by an isotonic $MgCl_2$ solution (75 g/L of for seawater of 35 PSU) (Somerfield and Warwick 2013). The disturbance events consisted of first homogenizing the microcosms and then replacing approximately 10g of sediment and associated fauna by azoic sediment. Nematodes in each aliquoted of sediment were counted under a stereomicroscope.

5.4.2 Sample processing

The sediment was washed through a 45um mesh sieve and the associated fauna extracted by flotation with Ludox TM 50 (specific density 1.18) (Heip et al. 1985). The retained material was stored in formaldehyde 4% and stained with Rose Bengal. Meiofauna was counted and identified under a stereomicroscope. Five percent of the nematodes per microcosm were randomly picked, evaporated slowly in anhydrous

glycerol and mounted on permanent slides for identification. Nematodes were identified to genus level (Warwick et al., 1998) and separated into morphospecies.

5.4.3 Microcosm data analyses

Univariate summary statistics for describing the experimental unities included: the total number of individuals (N), the total number of species (S), Margalef species richness (d'), Simpson diversity index (D), Simpson evenness (J'), relative dominance (C_{rel}) of the most abundant species and skewness ($R_{LogSkew}$, as a proxy for rarity). To characterize nematodes on the basis of their life history traits and functional groups, species were assigned to the 'colonizer – persister' ("*c-p"*) scale (from 1 to 5, Bongers and van de Haar 1990) and; to four to feeding types based on the morphology of the buccal cavity (Wieser 1953): 1A-selective deposit feeders, 1B non-selective deposit feeders, 2A- epigrowth feeders, and 2B-predators/omnivorous. The Index of Trophic Diversity was calculated as ITD=1-$\sum\Phi$, where Φ is the contribution of density of each trophic group to total nematode (Heip et al. 1985). The maturity index (MI) was determined as the weighted average of individual *c-p* values (Bongers and van de Haar 1990).

All univariate statistics were treated by one-way ANOVA followed by Tukey post hoc tests to access differences among control and treatments. Non-parametric multivariate analysis of variance (Anderson 2001) tested for significant treatments effects on the structure of assemblages (*adonis* function in "vegan" package, Oksanen et al. 2017). The analysis was done on Euclidean distances and further visualized by non-metric multidimensional scaling (nMDS).

5.4.4 Simulations

Simulations were performed using the function "*neut.simulate"* (R package "neutral.vp", Smith and Lundholm 2010). It consists of an individual-based model ruled by processes of birth and death and immigration. Our implementation of the model consisted of a single cell representing a homogeneous environment within which total population is regulated by carrying capacity K. In each time step, individuals give birth

to a single offspring and/or die with probabilities *b* and *d*. These parameters are further controlled by assigning species to a specific fitness value ranging from 1 to 10: the closer to ten is the fitness, the higher the ratio *b*/*d.* Penalties in birth and death rates are inversely related to the parameter *sel,* which ranges from 1 to 999: settling *sel*=1 decreases 10% of the base birth in species of fitness 10 and increases 10% of the base death in species of fitness 1. The equations used to calculate *b* and *d* are described in Smith and Lundholm (2010).

We performed simulations using the same number of species (S=42) and trait groups observed for "source sediment". Different equivalence assumptions were considered through five fitness constructions: F1- with equal values across all species; F2- with equal values across a same feeding type; F3- with equal values within a same *c-p* group; F4- with equal values within a same abundance group (log_2 geometric scale) and; F5- non-equivalent at all (unique fitness values across all species). All fitness constructions had the lower fitness values assigned for the species or group of species with the lowest abundance. All simulations were initialized with the species abundances averaged for the acclimation microcosm group and had the same temporal resolution of the laboratory experiment (27 cycles).

5.4.5 Simulating controls

We used inverse techniques of pattern-oriented modeling (Grimm 2005, Kramer-Schadt et al. 2007, Hartig et al. 2011) to find the parameters combination able to quantitatively predict responses of the microcosm control. We performed extensive simulations over the full factorial space of *b, d* (each varying from 0.01 to 1), K (varying from 1500 to 5000), *sel* (varying from 1 to 999) and fitness (*F1* to *F5*). A total of 2,600,599 simulations were done. Subsequently, a two-step filtering procedure was applied over all parameter combinations by accepting parameterizations only if they yield simultaneous agreement with the patterns observed for control microcosm:

1. In the first filtering, we selected the models matching nine observed univariate indexes: S, N, d', D, J, C_{rel}, $R_{LogSkew}$, MI, and ITD. For each summary statistic *i*, the threshold for the deviation measure *i* was defined as the 0.25 and 0.75 percentiles of the corresponding observed pattern. A redundancy analysis (Legendre et al. 2011) evaluated the

parameter space of the selected simulations. The predictor variables comprised base rates of birth, death, carrying capacity and fitness equivalence (expressed as $E = \sum \theta^2$, where θ is the density of each fitness level). Explanatory variables comprised all the nine summary statistics measured for the control microcosm group.

2. In the second filtering, we select simulations by comparing the multivariate structure of observed and simulated assemblages. For this purpose, non-metric multi-dimensional scaling ordination (nMDS) using the Euclidean distance (Legendre and Legendre 2012) was applied to species abundance data (observed and simulated). The filtering considered the 95% confidence ellipse around the observed data as criteria for selecting the models. The advantage of using the multivariate analysis as a second filter is that it accounts for simultaneous changes occurring in species abundance while preserving species identity (Legendre and Legendre 2012). Calibrating a model against multiple patterns at multiple levels increases the chance of capturing the internal organization of the real system (Wiegand et al. 2003).

5.4.6 Simulating treatments

Simulations mimicking the microcosms treatments consisted of a direct modification of the original model algorithm to allow addition and removal of species according to the temporal resolution of the laboratory experiment. Microcosms treatments were simulated using parameters selected in the two-step filtering described above. We simulate density-dependent immigration events by adding 100 individuals randomly sampled from a species pool *P*, which was determined from the species abundances averaged for the “source sediment”. Each disturbance event consisted of removing *N* individuals at random, with *N* determined by the counts of individuals removed from the microcosm set-up. Similarities in assemblage structure between simulated and observed treatments were assessed by nMDS of relative species abundance (Euclidean distance).

5.4.7 Model recalibration

Failure in simulating treatments would constitute a rejection of the model or a subsequent need for model recalibration. Lagarrigues et al. 2015 recently demonstrated that recalibrating the most influential parameters of a complex model in a Bayesian framework may help improve both accuracy and generality of model-based ecological predictions. It a recalibration context, we are primarily interested in tracking slight variations of parameters, rather than in obtaining a new accurate model calibration for the available data (Lagarrigues et al. 2015), especially as we had prior knowledge on the parameter space coming from the RDA analysis. In this study, we recalibrate the most influential parameter reveled by the RDA analysis, the other parameters being fixed at their values computed in the control parametrization.

5.5 RESULTS

5.5.1 Microcosms

Forty-two species of free-living marine nematodes were recorded in our study. Nematode densities varied from 1959 to 3500 individuals, with opportunist (cp=2) and persistent species (cp=5) representing 83% and 2% of the total abundance, respectively. The dominance was strong across all microcosms, with only two species (*Amphimonhystrella* sp.1 and *Dorylamoipsis* sp.1) contributing with more than 50% of the total abundance in each microcosm. All microcosms exhibited prevalence of non-selective deposit feeders (54% ± 8%), followed by selective deposit feeders (22% ± 5%), predators/omnivores (14% ± 5%) and epigrowth feeders (8% ± 4%).

One-way ANOVA showed significant effects of treatments on four summary statistics: total abundance (F=8. 14, p=0.02), number of species (F=22.2, p=<0.01), evenness (F=32.1, p<0.01) and percentage of predators/omnivores species (F=9.73, p=0.013). Tukey post hoc tests showed no differences between control and low immigration treatments (p<0.05). Both abundance and number of species decreased in the high immigration treatment, whereas evenness increased (Fig. 1a -c). The PERMANOVA analysis for the multivariate structure of nematode assemblages also showed the high immigration treatment significantly different from both control and low

immigration treatment (pseudo-F=1.91, p=0.01), which was evident from the nMDS plot (Fig 1d).

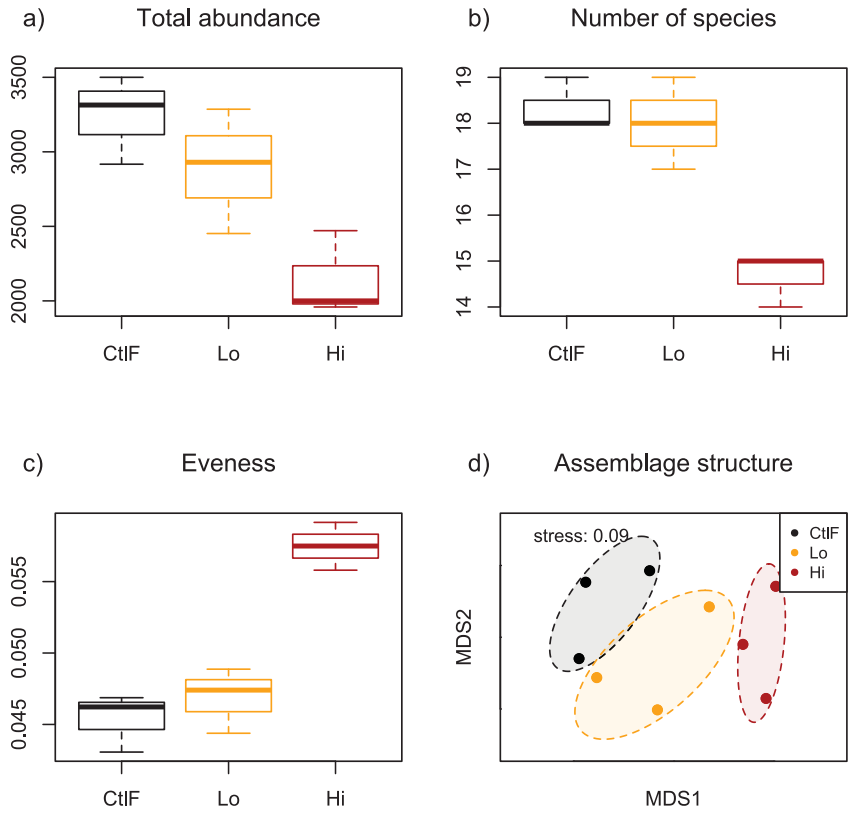


Fig. 1 – Effects of treatments on the microcosms. Box plot for (a) total abundance, (b) number of species, (c) Simpson evenness; (d) nMDS of nematodes species abundance data. CtlF – control; Lo – low immigration; Hi – high immigration.

5.5.2 Simulations

We analyzed over two million model simulations, with only 23 models matching the summary statistics observed for the control microcosm group. All the selected models had abundances ending at the settled carrying capacity K, which varied from 2500 to 3000. The selected simulations included all fitness constructions (F1 to F5) and had fitness equivalence E ranging from 0.1 to 1, with most of the models having E lower than 0.55 (21 models). The RDA analyses showed K and E as the most important parameters in determining the observed summary statistics (Fig. 2a). K was positively related to abundance and species richness and inversely related to the evenness index.

Ewas positively related to dominance and negatively related to diversity indices of Simpson and skewness.

Non-metric multidimensional analyses comparing simulated and observed assemblages selected two models (Fig 2b). These models (referred as $F3_{E053}$ and $F3_{E054}$) had the same K of 2500, similar fitness equivalence (0.53 and 0.54, respectively) and distinct base birth/death ratios (0.93 and 0.61, respectively). After using these models to simulate the low and high immigration treatments they produced very similar results (Fig 2c). The simulated community structure of both models differed from each other but were both within the 95% confidence ellipse calculated for the microcosm's treatment of low immigration.

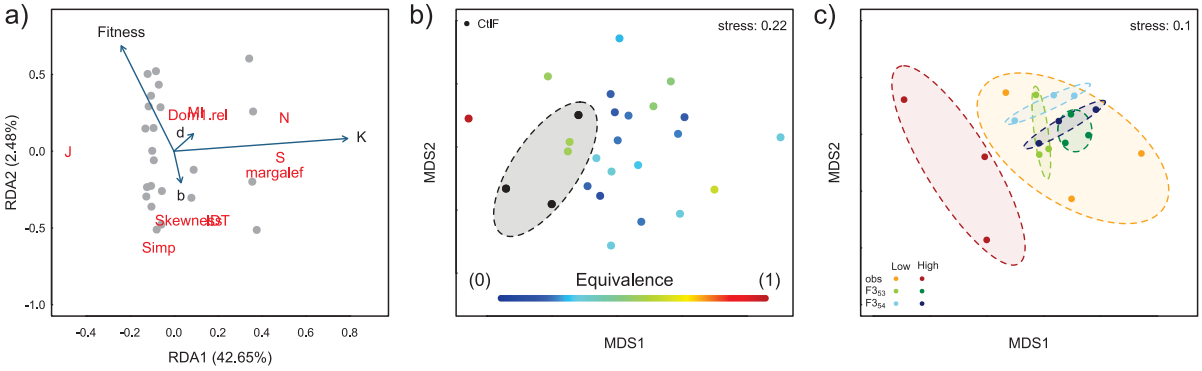


Fig. 2. (a) Redundancy analysis of the models selected by the univariate filters (control simulations). Vector length indicates the relative weight of a given variable, and direction indicates the correlation of that variable with each axis. (b-d) nMDS plots comparing the assemblage structure of (b) simulated and observed controls; (c) simulated and observed treatments.

5.5.3 Model recalibration

Given the importance of K towards indices that indicated responses of the microcosm's treatments (abundance, richness, and evenness), we recalibrate this parameter using the treatment data. Specifically, we adjusted K of the models $F3_{E053}$ and $F3_{E054}$ proportionally to the abundance observed in the microcosm treatments (K_{low} =2000 and K_{high} =1600). After the adjustment, the nMDS analysis showed that both $F3_{E053}$ and $F3_{E054}$ predicts the directional changes observed for microcosms treatments (Fig 3)

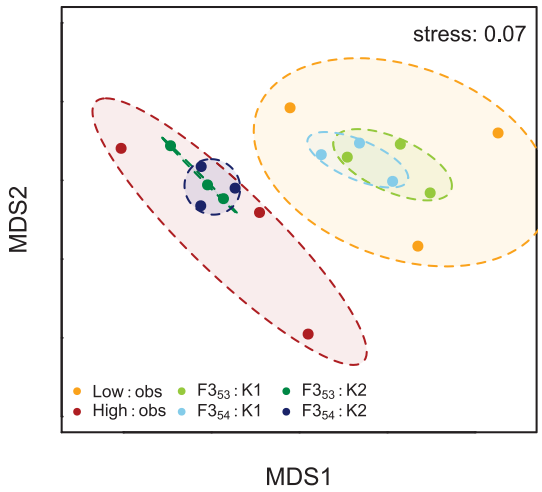


Fig 3. K-weighted simulations for treatments. "*Low:obs*" and "*High:obs*" are the microcosms treatments of low and high immigration, respectively. K1=2000, K2=1600.

5.6 DISCUSSION

The present article explored the ability of discrete-event simulations to predict the response of free-living marine nematodes to events of disturbance and immigration. Our simulations represented an exhaustive exploration of equivalence assumptions and parameter space, which was sufficient to generate competing models that closely resembled the observed patterns. Contrary to expectations, the microcosms experiment showed that high levels of immigration decrease both the abundance and richness of species while increasing the assemblage evenness. Control simulations suggested ecological equivalence among nematode species whose life history traits exhibit similar adaptations for colonizing and persisting in the system. Under the control parameters, the simulated treatments predicted the microcosm treatment of low dispersal but failed in predicting the high dispersal treatment.

Despite the wide range of parameters explored in this study (over two million of simulations), the two-step filtering procedure reduced dramatically the parameter space. The modeling process exploited the available biological data at all the steps: from the initial model construction to parameter estimation and prediction. The filter based on the univariate summary statistics did not account for an optimum' parameterization but to an entire 'cloud' of parameters whose included all fitness

assumptions. In the second filtering step, only two simulations matched the observed patterns, both predicting ecological equivalence among species within a same *c-p* group. This result corroborates the theory that equivalence across groups of traits can develop non-random patterns of co-distribution (neutral theory, Hubbell 2001). It means that nematode individuals within a same *c-p* group have similar competitive abilities and respond similarly to the same ecological drivers. A number of laboratory investigations have shown that colonizer-persister concept is a good indicator of disturbance acting on nematode community (Tahseen 2012, Wilson and Kakouli-Duarte 2013). However, in natural conditions, individual nematode genera within a cp-class may have different spatial niches resulting from small differences in life history traits including fecundity, tolerance to environmental stresses and interspecific competition for resources (Park et al. 2013). The present study highlights that under laboratory conditions these differences are too small, and individuals are exactly the same in their ecological properties such as the probability of giving birth and dying.

On the reduced parameter space, the simulated treatments did not predict the structured effects of disturbance associated with high immigration levels. After adjusting the most influential parameter (carrying capacity), both selected models ($F3_{E053}$ and $F3_{E054)}$ predicted the directional changes observed for microcosms treatments. These results suggest that nematodes assemblages remained near the carrying capacity and help to clarify the observed effects of microcosm treatments. Carrying capacity represents the community size that the resources of the environment maintain (‘carry’) without a tendency to either increase or decrease. Communities regulated by carrying capacity generally shows intense intraspecific competition (Begon et al. 2006). The more individuals arrive, the greater the variability in the use of resources by species, which increases the asymmetry of competition and consequently the assemblage evenness (Hobbs and Hanley 1990). When immigrants do not successfully reproduce, yet use valuable resources, they effectively reduce the carrying capacity and favor species able to use alternative resources (Hilderbrand 2003).

In this study, we showed that inverse modeling coupled with experimental data can provide valuable insights about processes driving manipulated communities. We have shown that careful recalibration processes may be required to improve the understanding of the model and generate more accurate predictions. The tight interplay between modeling process and experimental data help to explain unexpected

and contradictory findings and frame conclusions. In our study, carrying capacity was the key factor in predicting microcosms responses and can be seen as a factor that promotes stability for the assemblages, with regulatory properties classically characteristic of intraspecific competition (Begon et al. 2006).

6 GENERAL DISCUSSION

All studies gathered in this thesis share common approaches: inverse model designs, individual-based models (IBMs) and free-living nematodes as study case (Table 1). In the following sections, we summarize how these approaches allowed for (1) developing and validating new analytical methods and (2) confronting theory and data.

Table 1: Integrative framework of the studies presented in this thesis.

Integrating simulated data, field surveys and experimentally manipulated communities to predict marine ecological patterns				
Ph.D. thesis	Chapter 2	Chapter 3	Chapter 4	Chapter 5
Report	Methodological		Theoretical-empirical	
Design	Inverse models			
Model organism	Free-living nematodes from Araçá Bay			
Methodology	segRDA	IBM/pattern-oriented modelling		
Empirical system	Field survey			Microcosms
Explanatory model	Environmental gradient			Events of disturbance and immigration
information for modelling	community breakpoints	species-environment		species
Prediction	Community-level	Species-level		Community-level
Main remark	segRDA for modelling non-contiguous ecological responses	EB-NB framework for quantifying the importance of non-selection and selection processes	Nematodes have a larger non-selective component compared to macrofauna	Disturbance and high levels of immigration decreases the carrying capacity of local communities

6.1 NEW ANALYTICAL TOOLS FOR MODELING ECOLOGICAL COMMUNITIES.

Ordination is primarily a research tool for the interpretation of field data on communities and their environment (Ter Braak 1994). Particularly, redundancy analysis provides interesting statistics, such as the proportion of variance of the response data that is accounted for by the explanatory variables, and tests of significance of this statistic and of individual canonical eigenvalues (Makarenkov and Legendre 2002). The RDA is an extension of multiple regression analysis and thus it assumes a linear and constant trend of each species over the whole gradient (Legendre et al. 2011). However, there is no special reason why nature should linearly relate changes in species assemblages to changes in environmental variables. In **Chapter 2** we have shown that when there are two or more contrasting boundaries along a gradient and distinct community responses associated to it, the traditional RDA failed in detecting such discontinuities, returning a model with high unexplained variance.

To overcome difficulties imposed by the traditional RDA analysis, we proposed the **segRDA** package for modeling non-continuous responses of ecological communities. Following a inverse model design, segRDA uses information of the community at lower levels (i.e**. communities' breakpoints**) to break the original RDA model in pieces variable (i.e. the **pwRDA** analysis), thus minimizing the sum of squares of the differences between observed and predicted values of the response. Because IBMs simultaneously incorporate individual variations and environmental heterogeneity, they were fundamental for developing and validating the segRDA routine. Simulations results showed that pwRDA accounts for greater percentages of the total variance of the response variables than classical RDA. These results underscore the fact that constraining **hypotheses on parameters and processes by the observed patterns** increases the explanatory power of the model. However, if the transition between communities is **smooth**, the test of significance will point to the traditional RDA model as being the most appropriate; if the transition between communities is **sharp**, the test will point to the pwRDA model as the most appropriate one. It is important to note that, although the segRDA routine is a "inverse" approach, the pwRDA analysis itself is not: the "pieced" model consists of a set of equations, and the execution of the analysis consists of integrating them.

Chapter 3 also describes a new analytical method following a inverse model design: the EB-NB framework for exploring the relative importance of selection and dispersion processes in structuring ecological communities. This framework is based on a spatially-explicit individual-based model and therefore, compared to segRDA, **embody a more detailed description of the community and their environment for modeling**. The model resulted from the EB-NB framework (i.e. the composite model) explained the spatial distribution of most nematodes species along the Araçá bay. However, it is important to note that, pwRDA and EB-NB framework are intrinsically distinct from each other in terms of underlying assumptions about processes. A key difference between these two methods is that they derive **distinct prediction levels**: pwRDA and EB-NB framework make predictions at **community-level** and **species-level**, respectively.

An important aspect of the IBM implemented within the EB-NB framework is that it considers the state of each species at the measured environment (i.e, the niche position) to set the rules that determine the fate of every single individual. These rules are very simple and are based on parameters that govern the probability of taking various actions (e.g. birth, death, move away from the grid). In this manner, the IBMs mimic the element of chance in the natural world (Parrott and Kok 2000). A common criticism of IBMs is that it is very difficult to determine to what extent each rule in the model contributes to the dynamics that arise in different simulation scenarios and to assess whether these rules are based on valid assumptions (Van Dyke Parunak et al. 1998). The EB-NB framework surmounts these difficulties by using patterns displayed by the real system to validate the modeling process (**pattern-oriented modeling**). Therefore, this novel method enables to explore predictions from a wide range of scenarios (e.g. niche-based, neutral-based) that would be intractable through experiments or field observations alone.

6.2 CONFRONTING THEORY AND NEMATODE DATA.

Ecosystem surveys are well suited to verifying the applicability of models to complex natural systems (Srivastava et al. 2004). Chapters 2, 3 and 4 are based on the same field survey and model nematode response along the environmental gradient. In **Chapter 2**, we showed that **nematodes from Araçá bay have a non-continuous response to environmental data**, suggesting at least two contrasting

assemblages, with a transition zone between them. This result supports other recent investigation which defined two main habitats with distinct nematode assemblages at internal and external sublittoral area of the Araçá bay (Appendix A1).

Chapters 3 and 4 addressed the relative importance of non-selection (neutral-based) and selection (niche-based) assembly processes. **Chapter 3** showed that both mechanisms play a role in structuring nematodes assemblages, with **selection processes as the dominant structuring mechanism**. It means that species identity was an important factor to predict the spatial distribution of nematodes assemblages. The importance of environmental factors on the distribution of marine nematode species is well documented (Gallucci et al. 2009, Vieira and Fonseca 2013, Fonseca et al. 2014) although very little attention has been given to determine the niche dimension of nematode species. Our results also highlighted non-selection processes as an additional and important driver of nematodes assemblages, adding novelty to the understanding of assembly processes of free-living nematodes.

The relative importance of niche and neutral-based processes in explaining community structure may be related to both body size and life-history traits of organisms (Finlay 2002, Cottenie 2005, Beisner et al. 2006, Shurin et al. 2009). It has been suggested that environmental filtering is stronger in macrofauna, possibly because macrofauna are both less plastic in their fundamental niches and more selective in the settlement after dispersal (Gollner et al. 2015, McClain and Rex 2015, Ruiz-Abierno and Armenteros 2017). When the definition of the community is greatly expanded to incorporate species that differ vastly in body size and life history (as both Hubbell 2001 and Bell 2003 have done), it would seem essential to preserve those differences in an appropriate model. In this context, the EB-NB framework provides tools for constructing distinct composite models based on empirical niche measurements. In **Chapter 4,** we showed that the relative importance selection and non-selection processes indeed depends on the group of organisms under study: **nematodes have a larger non-selective component compared to macrofauna**. However, a significant part of macrofaunal species was not explained by any of our simulations. The most likely factors contributing to this low tractability are scale-effects and unmeasured environmental factors.

Compared to field surveys, artificial microcosms offer a greater control over experimental conditions and enable quick and precise experiments (i.e. have high tractability) (Srivastava et al. 2004). **Chapter 5** tested the ability of IBMs to predict the

response of free-living marine nematodes to events of disturbance and immigration. Simulations suggested **ecological equivalence** among nematode species whose **life history traits** exhibit similar adaptations for colonizing and persisting in the system. This result contrast with Chapters 3 and 4, where species identity was an important factor to predict the spatial distribution of the assemblages. The importance of carrying capacity for modeling natural and experimentally manipulated communities also showed contrasting behaviors. Carrying capacity represents the community size that the resources of the environment can just maintain without a tendency to either increase or decrease (Begon et al. 2006). In Chapter 2, trial simulations revealed little importance for carrying capacity for modeling the observed assemblage patterns. These results corroborate with a number of field investigations showing that nematode patterns are not constrained by resource limitation in enriched-muddy sediments (e.g. Admiraal et al. 1983, Montagna 1984, Radway 2008). In contrast, **Chapter 5** showed **carrying capacity as the key** factor in predicting the nematode responses and suggested that nematodes assemblages remain near the carrying capacity of the microcosms. Microcosms are by definition small in space in comparison to the system they are modeled after and therefore resource limitation is generally stronger under such conditions (Bengtsson et al. 2002).

It is accepted that the perceived importance of niche vs neutral processes is scale-dependent (Chase 2014). At small scales, the relative contribution of stochastic events to the overall structure of the community increases, and we perceive this system, which is highly niche-structured at larger scales, as largely neutrally structured at smaller scales (Chase 2014),

6.3 CONCLUSION

The results of this Ph.D. thesis corroborate that both deterministic and stochastic processes play an important role in assembling natural communities. The inverse modeling showed to be a powerful approach for dealing with local heterogeneity and/or complex community structure, thus enhancing our predictive power. Our efforts suggest that the tight interplay between modeling, field observations and experiments can lead to a better understanding of the processes governing community dynamics. To our knowledge, this thesis represents the first attempts to apply individual-based models in marine communities.

7 APPENDIX A1 – ADDITIONAL PUBLICATIONS

DEFINING SOFT BOTTOM HABITATS AND POTENTIAL INDICATOR SPECIES AS TOOLS FOR MONITORING COASTAL SYSTEMS: a CASE STUDY IN A SUBTROPICAL BAY

Helio Herminio Checon; Danilo Cândido Vieira[1,2]; Guilherme Nascimento Corte; Ediunetty C. P. M. Sousa; Gustavo Fonseca; Antônia Cecília Amaral Zacagnini

https://doi.org/10.1016/j.ocecoaman.2018.03.035

Corresponding author: hchecon@yahoo.com.br
phone: +55 16 993 626 902

Key-words: MacrofaunaMeiofaunaAraçá BayManagement

7.1 ABSTRACT

The definition of habitats and indicator species is a prerequisite for monitoring and conservation programs. Nonetheless, defining habitats in marine soft-bottom environments is challenging given their spatiotemporal dynamics and apparent homogeneity. The selection of indicator species is also complicated given the large number of occasional species usually presented in benthic communities. This study aims to elaborate a framework based on well-established analytical methodologies to identify soft-bottom habitats and select indicator species to support monitoring and conservation programs. The proposed framework consists of four steps: 1) perform a Redundancy Analysis (RDA) on the community data to identify the community structure response to environmental gradients; 2) conduct a kernel density analysis on the RDA biplot to determine the habitats; 3) use the indicator values analyses (IndVal) to select indicator species of each habitat; 4) run polynomial quantile regression analysis to find the optimum distribution of each indicator species. Such framework allows the determination of habitats based on the association of environmental and biological datasets, instead of relying solely on abiotic surrogates. As a case study, we used data of macro and meiofauna of a biodiverse coastal ecosystem in Southeast Brazil which is under anthropogenic pressure. Three main habitats were identified in the bay, and macro and meiofaunal assemblages were influenced by similar environmental variables. Nevertheless, macrofauna was more sensitive to changes in sediment composition, whereas meiofauna responded strongly to changes in total organic content and water depth. Macro- and meiofauna indicator taxa showed high specificity and fidelity values to each habitat, supporting their use in monitoring and conservation programs. The spatio-temporal organization of each habitat and the optimum distribution of each indicator species provide baseline knowledge to be used to monitor environmental changes in the study area and help in its conservation.

7.2 INTRODUCTION

All marine ecosystems are to some extent currently affected by human activities (Halpern et al. 2008), and it is estimated that almost half of the marine environment is already impacted by a combination of stressors such as ocean acidification, coastal hypoxia, and pollution (Gray et al., 2002; Defeo et al., 2009; Halpern et al., 2015). This unprecedented level of anthropogenic threats to marine systems has increased the need for biomonitoring and conservation programs (Crain et al., 2008; Halpern et al., 2008; Stelzenmüller et al., 2010).

The success of conservation efforts is highly dependent on the identification and protection of natural habitats which can act as biodiversity reservoirs and are important to ecosystem functioning and stability (Stevens and Connoly, 2004; Cogan et al., 2009). Once defined, a habitat can be used to plan monitoring programs. So far, the definition of habitats in marine benthic ecosystems usually relies on physical attributes and biogenic structures such as seagrass, rocky shores, and mussel beds (Banks and Skilleter, 2002; Seitz et al., 2014). Defining such habitats, however, is particularly challenging in highly dynamic and apparently homogeneous systems such as marine soft-bottoms (MacArthur et al., 2010).

The definition of habitats in marine soft-bottoms is usually linked to the less conspicuous variation in sediment properties (e.g., mud content, pebbles, and sorting coefficient) (Gray and Elliot, 2009). Such classification is normally done *a posteriori* to the data acquisition and based only on the environmental conditions. Nevertheless, the use of abiotic surrogates alone to map coastal habitats may generate unreliable results (Diaz et al., 2004; Stevens and Connolly, 2004). The organisms inhabiting the matrix of sediments exhibit complex interactions with the environmental characteristics and greatly influence the habitats conditions (McArthur et al., 2010). The consideration of such complex species-environment interaction is therefore crucial for properly delimiting a habitat (Diaz et al., 2004).

A complementary method to the habitat-based approach in conservation programs is the selection of indicator species (Carignan & Villard, 2002; De Cáceres et al., 2010; Siddig et al., 2016). These species show predictable responses to environmental conditions and can be used to assess the habitat conditions (Dufrêne and Legendre, 1997; Carignan and Villard, 2002; Niemi and McDonald, 2004, Siddig et al., 2016). An appropriate indicator species has to respond strongly to a particular

group of conditions, to which it will serve as an indicator (De Cáceres et al., 2010; Fonseca and Gallucci, 2016). The selection of indicator species for soft benthic communities is particularly important since the sedimentary habitat is dynamic, the number of species is high, and the identification of benthic biodiversity to species level is a major time-consuming activity (Warwick, 1993). On top of that, it is important to understand the optimum environmental conditions of each indicator species, in an attempt to facilitate monitoring programs (Anderson et al., 2008; Fonseca & Gallucci, 2016).

The aim of this study is to suggest a methodological framework to be used in monitoring and conservation programs of soft-bottom ecosystems. Using well established statistical analyses, we first identify potential habitats based on the responses of species assemblages to the environmental characteristics. Then, we select possible indicator species which can be used to assess future changes in each habitat. The congruence between the analyses is then used to delineate each habitat of soft-bottoms. We apply this methodological framework to the main groups of marine benthic fauna (meio- and macrofauna) in a biodiverse benthic ecosystem which is under recent threats due to the planned expansion of the neighboring port (Amaral et al., 2010; 2016). These threats reinforce the critical need to recognize and understand the local environmental dynamics to monitor and manage the area. We use both assemblages to comprehend the benthic environment better and to test the framework on the two groups most used in benthic monitoring studies (Semprucci & Balsamo, 2012; Bessa et al., 2014; Gorman et al., 2017). The study area is also a typical example of many vulnerable parts of the Brazilian coast, and as such, the outcomes of the present study are relevant for other regions.

7.3 MATERIALS AND METHODS

7.3.1 Study area

This work was done at Araçá Bay (23º 49'S, 45º 24'W), a coastal ecosystem (~500.000 m^2) located in the central area of the São Sebastião Channel, state of São Paulo, Southeast Brazil (Fig. 1). The area is environmentally heterogeneous, with many distinct features such as patches of different sedimentary textures, mangroves and rocky shores (Amaral et al., 2016;

Checon et al., 2017). The intertidal area has a gentle slope, with a maximum depth of 5 m, while further the bay reaches 30 m deep towards the channel. Araçá Bay is located within the Marine Environmental Protection Area of the Northern Litoral (APA Marinha do Litoral Norte), a conservation unit which aims to preserve biodiversity and natural processes and is recognized as one of the areas with the highest marine biodiversity on the Brazilian Coast (Amaral et al., 2010; 2016).

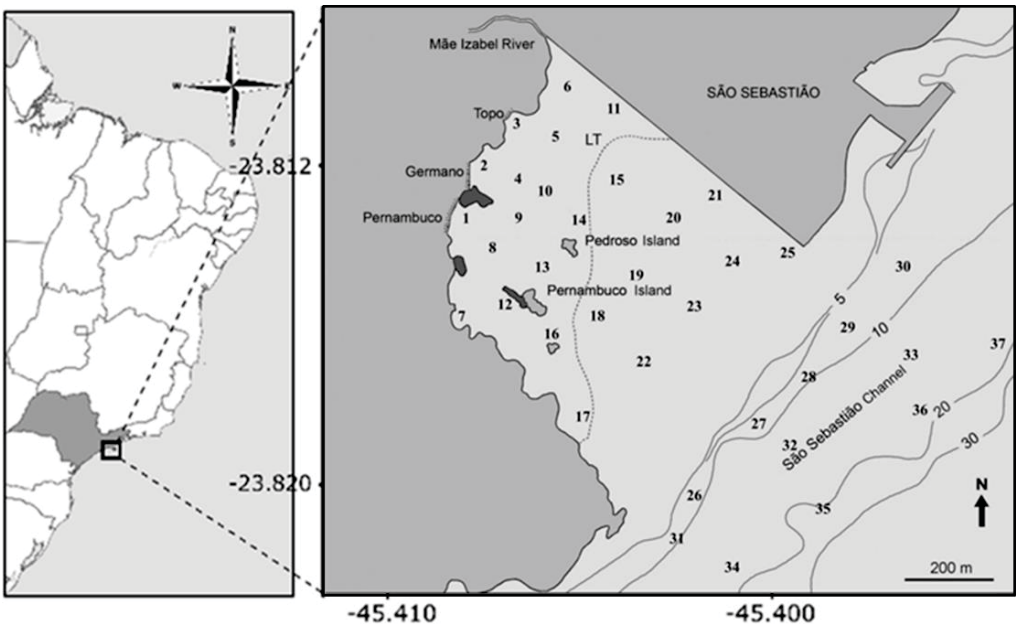


Fig. 1 - Study area. Map of the study area showing the 37 sampling stations along the intertidal (1 to 17) and internal sublittoral (18 to 25) at Araçá Bay, and external sublittoral at the São Sebastião Channel (26 to 37).

7.3.2 Sampling design

Sampling was performed during four periods (October 2012, February, June and September of 2013). Thirty-seven sampling stations were determined from the intertidal and shallow sublittoral area at the bay (< 5 m deep) to a depth of 25 m (São Sebastião Channel). Sampling stations were positioned to a) encompass habitat diversity (i.e., different sediment types and depths), and b) achieve a reasonable dispersion and spatial coverage (Fig. 1). The same locations (+/- 1 m) were sampled during each period using a GPS for orientation of sampling stations positions. Sampling was done manually at the intertidal and shallow sublittoral (< 3 m deep), and with the use of a multi-corer sampler for deeper sites. At each sampling

site, four samples were collected using a corer of 10 cm diameter and 20 cm depth (0.03 m²) for the evaluation of macrofauna, and one sample of 2.5 cm in diameter and 5 cm depth (19.6 cm²) for meiofauna. Particularly for macrofauna, the total area sampled in each sampling station is smaller than the 0.1 m² commonly employed in sublittoral studies (e.g. Petersen, box-corer, vanVeen) (Eleftheriou & Moore 2013), which may increase the effect of patchiness and sampling heterogeneity. However, we chose to collect all samples with a corer to obtain fully quantitative replicates, as well as to standardize sampling areas between the intertidal and sublittoral zones.

Additional samples were taken at each station to evaluate environmental parameters: Five samples of the top 1 cm of the sediment were taken using a corer measuring 2 cm in diameter to evaluate microphytobenthic biomass; and one sample of sediment was taken for granulometric analysis using a corer of 3 cm diameter and 20 cm depth.

7.3.3 Sample processing

Macrofauna samples were stored in plastic bags and posteriorly sieved with a 0.3 mm mesh. The fauna retained was sorted in taxonomic groups and fixed in 70% ethanol. All individuals were identified to the species level.

Meiofauna samples were immediately fixed in 4% formaldehyde, and posteriorly washed through a 45 µM mesh sieve and extracted by flotation with Ludox TM 50 (specific density 1.18) (Heip et al., 1985). The retained material was stored in formaldehyde 4% and stained with Rose bengal. Meiofauna counting and identification was done under a stereomicroscope. We selected only the nematode assemblage for further study, as they were the most abundant in the area. Nematodes were identified to genus level and further separated into morphospecies. From each sample, a total of 100 nematodes were randomly chosen, evaporated slowly in anhydrous glycerol and mounted on permanent slides for identification.

Microphytobenthic biomass was estimated from phaeopigments and chlorophyll *a* concentration according to Plante-Cuny (1973). Margalef pigment diversity index (Margalef, 1967), a ratio of total green pigments, was calculated. The index ranges from 2 to 8, increasing from young microphytobenthic communities to mature, oligotrophic ones. The granulometric analysis was carried out using the routine

sieving and pipetting techniques described by Suguio (1973) and sediment parameters were obtained using SysGran software, version 3.0 (Camargo, 2006) following the classifications of Folk & Ward (1957). Total organic carbon was evaluated using a modified Walkley-Black titration method, described by Gaudette et al. (1974).

7.3.4 Data analysis

We combined established statistical techniques to set up a framework to be used in studies that aim to characterize and define benthic habitats, and provide information to be applied in monitoring programs. The framework proposes a stepwise procedure, as shown in Fig. 2. Seven sites were excluded from analysis due to missing variables, resulting on a total of 141 sites.

First, we (i) performed a redundancy analysis (RDA) to evaluate the influence of environmental variables on community structure; then, (ii) we used a two-dimensional kernel density analysis to generate a contour map from the density distribution of points in RDA space and determine the habitats (i.e., areas with similar environmental characteristics and species composition). Due to the different sampling protocols, these analyses were performed separately for meiofauna and macrofauna. Given that marine meiofauna can reach extremely high densities, only a random fraction of the nematode assemblages was identified. The total number of nematodes of each species per sample was calculated by multiplying the total number of individuals per sample and the proportion of each morphospecies. For macrofauna, all individuals were identified. Both macrofaunal and nematode data were transformed using Hellinger function to minimize the importance of rare species (Legendre and Gallagher, 2001). Environmental data was checked for correlation and multicollinearity, using Spearman correlation and Variance Inflation Factor (VIF), respectively. The final model included 10 environmental variables with low collinearity: chlorophyll a, Margalef pigment diversity index, depth, total organic carbon, mean grain size, pebbles, coarse sands (as sum of very coarse, coarse and medium sands), fine sands, very fine sands and sorting coefficient. The

environmental gradient associated with each identified habitat was determined by checking the environmental variables ordination scores towards the respective clustering of sites.

Indicator species were determined using Indicator Values (IndVal) (Dufrêne and Legendre, 1997). This is a widely used tool to identify indicator species, which uses the species exclusive occurrence (i.e. specificity) and distribution in the sampling sites at a particular habitat (i.e. fidelity). From the species with high IndVal, we selected the ones with highest values (specificity and fidelity) for their particular habitat (in order to represent indicator species for every habitat) for further evaluation (mapping and relationship with environmental variables). Relations between the indicator species and correlated environmental variables were assessed using a non-linear quantile regression approach. Variables were selected based on their correlation with the habitats. Following the method proposed by Anderson (2008), abundances were plotted against the environmental variable of interest and quantile regression spline models were constructed for the 95th percentile (i.e., the value below which 95% of the abundances are expected to fall). The degree of the polynomial used on each model was calculated using the Akaike Information Criteria (AICc, Burnham and Anderson, 2002). Polynomials of degree 2, 3, 4 and 5 were created for each taxon with the best-fit model having the smallest AICc value (Anderson, 2008; Koenker, 2011). For each model, the value predicting the maximum abundance for a given environmental factor was taken as a measure of the estimated optimum. These values were subject to 999 sample pair bootstrapping and re-modelled from the original chosen model using bias-corrected percentiles. Confidence intervals of ninety-five percent were obtained from the distribution of bootstrapped sample pairs (Anderson, 2008).

Smoothed maps were generated to illustrate the distribution of species and habitats throughout the sampling periods. Maps were made using the inverse distance weighting (IDW) method. Data was interpolated by inverse distance weighting power equal to 2. The distribution of species and their representative habitats were contrasted to check for concordance in spatial and temporal distribution.

All analyses were done in R Software 3.3.1 (R Development Team, 2016), using the packages vegan (Oksanen et al., 2013), mixtools (Benaglia et al., 2009), MASS (Venables and Ripley, 2002), gstat (Pebesma, 2004), raster (Hijmans, 2016), maptools (Bivand and Lewin-Koh, 2017), sp (Pebesma and Bivand, 2005) and quantreg (Koenker, 2016).

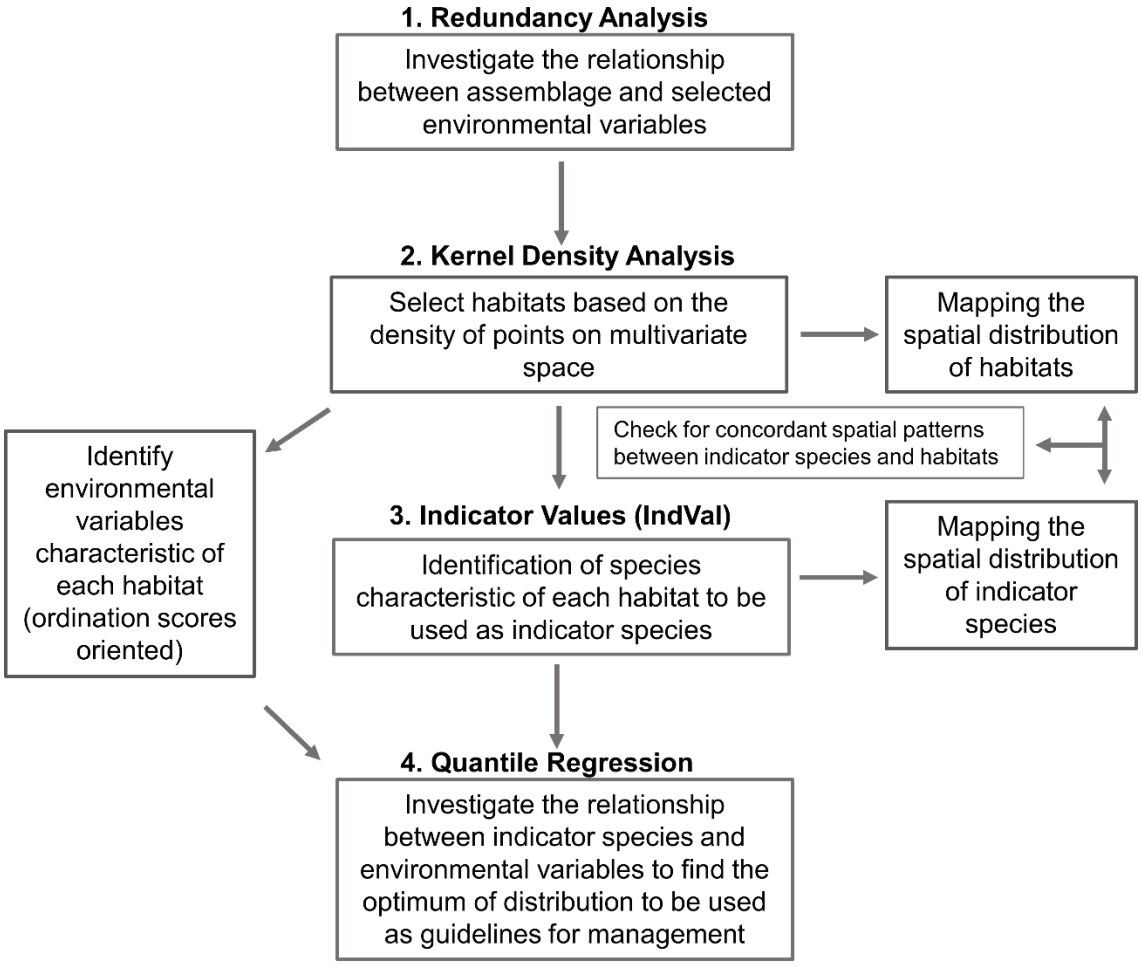


Fig. 2. Methodological framework used to define habitats and indicator species for monitoring purposes.

7.4 RESULTS

7.4.1 Defining soft bottom habitats

A total of 11270 individuals of 158 macrobenthic species, and an estimation of 16296 individuals of 195 nematode morphospecies were sampled. Macrobenthic assemblages were dominated by polychaetes, both in terms of abundance and species richness. The tanaidacean *Monokalliapseudes schubarti* (Mañé-Gárzon, 1949) was also very abundant. *Comesoma* sp.1 and *Dorylaimopsis* sp.1 were the most abundant species of nematodes.

Redundancy analysis models showed that the environmental variables were responsible for 22 % and 32% of the variation in species distribution of macrofauna (Model, $F_{10,130}$ = 3.723, $p < 0.001$) and nematode (Model, $F_{10,130}$ = 6.105, $p < 0.001$) assemblages, respectively. With the exception of pebble percentages, all individual

variables significantly explained macro and nematode distribution. Depth, coarse sands, chlorophyll a and very fine sands were, in order of importance, the most important variables for both groups. A summary of the values obtained for each variable is presented in Table 1.

Table 1. Summary of environmental characterization of Araçá Bay in each campaign. Total organic carbon (TOC) and sedimentary fractions are given as percentages. Sorting coefficient and mean grain size are given on phi (ϕ) scale.

	Campaigns				Range
	1st	2nd	3rd	4th	
Clorophyll a	82.6 ± 46.4	79.27 ± 54.9	75.6 ± 55.7	96.09 ± 72.1	10.1 - 264.2
Phaeopigments	91.82 ± 77.3	87.16 ± 79.5	79.08 ± 75.8	121.13 ± 132.9	2.4 - 515.8
Margalef Index	3.09 ± 0.45	3.43 ± 0.8	3.42 ± 0.8	3.16 ± 0.7	2.4 - 5.8
Depth (m)	-4.13 ± 7.7	-4.09 ± 7.5	-3.35 ± 6.7	-3.66 ± 6.8	0.8 - (-23.2)
TOC	0.8 ± 0.8	0.75 ± 0.6	0.81 ± 0.5	0.71 ± 0.7	0 -2.71
Peebles	2.47 ±7.7	1.62 ± 3.6	3.67 ± 7.3	1.52 ± 3.5	0 - 44.5
Very Coarse Sand	2.41 ± 4.6	2.6 ± 5.6	3.08 ± 4.8	2.41 ± 6.9	0 - 37.3
Coarse Sand	3.43 ± 8.57	3.36 ± 6.3	3.73 ± 5.9	3.33 ± 8.8	0 - 47.79
Medium Sand	5.2 ± 7.52	5.22 ± 7.9	10.5 ± 14.9	7.35 ± 15.4	0 - 62.88
Fine Sand	11.4 ± 13.35	8.95 ± 9.2	16.9 ± 15.4	10.3 ± 14.3	0.32 - 56.62
Very Fine Sand	46.39 ± 25.3	50.62 ± 26.3	56.7 ± 26.9	45.85 ± 26.4	3.52 - 99.15
Silt/Clay	28.59 ± 27.1	27.5 ± 28.2	5.30 ± 6.6	29.16 ± 28.7	0 - 84.95
Sorting	1.43 ± 0.7	1.21 ± 0.6	1.78 ± 0.6	1.51 ± 0.6	0.27 - 3.43
Mean Grain Size	3.87 ± 1.5	3.57 ± 1.1	3.77 ± 1.2	3.97 ± 1.7	(-0.09) - 6.85

Based on the ordination results, three habitats types were recognized for macrofauna and two habitat types for meiofauna (Fig. 2). Habitats were recognized using 0.5 density as cut-criteria. However, for macrofauna, a secondary step was made using a 1.0 cut-criteria to separate the two clusters of sites located at the opposing sides of the ordination axis (H2 and H3) (Fig. 3a, 3c). For macrofauna, Habitat 1 was characterized by few sites mostly restricted to the upper intertidal area of the bay and with a high contribution of coarse sands. Habitat 2 was characterized by most of the sites in the shallower area, with higher primary production and predominance of fine and very fine sand fractions. Lastly, Habitat 3 was characterized mainly by sites located on the external sublittoral area, from 5 m depth to the deeper areas of the channel, with higher percentages of silt/clay (positively related to water depth) and total organic carbon content. Transition sites are found between Habitats 2 and 3 (Fig 4). Nematode

were also structured in habitats 2 and 3, with Habitat 2 extending up to the intertidal zone. The samples characterized as Habitat 1 for macrofauna did not appear as a separate Habitat for nematodes (Fig. 5). For both groups, few sites were not categorized in any habitat and were considered as “no habitat” sites (Figs. 4, 5).

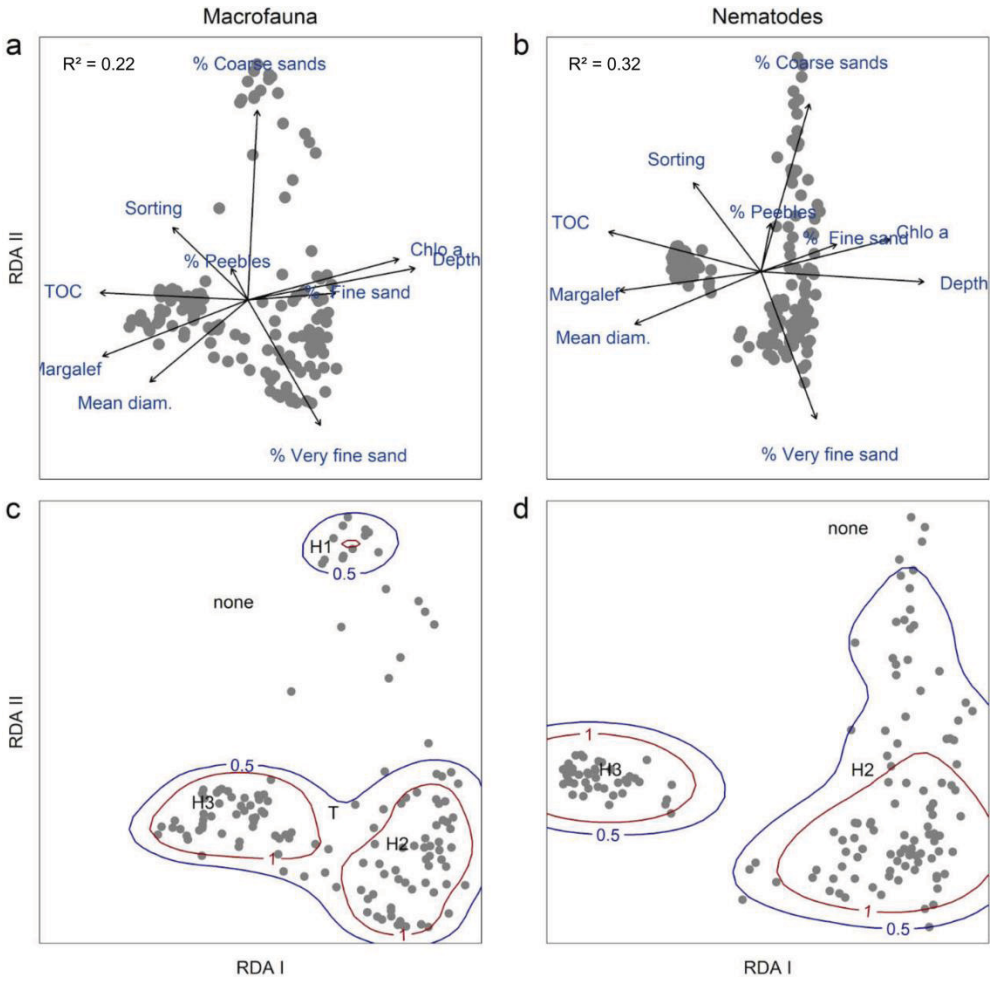


Fig. 3. Redundancy analysis and kernel density analysis. (a, b) RDA results for macrofauna and nematodes**; (c, d)** Kernel density plots showing the three main macrofaunal habitats (H1, H2 and H3) and the two main nematodes habitats (H2 and H3) . “T”: Transition between groups. “none”: non-grouped sites

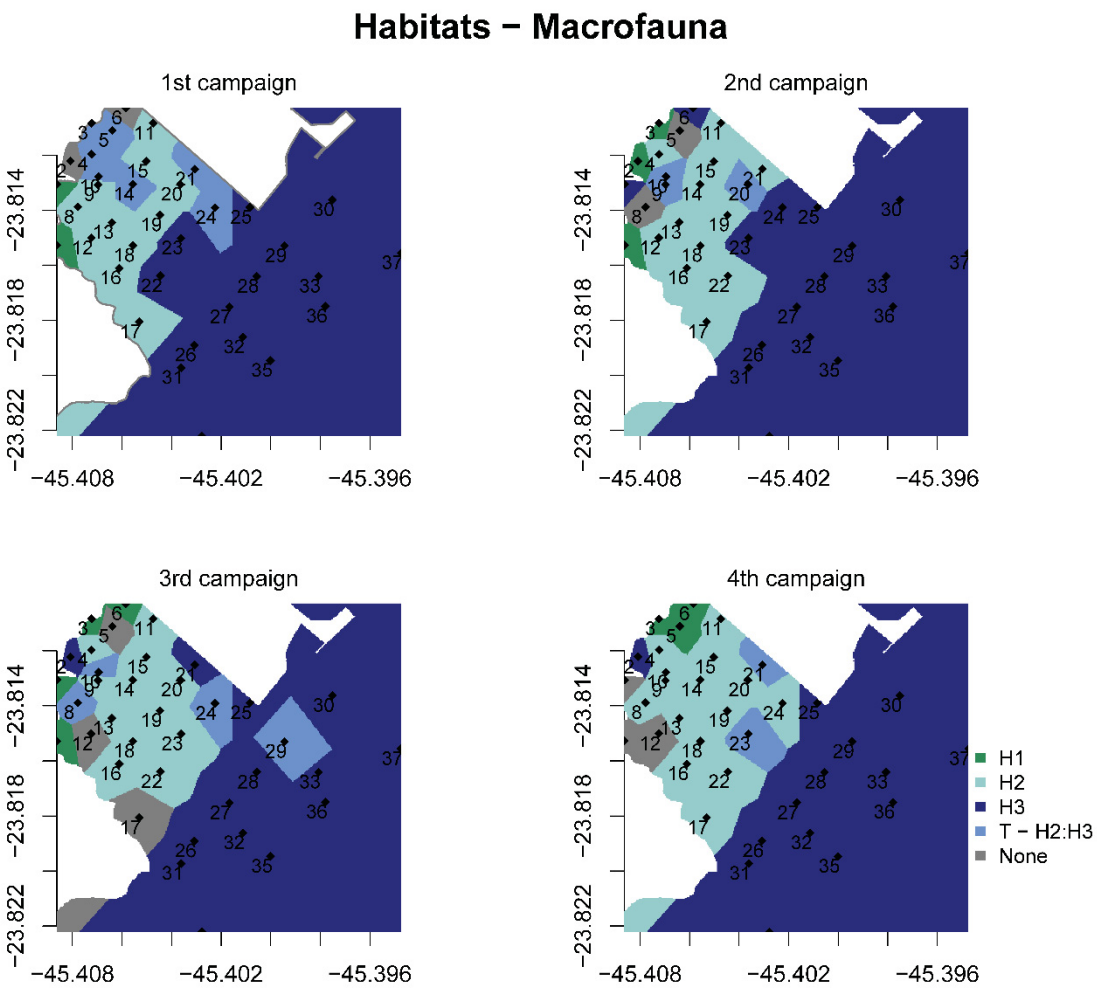


Fig. 4. Macrofauna habitats. Spatial interpolation maps showing distribution of the three macrofaunal habitats (H1, H2, and H3) found at Araçá Bay during each sampling campaign. T: sites with intermediate characteristics between H2 and H3. None: unclassified sites.

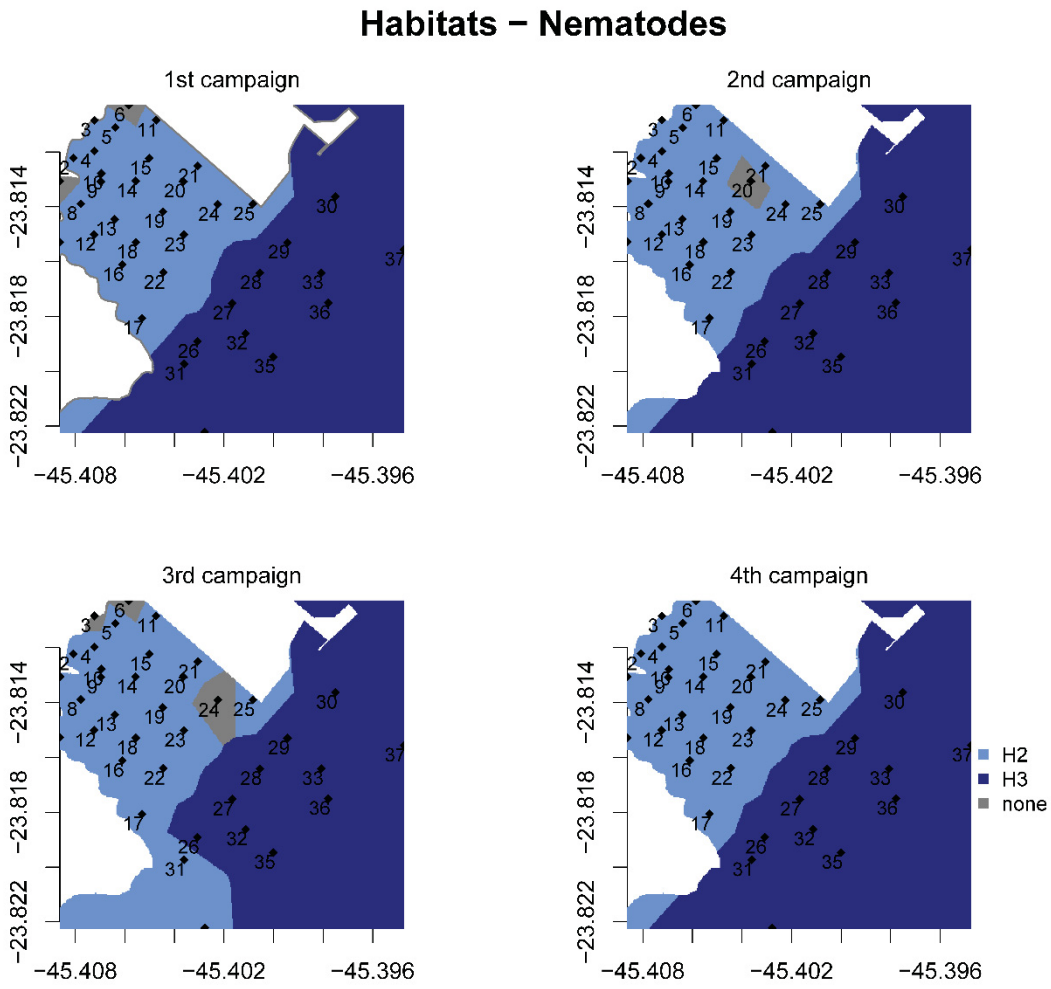


Fig. 5. Nematodes habitats. Spatial interpolation maps showing distribution of the two nematodes habitats (H2 and H3) found at Araçá Bay during each sampling campaign. None: unclassified sites.

7.4.2 Indicator species

For each habitat, IndVal identified species with high specificity and fidelity (Table 2). For macrofaunal species, Habitat 1 was represented by the polychaetes *Capitella nonatoi* (Silva and Amaral, 2017) and *Laeonereis culveri* (Webster, 1879), whereas Habitat 2 was represented by the gastropod *Olivella minuta* (Link, 1807) and the tanaidacean *Monokalliapseudes schubarti*. For Habitat 3, the cnidaria *Protankyra benedeni* (Östergen, 1898) and the polychaete *Neanthes bruaca* (Lana and Sovierzovski, 1987) had a high specificity, but low fidelity, meaning that they are found predominantly at Habitat 3, but did not have a broad distribution within it. For meiofaunal species, *Comesoma* sp. 1, *Viscosia* sp. 1 and to a lesser degree *Subsphaerolaimus* sp. 1 were strongly associated with Habitat 2. At Habitat 3,

Terschellingia sp. 5, *Sabatiera* sp. 1 and *Aponema* sp. 1 were found with very high specificity and fidelity.

Table 2. Indicator values (IndVal) of species with highest proportion of specificity (A) and fidelity (B) at each habitat. p-values are given by permutational analysis (n=999).

	IndVal Components		
Macrofauna	A	B	P-Value
Habitat 1			
Capitella nonatoi	0.99	1.00	<0.001
Laeonereis culveri	0.99	0.83	<0.001
Capitella neoaciculata	0.99	0.75	<0.001
Habitat 2			
Monokaliapseudes schubarti	0.97	0.66	<0.001
Olivella minuta	0.96	0.64	<0.001
Scoloplos (Leodamas) sp. A	1.00	0.60	<0.001
Habitat 3			
Neanthes bruaca	0.94	0.24	0.02
Protankyra benedeni	1.00	0.18	0.02
Pinnixa sayana	0.95	0.12	0.03
Nematodes			
Habitat 2			
Comesoma sp. 1	0.94	0.83	<0.001
Viscosia sp. 1	0.98	0.72	<0.001
Subsphaerolaimus sp. 1	0.98	0.67	<0.001
Habitat 3			
Terschellingia sp. 5	0.97	1	<0.001
Sabatieria sp. 1	0.99	0.97	<0.001
Aponema sp. 1	0.99	0.9	<0.001

The macrofaunal species *C. nonatoi, O. minuta*, *M. schubartii*, *N. bruaca* and the nematodes *Terschellingia* sp.5, *Sabatiera* sp. 1, *Comesoma* sp. 1 and *Viscosia* sp. 1 were chosen as indicator species for further analysis, based on their habitat-specific IndVal values. Species mapping showed that the indicator species distribution matches those found for their respective habitats. *Capitella nonatoi* distribution was concordant with spatial distribution and temporal fluctuation of Habitat 1, found only in the upper intertidal. *Olivella minuta* and *M. schubarti* spatial distribution was stable among sampling periods, occupying a large area of Habitat 2. However, their fidelity was lower

than the specificity due to variation on the spatial distribution of these species, with *O. minuta* occupying the inner sublittoral and *M. schubartii* the intertidal area. In Habitat 3, *N. bruaca* had a high specificity but low fidelity, due to the strong spatiotemporal variation, being restricted to a few stations during the 2nd and 4th campaign (Fig. 6). For meiofaunal species, *Terschellingia* sp. 5 and to a lesser degree *Sabatiera* sp. 1, occurred with strong spatial concordance and little temporal variation at Habitat 3. *Comesoma* sp. 1 and *Viscosia* sp. 1 were well represented at Habitat 2 sites, but with different spatial distribution, the former found at the southern and the latter at the northern portion of the intertidal area of the bay. Temporal variation was observed, and distribution of both species was more restricted during the 3rd and 4th campaign, respectively (Fig. 7).

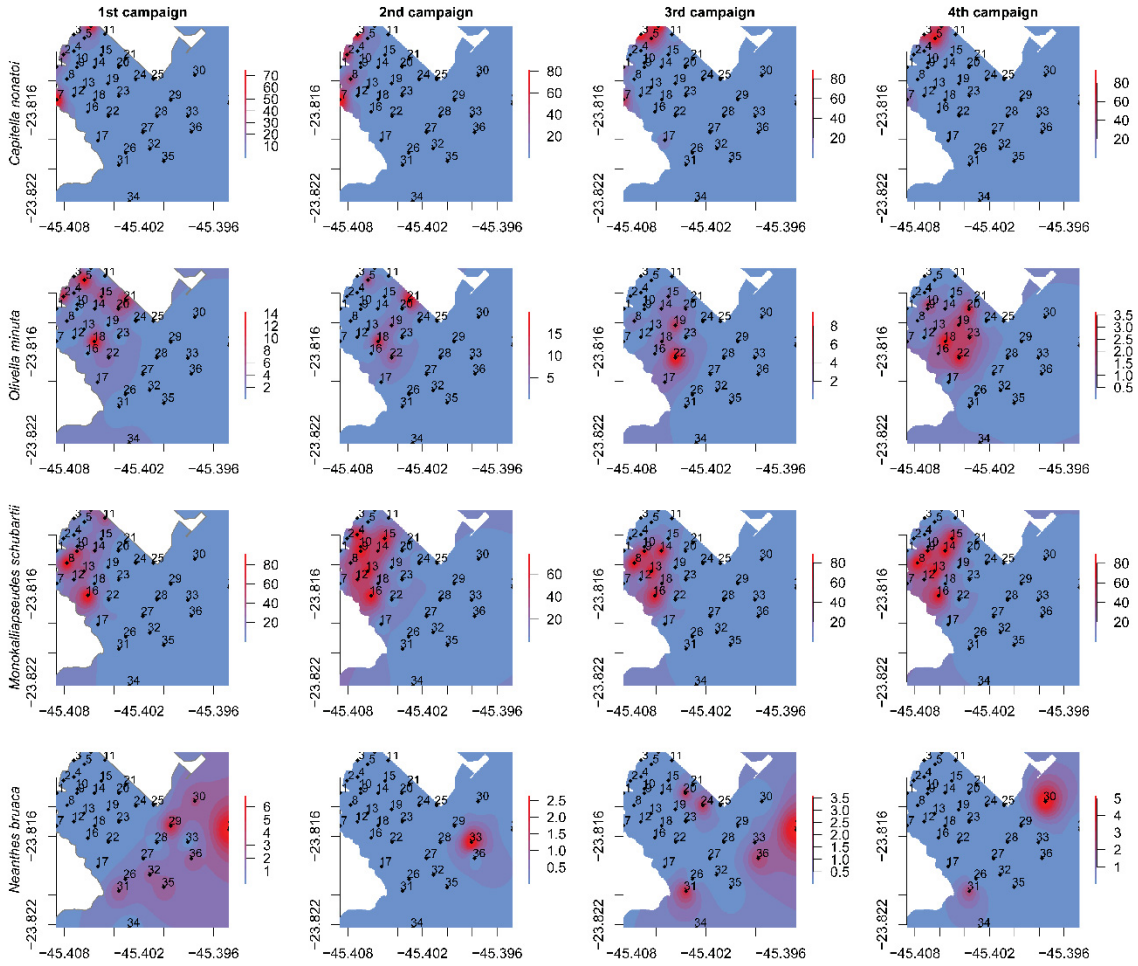


Fig. 6. Macrofauna indicator species. Spatio-temporal distribution of the habitat indicator species at Araçá Bay.

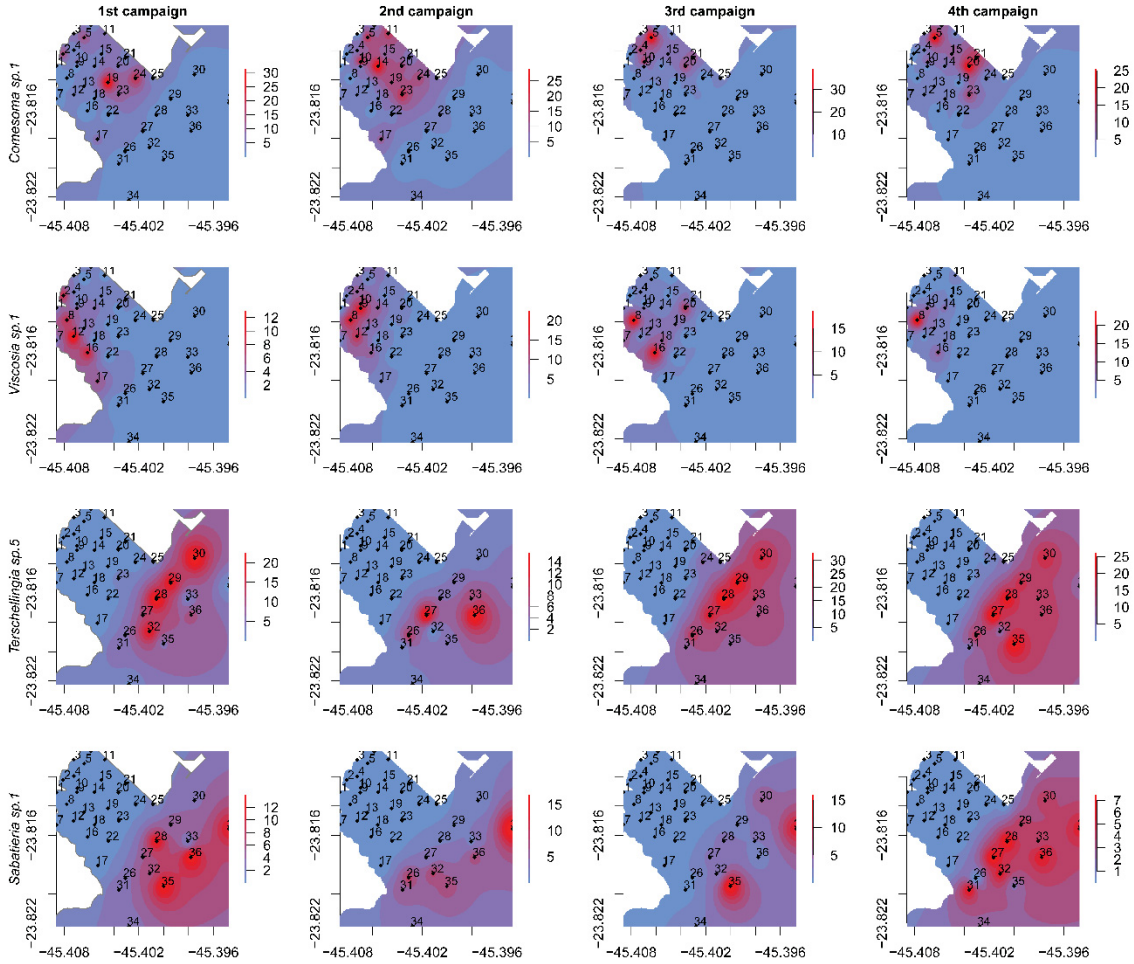


Fig. 7. Nematodes indicator species. Spatio-temporal distribution of the habitat indicator species at Araçá Bay.

Three environmental variables were selected to investigate their relations with the respective indicator species distributions (i.e. coarse sands; chlorophyll *a* and total organic carbon). The selection of these variables was based on the ordination scores and the common application in benthic studies, which can ease their applicability in monitoring programs. Most species showed a unimodal relationship with the most characteristic environmental variables of each habitat, but *Capitella nonatoi*, *Monokalliapseudes schubartii* and *Neanthes bruaca* exhibited a bi-modal relationship. For all the macrofauna species and *Viscosia* sp. 1 the estimated optimal interval with the selected environmental variable were larger than the range sampled in the bay, with high uncertainty around the optimum. *Comesoma* sp. 1, *Terschellingia* sp. 5 and *Sabatieria* sp. 1 showed a unimodal pattern with its optimum within the observed range. (Fig. 8).

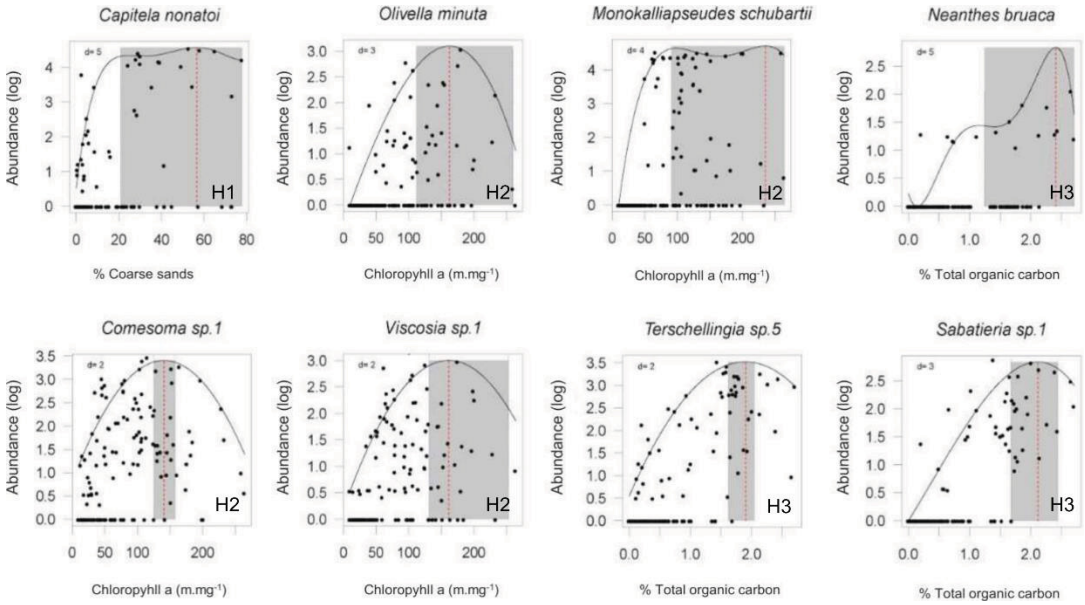


Fig. 8. Quantile regressions. Relationship between indicator species abundance (log y + 1) and environmental variables characteristic of each habitat. The red line represents the estimated optimum value for the species, and the gray area illustrates their 95% confidence interval. d: polynomial degree. Corresponding habitat for each species is shown on the lower right corner.

7.5 DISCUSSION

7.5.1 The analytical framework

Given the natural dynamics and apparent homogeneity of sedimentary systems, benthic habitats are not promptly recognized in the field. As such, the habitat boundaries, and related environmental variables have to be established *a posteriori* of the sampling campaign. In this scenario, the use of RDA combined with kernel density analysis permits one to infer whether the samples are structured in distinct clusters (i.e. habitats). We used RDA because it is based on the animal-environment relationship, whereas other methods commonly used in benthic protocols, such as PCA and cluster analysis (Verfaille et al., 2009; Brown et al., 2011), consider only abiotic surrogates, an approach not recommended to define and map benthic habitats (Diaz et al., 2004; Stevens & Connoly, 2004). Since kernel analysis is based on the principle of mapping the concentration of dots in a plot, this framework is more effective, but not restricted to, when a large number of samples are collected.

Once the habitats are identified, they can be used as categories for the IndVal analysis. If the analysis selects potential indicator species, i.e., species with high fidelity and specificity, it is an indication corroborating the presence of the habitat and that the indicator taxa can be used to monitor environmental changes (Dufrêne & Legendre, 1997). Finally, the use of

quantile regression is important to determine the optimum value and range of a species to a specific environmental variable. This information can be used to identify and predict environmental changes according to variations in species distribution and abundance (Anderson, 2008; Keeley et al., 2012). Density variations of an indicator species under present optimum conditions may indicate that the environmental features are changing towards non-optimum conditions. Such changes can be attributed to anthropic disturbance; however, it may also be caused by natural fluctuations or biotic interactions. Thus, in order to identify the source of such changes, it is necessary to monitor the environment and anthropic activities at the area.

7.5.2 Detecting the main habitats in the Araçá Bay

Three main soft-bottom habitats were found at Araçá Bay. Habitat 1 is located in the upper intertidal area and environmentally defined by a high percentage of coarse sands, which is not usually characteristic of tide-dominated environments such as the Araçá Bay. The high percentage of coarse sediments found here is likely related to sediment resuspension, intrusion due to past anthropic activities (e.g., dredging, disposal from nearby roads and harbor construction and expansion, Amaral et al., 2010; Mani-Peres et al., 2016), and the influence of weathering process on the cobbles and boulders of nearby rocky shores. The internal area of the bay is largely homogeneous forming a single habitat (Habitat 2). This shallow sublittoral area is characterized by a gradual decrease in water depth, a higher content of microphytobenthic pigments and fine and very fine sands, in comparison to the other two habitats. This sedimentary property is characteristic for tide dominated sandflats (Dyer et al., 2000; Le Hir et al., 2000). Finally, at the external sublittoral area, Habitat 3 is mainly environmentally characterized by high silt/clay content (positively correlated to TOC and increasing water depth) and total organic carbon. This high mud and organic carbon content is expected due to the hydro and morphological characteristics of the continental side of the São Sebastião channel, such as weak currents, low channel depth and natural inputs (i.e. riverine) of anthropogenic sources (i.e domestic and industrial sewage inputs) (Barcellos and Furtado, 2006, Alcántara-Carrió et al., this issue). The predominance of silt fractions is a result of the low availability of sand transport in the area (Alcantará-Carrió et al., 2017).

These three main habitats showed relatively stable spatiotemporal distribution throughout the sampling campaigns. Habitat 1, despite temporal fluctuations, was always restricted to the upper intertidal area. The consideration of temporal patterns is an important

feature for management of coastal areas, especially considering the dynamic nature of these ecosystems (Paiva, 2001; Arkema et al., 2006), and the relative stability on the distribution of habitats confers reliability to the monitoring of the area over different temporal scales.

Water depth and sedimentary composition were the most important environmental variables explaining the distribution of macro- and meiofaunal (nematodes) benthic assemblages at Araçá Bay. The importance of these variables in structuring marine benthic communities is well-known and has been highlighted in previous studies (e.g. Flach et al., 2002; Verfaillie et al., 2009; Corte et al., 2017). The habitats were also similar to the sedimentary facies described for the bay (Alcántara-Carrió et al., this issue), further reinforcing the importance of sedimentary features on benthic distribution. The variation explained by ordination analysis, however, was not high. Low percentages of variance explained when investigating whole communities are common and have been observed in various systems (e.g., Provete et al., 2014; Rodil et al., 2017), probably due to the great number of variables (species) analyzed. It can also arise as a result of unmeasured environmental variables (e.g. hydrodynamic conditions, nutrient proxies, contaminants), as well as local biotic interactions and seasonality (Quillien et al., 2015; Rodil et al., 2017). Additionally, the bay is composed by three distinct habitats, where species are responding differently to the environmental condition at each habitat. For instance, while at habitat 1 species are predominantly structured by the percentages of coarse sediments, at habitat 3 the percentage of organic matter is the structuring factor. The combination of multiple habitats within the same RDA analysis also reduces the power of explanation.

Habitat configuration was similar for macro and meiofaunal assemblages, suggesting a similar influence of these assemblages to environmental features (Corte et al., 2017). The presence of coarse sands at the upper intertidal area, however, creates a habitat for macrobenthic but not for nematode assemblages. Although coarse sand fraction has been shown to also modify species composition and increase nematode species richness (Semprucci et al., 2010; Vanaverbeke et al., 2011; Patricio et al., 2012); our results show that this effect may not be as strong as the one found for the macrofauna. As suggested, macrofauna might be more affected by spatial variability of habitat characteristics than nematode species (Semprucci et al., 2013). It is worth noting, however, that nematodes are found to have a higher variability than macrofauna at smaller scales (within < 0.1 km) (Schartzberger et al., 2008; Semprucci et al. 2010), and thus studies at different scales may found contrasting results.

7.5.3 Selecting indicator species of soft bottom communities

Indicator species analysis were able to identify many species to be used as indicators of habitat conditions (Supplementary Material), but we chose the most representative of each habitat for further modelling. Among those taxa, we selected four macrofaunal (*Capitella nonatoi*, *Olivella minuta*, *Monokalliapseudes schubarti* and *Nenathes bruaca*) and four meiofaunal (*Comesoma* sp. 1, *Viscosia* sp. 1, *Terschellingia* sp. 5 and *Sabatiera* sp. 1). Meiofaunal (nematodes) taxa, however, were identified only as morphospecies because of the still limited knowledge regarding nematode taxonomy and the lack of specialists in Brazil (Fonseca et al., 2014; Fonseca et al., 2017). Consequently, their proper use as indicator taxa is limited and depends on further taxonomic studies to aid identification procedures. Nonetheless, our results show that nematode taxa can be used as indicator taxa with the applied framework.

Given the framework used to establish indicator species, these taxa can be used to monitor changes in the conditions associated with each habitat. It is important, however, to emphasize that they are not direct indicator of disturbance or of any other anthropogenic interference within this area. Some of these species are indeed associated with opportunistic behavior. For instance, *Capitella nonatoi* is part of the *Capitella capitata* complex (Silva et al., 2017), a group of species kwown as *r*-strategists and early colonizers of disturbed environments (Tsutsumi, 1987) usually associated to organically enriched sediments (Pearson & Rosenberg, 1978; Rivero et al., 2005). *Monokalliapseudes schubarti* is also suggested as being an opportunistic species due to its high fecundity and fast growth (Leite et al., 2003). In regards to nematodes, there is a more limited knowledge regarding indicator species (Kennedy and Jacoby, 1999; Semprucci & Balsamo, 2012). Nonetheless, *Terschellingia* species have been suggested as being tolerant to stressful conditions, especially hypoxia (Armenteros et al., 2010; Boufahja et al., 2016), and are commonly found in areas with organic enrichment processes (Moreno et al., 2008; Armenteros et al., 2010). However, we cannot tease apart whether they occur naturally in the bay or are already responding to past perturbations in the area (e.g. dredging, sewage outfall) (Mani Peres et al., 2016; Amaral et al., 2016). Chemical data, for instance, do not suggest a significant ongoing organic and inorganic contamination at the area (Appendix A1, Kim et al., this issue). Thus, these species cannot be used as proxies of biodiversity, which is a common application of indicator species, but only to monitor environmental characteristics (Bustos-Baez and Frid, 2003; Siddig et al. 2016).

All selected indicator taxa showed high specificity, however, fidelity was higher for meiofaunal taxa, aside from *Capitella nonatoi* at Habitat 1. The low fidelity of macrofauna may be attributed to varied reasons. At Habitat 2, the two indicator species exhibited a distinct, yet

complementary, spatial pattern. *Monokalliapseudes schubarti* was more present in the intertidal area, likely due to the higher content of very fine sand at the internal sublittoral, known to restrict their distribution (Leite et al., 2003). On the other hand, *O. minuta* was found mainly in the internal sublittoral and lower levels of intertidal probably because its lower desiccation tolerance (Arruda and Amaral, 2003). This complementary distribution shows that both species should be considered in order to monitor and understand changes in Habitat 2. In Habitat 3, *Neanthes bruaca* had a strong spatio-temporal variation, which reflected in low fidelity of macrofauna. This result suggests that its use as an indicator in Habitat 3 is limited, as their absence in the habitat cannot confidently indicate changes. These situations show that the cause of low fidelity should be investigated in order to verify the suitability of an indicator taxon. Nonetheless, all macrofaunal indicator species had high specificity, and thus their presence in other areas can indicate modification. The selected indicator species were also persistent in the landscape and showed a relative temporal stability, which is a desirable feature of an indicator species (Hilty and Merenlender, 2000; Siddig et al., 2016).

The precision of the estimated optimum was better for meiofaunal taxa than macrofauna, except for *O. minuta*. This precision can be seen as a measure of reliability between changes in environmental variables and species response (Anderson, 2008). The lower precision found for macrofaunal taxa may be due to unmeasured variables playing a role on the species distribution (Anderson, 2008) or low abundances (Keeley et al., 2012), as is the case for *N. bruaca*. Macrofauna species also had the upper boundary of the optimum interval outside of the range of environmental variation, aside from exhibiting bimodality (except for *O. minuta*), which makes the variation in abundances less straightforward to interpret (Anderson 2008). This wide range indicates that these species may occur naturally beyond the obtained range of environmental variability. Even so, in such instances, the lower boundary of optimum can be used to predict values where the abundance of indicator species may start to decrease.

7.6 CONCLUSION

The applied framework was able to identify benthic habitats, select indicator species, as well as to determine their optimum relationship with environmental variables. The results presented here have implications for the management of the Araçá Bay and can be further used in other areas. Particularly for the Araçá Bay, we now understand the habitat dynamics and the most appropriate indicator species of each habitat. These species can now be used in monitoring

programs as indicators of possible environmental changes, especially considering the expected anthropic interference at the area. The advantage of the framework applied here is that, differently from commonly used methods, the habitat is identified not only by abiotic features, but by the relationship between species and environment. The proposed approach gives a complete assessment of habitat dynamics and environmental conditions, including the identification of potential indicator species, and therefore providing critical information necessary in conservation and management efforts.

7.7 ACKNOWLEDGEMENTS

We are thankful to Dr. Paulo Paiva, Dr. Ronaldo Christofoletti, Dr. Paulo Gomes Sumida, Dr. Andre Garraffoni and the anonymous reviewers for providing comments on the manuscript. We also acknowledge the Instituto de Biologia da Universidade Estadual de Campinas, Instituto Oceanográfico da Universidade de São Paulo and Centro de Biologia Marinha da Universidade de São Paulo for providing logistic support. Special thanks to AS Godoy, RK Murayama, CF Silva, D Gomes, RF Daolio, N Padovanni for providing support with field sampling; and Y Shah Esmaeili for language review. Funding: This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) through the BIOTA-Araçá project (2011/50317-5) and scholarship grants to GNC (2016/10810-8); Coordenação de Amparo ao Pessoal de Ensino Superior (Capes) for scholarship grants to HHC and DCV; Conselho Nacional de Desenvolvimento Científico (CNPq) through a productivity grant to ACZA (306534/2015-0).

References

Alcántara-Carrió, J., Mahiques, M.M., Goya, S.C., Fontán-Bouzas, A., *In press*. Sedimentary dynamics of a subtropical tidal flat sheltered inside a coastal channel (Araçá Bay, SE, Brazil). Ocean & Coastal Management. doi.org/10.1016/j.ocecoaman.2017.11.011

Alcántara-Carrió, J., Sasaki, D.K., Mahiques, M.M., Taborda, R., Souza, L.A.P., 2017. Sedimentary constraints on the development of a narrow deep strait (São Sebastião Channel, SE Brazil). Geo-Marine Letters: 1-14.

Amaral, A.C.Z., Migotto, A.E., Turra, A., Schaeffer-Novelli, Y., 2010. Araçá: biodiversidade, impactos e ameaças. Biota Neotropica 10: 219-262.

Amaral A.C.Z., Turra A., Ciotti A.M., Wongtschowski C.L.D.B.R., Schaeffer-Novelli Y., 2016. Life in Araçá Bay: diversity and importance. 3. ed. Lume, 100 pp. Available: www.bibliotecadigital.unicamp.br/document/?code=73819&opt=1.

Anderson, M.J., 2008. Animal-sediment relationships re-visited: Characterising species' distribution along and environmental gradient using canonical analysis and quantile regressions. Journal of Experimental Marine Biology and Ecology 366: 16-27.

Armenteros, M., Pérez-García, J.A., Ruiz-Abierno, A., Diaz-Asencio, L., Helgera, Y., Vincx, M., Decraemer, W., 2010. Effects of organic enrichment on nematode assemblages on a microcosm experiment. Marine Environmental Research 70: 374-382.

Arkema, K.K., Abramson, S.C., Dewsbury, B.M., 2006. Marine ecosystem-based management: from characterization to implementation. Frontiers in Ecology and the Environment 4: 525-532.

Arruda, E.P., Amaral, A.C.Z., 2003. Spatial distribution of mollusks in the intertidal zone of sheltered beaches in southeastern of Brazil. Revista Brasileira de Zoologia 20: 291-300.

Banks, S.A., Skilleter, G.A., 2002. Mapping intertidal habitats and an evaluation of their conservation status in Queensland, Australia. Ocean & Coastal Management 45: 485-509.

Barcellos, R.L., Furtado, V.V., 2006. Organic matter contents and modern sedimentation at São Sebastião Channel, São Paulo. Journal of Coastal Research SI 39: 1073-1077

Benaglia, T., Chauveau, D., Hunter, D.R., Young, D., 2009. mixtools: an R package for analyzing finite mixture models. Journal of Statistical Software 32: 1-29.

Bessa, F., Gonçalves, S.C., Franco, J.N., Andre, J.N., Cunha, P.P., Marques, J.C., 2014. Temporal changes in macrofauna as response indicator to potential human pressures on sandy beaches. Ecological Indicators 41: 49-57.

Bilyard, G.R., 1987. The value of benthic infauna in marine pollution monitoring studies. Marine Pollution Bulletin 18: 581-585.

Bivand, R., Lewin-Koh, N., 2017. maptools: tools for reading and handling spatial objects. Available at https://CRAN.R-project.org/package=maptools

Boufahja, F., Semprucci, F., Beyrem, H., 2016. An experimental protocol to select nematode species from an entire community using progressive sedimentary enrichment. Ecological Indicators 60: 292-309.

Brown, C.K., Smith, S.J., Lawton, P., Anderson, J.T., 2011. Benthic habitat mapping: a review of progress towards improving understanding of the spatial ecology of the seafloor using acoustic techniques. Estuarine, Coastal and Shelf Science 92: 502-520.

Burnham, K.P., Anderson, M.J., 2002. Model selection and multi-model inference: a practical information-theoretical approach. Springer-Verlag New York, 488 pp.

Bustos-Baez, S., Frid, C. 2003. Using single indicator species to assess the state of macrobenthic communities. Hydrobiologia 80: 1-11.

Camargo, M.G., 2006. Sysgran: um sistema de código aberto para análises granulométricas do sedimento. Revista Brasileira de Geociências 36: 371−378.

Carignan, V., Villard, M.A., 2002. Selecting indicator species to monitor ecological integrity: a review. Environmental Monitoring and Assessment 78: 45-61.

Checon, H.H., Corte, G.N., Muniz, P., Brauko, K.M., Di Domenico, M.. Bícego, M.C., Siegle, E., Figueira, R.C.L., Amaral, A.C.Z., 2018. Unrevaling the performance of the benthic index AMBI in a subtropical bay: the effects of data transformation and exclusion of low-reliability sites. Marine Pollution Bulletin 126: 438-448.

Checon, H.H., Corte, G.N., Silva, C.F., Schaeffer-Novelli, Y., Amaral, A.C.Z. 2017. Mangrove vegetation decreases density but does not affect species richness and trophic structure of intertidal polychaete assemblages. Hydrobiologia 795: 169- 179.

Cogan, C.B., Todd, B.J., Lawton, P., Noji, T.T., 2009. The role of marine habitat mapping in ecosystem-based management. ICES Journal of Marine Science 66: 2033-2042.

Corte, G.N., Checon, H.H., Vieira, D.C., Fonseca, G., Amaral, A.C.Z, 2017. Cross-taxon congruence in benthic communities: searching for surrogates in marine sediments. Ecological Indicators 78: 173-182.

Crain, C.M., Kroker, K.J., Halpern, B.S., 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters 11:1304–1315

De Cáceres, M., Legendre, P., Moretti, M., 2010. Improving indicator species analysis by combining group of sites. Oikos 119: 1674-1684.

De Cáceres, M., Legendre, P., Wiser, S.K., Brotons, L., 2012. Using species combinations in indicator values analysis. Methods in Ecology and Evoution 3: 973-982.

Defeo, O., McLachlan, A., Schoeman, D.S., Schlacher, T.A., Dugan, J., Jones, A., Lastra, M., Scapini, F., 2009. Threats to sandy beach ecosystems: a review. Estuarine, Coastal and Shelf Science: 1-12.

Diaz, R.J., Solan, M., Valente, R.M., 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. Journal of Environmental Management 73: 165-181.

Dufrêne, M., Legendre, P., 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs 67: 345-366.

Dyer, K.R., Christie, M.C., Wright, E.W., 2000. The classification of intertidal mudflats. Continental Shelf Research 20: 1039-1060.

Eleftheriou, A., Moore, D.C., 2013. Macrofauna techniques. In: Eleftheriou, A. (Ed) Methods for the study of marine benthos. 4th edition. Wiley-Blackwell, p 175-251.

Flach, E., Muthumbi, A., Heip, C., 2002. Meiofauna and macrofaunal community structure in relation to sediment composition at the Iberian margin compared to the Goban Spur (NE Atlantic). Progress in Oceanography

Folk, R.L., Ward, W.C., 1957. Brazos River bar: a study in the significance of grain size parameters. Journal of Sedimentary Petrology 27: 3-26.

Fonseca, G., Fontaneto, D., Di Domenico, M., 2017. Addressing biodiversity shortfalls in meiofauna. Journal of Experimental Marine Biology and Ecology: 1-13.

Fonseca, G., Gallucci, F., 2016. The need of hypothesis-driven designs and conceptual models in impact assessment studies: An example from the free-living marine nematodes. Ecological Indicators 71: 79-86.

Fonseca, G., Maria, T.F., Kandratavicius, N., Venekey, V., Gheller, P.F., Gallucci, F. 2014. Testing for nematode-granulometry relationships. Marine Biodiversity 44: 435-443.

Fonseca, G., Norenburg, J., Di Domenico, M., 2014. Diversity of marine meiofauna on the coast of Brazil. Marine Biodiversity 44: 459-462.

Gaudette, H.E., Flight, W.R., Toner, L., Folger, D.W., 1974. An inexpensive titration method for the determination of organic carbon in recent sediments. Journal of Sedimentary Research 44: 249-253.

Gorman, D., Corte, G.N., Checon, H.H., Amaral, A.C.Z., Turra, A., 2017. Optimizing coastal and marine spatial planning through the use of high-resolution benthic sensitivity models. Ecological Indicators 82: 23-31.

Gray, J.S., 2002. Species richness of marine soft-sediments. Marine Ecology Progress Series 244: 285-297.

Gray, J.S., Elliot, M., 2009. Ecology of Marine Sediments. 2nd Edition. Oxford University Press, 256 pp.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., Fujita, R., Heinemann, D., Lenihan, H.S., Madin, E.M.P., Perry, M.T., Selig, E.R., Spalding, M., Steneck, R., Watson, R., 2008. A global map of human impact on marine ecosystems. Science 319: 948–952.

Halpern, B.S., Frazier, M., Potapenko, J., Casey, K.S., Koenig, K., Longo, C., Lowndes, J.S., Rockwood, R.C., Selig, E.R., Selkoe, K.S., Walbridge, S., 2015. Spatial and temporal changes on the cumulative human impacts on the world's oceans. Nature Communications 6: 7615.

Heip, C., Vincx, M., Vranken, G., 1985. The ecology of marine nematodes. Ocenography and Marine Biology Annual Review 23: 399-489

Hijmans, R.J., 2016. raster: geographic data analysis and modelling. Available at https://CRAN.R-project.org/package=raster

Hilliard, J., Hajduk, M., Schulze, A., 2016. Species delineation in the Capitella species complex (Annelida: Capitellidae): geographic and genetic variation in the northern Gulf of Mexico. Invertebrate Biology 135: 415-422.

Hilty, J., Merenlender, A., 2000. Faunal indicator taxa selection for monitoring ecosystem health. Biological Conservation 92: 185-197.

Hyland, J., Balthis, L., Karakassis, I., Magni, P., Petrov, A., Shine, J., Vestergaard, O., Warwick, R., 2005. Organic carbon content of sediments as an indicator of stress on marine benthos. Marine Ecology Progress Series 295: 91-103.

Keeley, N.B., Macleod, C.K., Forrest, B.M., 2012. Combining best professional judgement and quantile regression splines to improve characterization of macrofaunal responses to enrichment. Ecological Indicators 12: 154-166.

Kennedy, A.D., Jacoby, C.A., 1999. Biological indicators of marine environmental health: meiofauna – a neglected benthic component? Environmental Monitoring and Assessment 54: 47-68.

Kim, B.S.M., Bícego, M.C., Taniguchi, S., Siegle, E., Oliveira, R., Alcántara-Carrió, J., Figueira, R.C.L., *In press.* Organic and inorganic contamination in sediments from Araçá Bay, São Sebastião, Brazil. Ocean and Coastal Management. doi.org/10.1016/j.ocecoaman.2017.12.028

Koenker, R., 2011. Additive models for quantile regressions: model selection and confidence bandaids. Brazilian Journal of Probability and Statistics 25: 239-262.

Koenker, R., 2016. Quantreg: quantile regressions. Available at https://CRAN.R-project.org/package=quantreg

Le Hir P., Roberts, W., Cazaillet, O., Christie, M., Bassoullet, P., Bacher, C., 2000. Characterization of intertidal flat hydrodynamics. Continental Shelf Research 20: 1433-1459.

Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129: 271-280.

Leite, F.P.F., Turra, A., Souza, E.C.F., 2003. Population biology and distribution of the tanaid *Kalliapseudes schubarti* Mañe-Garzon, 1949, in an intertidal flat in Southeastern Brazil. Brazilian Journal of Biology 63: 469-479.

Mani-Peres, C.M., Xavier, L.Y., Santos, C.R., Turra, A., 2016. Stakeholders perception of local environmental changes as a tool for impact assessments in coastal zones. Ocean and Coastal Management 119: 135-145.

McArthur, M.A., Brooke, B.P., Przeslawski, R., Ryan, D.A., Lucieer, V.L., Nichol, S., McCallum, A.W., Mellin C., Cresswell, I.D., Radke, L.C., 2010. On the use of abiotic surrogates to describe marine benthic biodiversity. Estuarine, Coastal and Shelf Science, 88: 21-32.

MacIntyre, H., Geider, R., Miller, D., 1996. Microphytobenthos: the ecological role
of the "secret garden" of unvegetated, shallow-water marine habitats. I.
Distribution, abundance and primary production. Estuaries and Coasts 19: 186-201

Margalef, R., 1968. Perspectives in ecological theory. University of Chicago Press, 111 pp.

Moreno, M., Ferrero, T.J., Gallizia, I., Vezzulli, L., Albertelli, G., Fabiano, M., 2008. An assessment of the spatial heterogeneity of environmental disturbance within an enclosed harbor through the analysis of meiofauna and nematode assemblages. Estuarine, Coastal and Shelf Science 77: 565-576.

Niemi, G.J., McDonald, M.E., 2004. Application of ecological indicators. Annual Review of Ecology, Evolution and Systematics 35: 89-111.

Paiva, P.C., 2001. Spatial and temporal variation of a nearshore benthic community in southern Brazil: Implications for the design of monitoring programs. Estuarine, Coastal and Shelf Science 52: 423-433.

Patricio, J., Adão, H., Neto, J.M., Alves, A.S., Traunspurger, W., Marques, J.C., 2012. Do nematode and macrofauna assemblages provide similar ecological assessment information. Ecological Indicators 14: 124-137.

Pearson, T.H., Rosenberg, R., 1978. Macrobenthic sucession in relation to organic enrichment and pollution on the marine environment. Oceanography and Marine Biology: an Annual Review 16: 229-311.

Pebesma, E.J., 2004. Multivariable statistics in S: the gstat package. Computers & Geosciences 30: 683-691.

Pebesma, E.J., Bivand, R., 2005. Classes and methods for spatial data in R. R News 5, Available at https://cran.r-project.org/doc/Rnews/

Plante-Cuny, M.R. 1973. Recherches sur la production primaire benthique en milieu marin tropical. I. Variations de la production primaire et des teneurs en pigments photosynthétiques sur quelques fonds sableaux. Valeur des résultats obtenus par la méthode du 14C. Cahiers ORSTOM 11:317-348.

Provete, D.B., GOnçalves-Souza, T., Garey, M.V., Martins, I.A., Rossa-Feres, D.C., 2014. Broad-scale spatial patterns of canopy cover and pond morphology affect the structure of a Neotropical amphibian metacommunity. Hydrobiologia 734: 69-79.

R Core Team, 2016. R: A language and environment for statistical computing. Vienna, Austria. Available at: https://www.R-project.org/.

Rivero, M., Elías, R., Vallarino, E., 2005. First survey of macroinfauna in the Mar del Plata Harbor (Argentina), and the use of polychaetes as pollution indicators. Revista de Biologia Marina y Oceanografia 40: 101–108.

Rodil, I., Lucena-Moya, P., Jokinen, H., Ollus, V., Wennhage, H., Villñas, A., Norkko, A., 2017. The role of dispersal mode and habitat specialization for metacommunity structure of shallow beach invertebrates. PlosOne 12: e0172160.

Schratzberger, M., Maxwell, T.A.D., Warr, K., Ellis, J.R., Rogers, S.I., 2008. Spatial variability of infauna nematodes and polychaete assemblages in two muddy subtidal habitats. Marine Biology 153: 621-642.

Seitz, R.D., Wennhage, H., Bergström, U., Lipcius, R.N., Ysebaert, T., 2014. Ecological values of coastal habitats for commercially and ecologically important species. ICES Journal of Marine Science 71: 648-655.

Semprucci, F., Balsamo, M., 2012. Free-living marine nematodes as bioindicators: past, present, and future perspectives. Environmental Research Journal 6: 17-36.

Semprucci, F., Colantoni, P., Baldelli, G., Rocchi, M., Balsamo, M., 2010. The distribution of meiofauna on back-reef sandy platforms in the Maldives (Indian Ocean). Marine Ecology: an evolutionary perspective 31: 592-607.

Semprucci, F., Frontalini, F., Covazzi-Harriague, A., Coccioni, R., Balsamo, M., 2013. Meio- and macrofauna in the marine area of the Monte St. Bartolo Natural Park (Central Adriatic Sea, Italy). Scientia Marina 77: 189-199.

Siddig, A.A.H., Ellison, A.M., Ochs, A., Villar-Leeman, C., Lau, M.K., 2016. How do ecologists select and use indicator species to monitor ecological change? Insight form 14 years of publication on Ecological Indicators. Ecological Indicators 60: 223-230.

Silva, C.F., Seixas, V.C., Barroso, R., Di Domenico, M., Amaral, A.C.Z., Paiva, P.C., 2017. Demystifying the *Capitella capitata* complex (Annelida, Capitellidae) diversity by morphological and molecular data along the Brazilian coast. PlosOne 12: e0177760.

Stelzenmüller, V., Lee., J., South, A., Rogers, S.I., 2010. Quantifying cumulative impacts of human pressures on the marine environment: a geospatial modelling framework. Marine Ecology Progress Series 398: 19-32.

Stevens, T., Connoly, R.M., 2004. Testing the utility of abiotic surrogates for marine habitat mapping at scales relevant to management. Biological Conservation 119: 351-362.

Suguio, K., 1973. Introdução à sedimentologia. São Paulo: EDUSP.

Teodoro, A.C., Duleba, W., Gubitoso, S., Prada, S.M., Lamparelli, C.C., Bevilacqua, J.E., 2010. Analysis of foraminifera assemblages and sediment geochemical propertiesto characterise the environment near Araçá and Saco da Capela domestic sewage submarine outfalls of São Sebastião Channel, São Paulo State, Brazil. Marine Pollution Bulletin 60: 536-553.

Tsutsumi, H., 1987. Population dynamics of *Capitella capitata* (Polychaeta; Capitellidae) in an organically polluted cave. Marine Ecology Progress Series 36: 139-149.

Tyson, R.V., 1995. Abundance of organic matter in sediments: TOC, hydrodynamic equivalence, dilution and flux effects. In: Tyson, R.V. *Sedimentary organic matter* (pp 81-118). Amsterdan: Springer

Vanaverbeke, J., Merckx, B., Degraer, S., Vincx, M., 2011. Sediment-related distribution patterns of nematodes and macrofaunal: two sides of the benthic coin? Marine Environmental Research 71: 31-40.

Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S. Springer: New York. 501 pp.

Verfaillie, E., Degraer, S., Schelfaut, K., Willems, W., Van Lecker, V., 2009. A protocol for classifying ecologically relevant marine zones, a statistical approach. Estuarine, Coastal and Shelf Science 83: 175-185.

Warwick, R.M., 1993. Environmental impact studies on marine communities: pragmatical considerations. Australian Journal of Ecology 18: 63-80.

Wiberg, P.L., Sherwood, C.R., 2008. Calculating wave-generated bottom orbital velocities from surface-wave parameters. Computers and Geosciences 34: 1243-1262.

Zettler, M.F., Proffitt, C.E., Darr, A., Degraer, S., Devriese, L., Greathhead, C., et al. 2013. On the myths of indicator species: issues and further considerations on the use of static concepts for ecological applications. PlosOne 8: e78219.

REFERÊNCIAS

Adler, F. R. 2011. The effects of intraspecific density dependence on species richness and species abundance distributions. Theoretical Ecology 4:153–162.

Admiraal, W., L. A. Bouwman, L. Hoekstra, and K. Romeyn. 1983. Qualitative and Quantitative Interactions between Microphytobenthos and Herbivorous Meiofauna on a Brackish Intertidal Mudflat. Internationale Revue der gesamten Hydrobiologie und Hydrographie 68:175–191.

Akima, H., and A. Gebhardt. 2016. akima: Interpolation of Irregularly and Regularly Spaced Data.

Amaral, A. C. Z., A. Turra, A. M. Ciotti, C. L. B. Rossi-Wongtschowski, and Y. Scharffer-Novelli. 2016. Vida na Baía do Araçá: diversidade e importância. Second. Lume.

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology 26:32–46.

Anderson, M. J. 2006. Distance-Based Tests for Homogeneity of Multivariate Dispersions. Biometrics 62:245–253.

Armenteros, M., J. A. Pérez-García, A. Ruiz-Abierno, L. Díaz-Asencio, Y. Helguera, M. Vincx, and W. Decraemer. 2010. Effects of organic enrichment on nematode assemblages in a microcosm experiment. Marine Environmental Research 70:374–382.

Arya, S., D. Mount, S. E. Kemp, and G. Jefferis. 2017. RANN: Fast Nearest Neighbour Search (Wraps ANN Library) Using L2 Metric.

Balsamo, M., F. Semprucci, F. Frontalini, and R. Coccioni. 2012. Meiofauna as a tool for marine ecosystem biomonitoring. Page Marine Ecosystems, A. Cruzado (Ed.), InTech, 310pp.

Bar-Massada, A., R. Kent, and Y. Carmel. 2014. Environmental heterogeneity affects the location of modelled communities along the niche-neutrality continuum. Proceedings of the Royal Society B: Biological Sciences 281:20133249–20133249.

Barbujani, G., N. L. Oden, and R. R. Sokal. 1989. Detecting Regions of Abrupt Change in Maps of Biological Variables. Systematic Zoology 38:376.

Basille, M., C. Calenge, É. Marboutin, R. Andersen, and J. M. Gaillard. 2008. Assessing habitat selection using multivariate statistics: Some refinements of the ecological-niche factor analysis. Ecological Modelling 211:233–240.

Beckage, B., L. Joseph, P. Belisle, D. B. Wolfson, and W. J. Platt. 2007. Bayesian change-point analyses in ecology. New Phytologist 174:456–467.

Begon, M., C. Townsend, and J. Harper. 2006. Ecology: from individuals to ecosystems. Page Ecological Applications.

Bell, G. 2000. The Distribution of Abundance in Neutral Communities. The American Naturalist 155:606–617.

Bell, G. 2003. The interpretation of biological surveys. Proceedings of the Royal Society B: Biological Sciences 270:2531–2542.

Bell, G. 2005. The co-distribution of species in relation to the neutral theory of community ecology. Ecology 86:1757–1770.

Bemvenutti, C., and S. E. & Netto. 1998. Distribution and seasonal patterns of the sublittoral benthic macrofauna of Patos Lagoon (South Brazil). Revista Brasileira de Biologia 58:211–221.

Bengtsson, J., K. Engelhardt, P. Giller, S. Hobbie, D. Lawrence, J. Levine, M. Vila, and V. Wolters. 2002. Slippin' and slidin' between the scales: the scaling components of biodiversity-ecosystem functioning relations. Pages 209–220 *in* M. Loreau and P. Naeem, S. & Inchausti, editors. Biodiversity and Ecosystem Functioning: Synthesis and Perspectives. Oxford University Press, Oxford.

Biswas, S. R., and H. H. Wagner. 2012. Landscape contrast: A solution to hidden assumptions in the metacommunity concept? Landscape Ecology 27:621–631.

Boeckner, M. J., J. Sharma, and H. C. Proctor. 2009. Revisiting the meiofauna paradox: dispersal and colonization of nematodes and other meiofaunal organisms in low- and high-energy environments. Hydrobiologia 624:91–106.

Bongers, T., and J. van de Haar. 1990. On the potential of basing an ecological typology of aquatic sediments on the nematode fauna: An example from the River Rhine. Hydrobiological Bulletin 24:37–45.

Bonthoux, S., and G. Balent. 2015. Bird metacommunity processes remain constant after 25 years of landscape changes. Ecological Complexity 21:39–43.

Ter Braak, C. J. F. 1994. Canonical community ordination. Part I: basic theory and linear methods. Ecoscience 1:127–140.

Brown, B. L., E. R. Sokol, J. Skelton, and B. Tornwall. 2017. Making sense of metacommunities: dispelling the mythology of a metacommunity typology. Oecologia 183:643–652.

Brown, J. H. 1984. On the Relationship between Abundance and Distribution of Species. The American naturalist 124:255–279.

Burrough, P. A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Journal of Quaternary Sciencee:193.

Camargo, M. G. 2006. SysGran: um sistema de código aberto para análises granulométricas do sedimento. Revista Brasileira de Geociências. 36:371–378.

Channell, R., and M. V. Lomolino. 2000. Dynamic biogeography and conservation of endangered species. Nature 403:84–86.

Chase, J. M. 2014. Spatial scale resolves the niche versus neutral theory debate. Journal of Vegetation Science 25:319–322.

Chase, J. M., and M. A. Leibold. 2003. Ecological niches: linking classical and contemporary approaches. University of Chicago Press.

Chase, J. M., and J. a Myers. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. Philosophical Transactions of the Royal Society B: Biological Sciences 366:2351–2363.

Checon, H. H., D. C. Vieira, G. N. Corte, E. C. P. M. Sousa, G. Fonseca, and A. C. Z. Amaral. 2018. Defining soft bottom habitats and potential indicator species as tools for monitoring coastal systems: A case study in a subtropical bay. Ocean & Coastal Management.

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. Austral Ecology 18:117–143.

Codling, E. A., and A. J. Dumbrell. 2012. Mathematical and theoretical ecology: linking models with ecological processes. Interface Focus 2:144–149.

Cornelius, J. M., and J. F. Reynolds. 1991. On Determining the Statistical Significance of Discontinuities with Ordered Ecological Data. Ecology 72:2057–2070.

Corte, G. N., H. H. Checon, G. Fonseca, D. C. Vieira, F. Gallucci, M. Di Domenico, and A. C. Z. Amaral. 2017. Cross-taxon congruence in benthic communities: Searching for surrogates in marine sediments. Ecological Indicators 78:173–182.

Coull, B. C. 1999. Role of meiofauna in estuarine soft-bottom habitats*. Austral Ecology 24:327–343.

Courbaud, B., V. Lafond, G. Lagarrigues, G. Vieilledent, T. Cordonnier, F. Jabot, and F. de Coligny. 2015. Applying ecological model evaludation: Lessons learned with the forest dynamics model Samsara2. Ecological Modelling 314:1–14.

Dolédec, S., D. Chessel, and C. Gimaret-Carpentier. 2000. Niche separation in community analysis: a new method. Ecology 81:2914–2927.

Dray, S., and A. B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22:1–20.

Van Dyke Parunak, H., R. Savit, and R. L. Riolo. 1998. Agent-Based Modeling vs. Equation-Based Modeling: A Case Study and Users' Guide. Pages 10–25 Proceedings of Multi-agent systems and Agent-based Simulation (MABS'98).

Eleftheriou, A., and A. McIntyre. 2007. Methods for the Study of Marine Benthos. Page (A. Eleftheriou and A. McIntyre, Eds.) Methods for the Study of Marine Benthos: Third Edition. Third. Blackwell Science Ltd.

Erdos, L., Z. Bátori, C. S. Tölgyesi, and L. Körmöczi. 2014. The moving split window (MSW) analysis in vegetation science - An overview. Applied Ecology and Environmental Research 12:787–805.

Ertel, J. E., and E. B. Fowlkes. 1976. Some Algorithms for Linear Spline and Piecewise Multiple Linear Regression. Journal of the American Statistical Association 71:640–648.

Etienne, R. S., and D. Alonso. 2007. Neutral Community Theory: How Stochasticity and Dispersal-Limitation Can Explain Species Coexistence. Journal of Statistical Physics 128:485–510.

Farjalla, V. F., D. S. Srivastava, N. A. C. Marino, F. D. Azevedo, V. Dib, P. M. Lopes, A. S. Rosado, R. L. Bozelli, and F. A. Esteves. 2012. Ecological determinism increases with organism size. Ecology 93:1752–1759.

Folk, R. L., and W. C. Ward. 1957. Brazos River bar [Texas]; a study in the significance of grain size parameters. Journal of Sedimentary Research 27:3–26.

Fonseca, G., T. F. Maria, N. Kandratavicius, V. Venekey, P. F. Gheller, and F. Gallucci. 2014. Testing for nematode–granulometry relationships. Marine Biodiversity 44:435–443.

Fox, J., and S. Weisberg. 2011. An R companion to applied regression. Sage Publications.

Fraser, L. H., and P. Keddy. 1997. The role of experimental microcosms in ecological research. TREE 12:478–481.

Fukumori, K., G. Livingston, and M. A. Leibold. 2015. Disturbance mediated colonization-extinction dynamics in experimental protist metacommunities. Ecology:150522233701009.

Gallucci, F., I. B. de Castro, F. C. Perina, D. M. de Souza Abessa, and A. de Paula Teixeira. 2015. Ecological effects of Irgarol 1051 and Diuron on a coastal meiobenthic community: A laboratory microcosm experiment. Ecological Indicators 58:21–31.

Gallucci, F., T. Moens, and G. Fonseca. 2009. Small-scale spatial patterns of meiobenthos in the Arctic deep sea. Marine Biodiversity 39:9–25.

Gascón, S., I. Arranz, M. Cañedo-Argüelles, A. Nebra, A. Ruhí, M. Rieradevall, N. Caiola, J. Sala, C. Ibàñez, X. D. Quintana, and D. Boix. 2016. Environmental filtering determines metacommunity structure in wetland microcrustaceans. Oecologia 181:193–205.

Gaudette, H. E., and W. R. Flight. 1974. An Inexpensive Titration Method for the Determination of Organic Carbon in Recent Sediments. SEPM Journal of Sedimentary Research Vol. 44.

Giere, O. 2009. Meiobenthology. Springer Berlin Heidelberg, Berlin, Heidelberg.

Giesy, J., and P. Allred. 1985. Replicability of aquatic multipecies test systems.

Gilbert, B., and M. J. Lechowicz. 2004. Neutrality, niches, and dispersal in a temperate forest understory. Proceedings of the National Academy of Sciences 101:7651–

7656.

Gingold, R., T. Moens, and A. Rocha-Olivares. 2013. Assessing the Response of Nematode Communities to Climate Change-Driven Warming: A Microcosm Experiment. PLoS ONE 8.

Gollner, S., B. Govenar, C. R. Fisher, and M. Bright. 2015. Size matters at deep-sea hydrothermal vents: Different diversity and habitat fidelity patterns of meio- and macrofauna. Marine Ecology Progress Series 520:57–66.

Goslee, S. C., and D. L. Urban. 2007. The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software 22:1–19.

Gravel, D., C. D. Canham, M. Beaudet, and C. Messier. 2006. Reconciling niche and neutrality: the continuum hypothesis. Ecology Letters 9:399–409.

Gravetter, F. J., and L. B. Wallnau. 2016. Statistics for the Behavioral Sciences. Tenth. Cengage Learning.

Gray, J. 2002. Species richness of marine soft sediments. Marine Ecology Progress Series 244:285–297.

Gray, J. S., A. Bjorgesaeter, and K. I. Ugland. 2005. The impact of rare species on natural assemblages. Journal of Animal Ecology 74:1131–1139.

Grimm, V. 2002. Bottom-Up Simulation Models in Ecology 15.

Grimm, V. 2005. Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. Science 310:987–991.

Grimm, V., D. Ayllón, and S. F. Railsback. 2017. Next-Generation Individual-Based Models Integrate Biodiversity and Ecosystems: Yes We Can, and Yes We Must. Ecosystems 20:229–236.

Grimm, V., and S. F. Railsback. 2012. Pattern-oriented modelling: a “multi-scope” for predictive systems ecology. Philosophical Transactions of the Royal Society B: Biological Sciences 367:298–310.

Grüner, N., C. Gebühr, M. Boersma, U. Feudel, K. H. Wiltshire, and J. A. Freund. 2011. Reconstructing the realized niche of phytoplankton species from environmental data: fitness versus abundance approach. Limnology and Oceanography: Methods 9:432–442.

Gubitoso, S., W. Duleba, A. C. Teodoro, S. M. Prada, M. M. Rocha, C. C. Lamparelli, J. E. Bevilacqua, and D. O. Moura. 2008. Estudo geoambiental da região circunjacente ao emissário submarino de esgoto do Araçá, São Sebastião (SP). Revista Brasileira de Geociências 38:467–475.

Hallam, T. G., and S. A. Levin. 1986. Mathematical Ecology. An introduction. Page Biomathematics.

Hartig, F., J. M. Calabrese, B. Reineking, T. Wiegand, and A. Huth. 2011. Statistical

inference for stochastic simulation models - theory and application. Ecology Letters 14:816–827.

Heino, J. 2005. Positive relationship between regional distribution and local abundance in stream insects: a consequence of niche breadth or niche position? Ecography 28:345–354.

Heino, J. 2013. Does dispersal ability affect the relative importance of environmental control and spatial structuring of littoral macroinvertebrate communities? Oecologia 171:971–980.

Heino, J., A. S. Melo, T. Siqueira, J. Soininen, S. Valanko, and L. M. Bini. 2015. Metacommunity organisation, spatial extent and dispersal in aquatic systems: Patterns, processes and prospects. Freshwater Biology 60:845–869.

Heip, C., M. Vincx, and G. Vranken. 1985. The ecology of marine nematodes. Oceanogr Mar Biol Annu Rev 23:399–489.

Henderson, P. A., and A. E. Magurran. 2010. Linking species abundance distributions in numerical abundance and biomass through simple assumptions about community structure. Proceedings. Biological sciences / The Royal Society 277:1561–1570.

Hilderbrand, R. H. 2003. The roles of carrying capacity , immigration , and population synchrony on persistence of stream-resident cutthroat trout 110:257–266.

Hirzel, A. H., and G. Le Lay. 2008. Habitat suitability modelling and niche theory. Journal of Applied Ecology 45:1372–1381.

Hobbs, N. T., and T. A. Hanley. 1990. Habitat Evaluation - Do Use Availability Data Reflect Carrying-Capacity. Journal of Wildlife Management 54:515–522.

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

Hufkens, K., P. Scheunders, and R. Ceulemans. 2009. Ecotones in vegetation ecology: Methodologies and definitions revisited. Ecological Research 24:977–986.

Jackson, L. J., A. S. Trebitz, and K. L. Cottingham. 2000. An Introduction to the Practice of Ecological Modeling. BioScience 50:694.

Josefson, A. B. 2016. Species Sorting of Benthic Invertebrates in a Salinity Gradient – Importance of Dispersal Limitation. PLOS ONE 11:e0168908.

Keddy, P. A. 1992. Assembly and response rules: two goals for predictive community ecology. Journal of Vegetation Science 3:157–164.

Kent, M., R. Moyeed, C. L. Reid, R. Pakeman, and R. Weaver. 2006. Geostatistics, spatial rate of change analysis and boundary detection in plant ecology and biogeography. Progress in Physical Geography 30:201–231.

Körmöczi, L., Z. Bátori, L. Erdős, C. Tölgyesi, M. Zalatnai, and C. Varró. 2016. The role of randomization tests in vegetation boundary detection with moving split-window analysis. Journal of Vegetation Science 27:1288–1296.

Kramer-Schadt, S., E. Revilla, T. Wiegand, and V. Grimm. 2007. Patterns for parameters in simulation models. Ecological Modelling 204:553–556.

Labrune, C., A. Grémare, J. M. Amouroux, R. Sardá, J. Gil, and S. Taboada. 2007. Assessment of soft-bottom polychaete assemblages in the Gulf of Lions (NW Mediterranean) based on a mesoscale survey. Estuarine, Coastal and Shelf Science 71:133–147.

Labrune, C., A. Grémare, J. M. Amouroux, R. Sardá, J. Gil, and S. Taboada. 2008. Structure and diversity of shallow soft-bottom benthic macrofauna in the Gulf of Lions (NW Mediterranean). Helgoland Marine Research 62:201–214.

Lagarrigues, G., F. Jabot, V. Lafond, and B. Courbaud. 2015. Approximate Bayesian computation to recalibrate individual-based models with population data: Illustration with a forest simulation model. Ecological Modelling 306:278–286.

Lambshead, P. J. D., P. J. D. Lambshead, G. Boucher, and G. Boucher. 2003. Marine nematode deep-sea biodiversity - hyperdiverse or hype? Journal of Biogeography 30:475–485.

Latombe, G., C. Hui, and M. A. Mcgeoch. 2015. Beyond the continuum : a multi-dimensional phase space for neutral – niche community assembly. Proceedings of the Royal Society B: Biological Sciences:8–10.

Latombe, G., L. Parrott, and D. Fortin. 2011. Levels of emergence in individual based models: Coping with scarcity of data and pattern redundancy. Ecological Modelling 222:1557–1568.

Legendre, P., and L. Legendre. 1998. Numerical ecology: second English edition. Developments in environmental modelling 20.

Legendre, P., and L. Legendre. 2012. Numerical Ecology. Page Ecology. Third. Developments in Environmental Modeling.

Legendre, P., J. Oksanen, and C. J. F. ter Braak. 2011. Testing the significance of canonical axes in redundancy analysis. Methods in Ecology and Evolution 2:269–277.

Leibold, M. A. 1995. The Niche Concept Revisited: Mechanistic Models and Community Context. Ecology 76:1371–1382.

Lerman, P. M. 1980. Fitting Segmented Regression Models by Grid Search. Journal of the Royal Statistical Society. Series C (Applied Statistics) 29:77–84.

Liu, J., S. Wu, and J. V. Zidek. 1997. On Segmented Multivariate Regression. Statistica Sinica 7:497–525.

Logue, J. B., N. Mouquet, H. Peter, and H. Hillebrand. 2011. Empirical approaches to

metacommunities: a review and comparison with theory. Trends in ecology & evolution 26:482–91.

Ludwig, J. a, and J. M. Cornelius. 1987. Locating Discontinuities along Ecological Gradients. Ecology 68:448–450.

MacArthur, R. H. 1972. Geographical Ecology: Patterns in the Distribution of Species. Princeton University Press, New Jersey.

Magurran, A. E., and P. A. Henderson. 2003. Explaining the excess of rare species in natural species abundance distributions. Nature 422:714–716.

Makarenkov, V., and P. Legendre. 2002. Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. Ecology 83:1146–1161.

Maria, A. 1997. Introduction to modeling and simulation. Pages 7–13 Proceedings of the 29th conference on Winter simulation - WSC '97. ACM Press, New York, New York, USA.

Marilleau, N., C. Lang, and P. Giraudoux. 2018. Coupling agent-based with equation-based models to study spatially explicit megapopulation dynamics. Ecological Modelling 384:34–42.

Matějíček, L. 2002. Ecological Modelling Using the Spreadsheet Model TabSim:354–359.

McClain, C. R., and M. a. Rex. 2015. Toward a Conceptual Understanding of β-Diversity in the Deep-Sea Benthos. Annual Review of Ecology, Evolution, and Systematics:623–642.

Moens, T., U. Braeckman, S. Derycke, G. Fonseca, F. Gallucci, R. Gingold, K. Guilini, J. Ingels, D. Leduc, J. Vanaverbeke, C. Van Colen, A. Vanreusel, and M. Vincx. 2013. 3. Ecology of free-living marine nematodes. Pages 109–152 Nematoda. DE GRUYTER, Berlin, Boston.

Montagna, P. 1984. In situ measurement of meiobenthic grazing rates on sediment bacteria and edaphic diatoms. Marine Ecology Progress Series 18:119–130.

Morgan, M. S. 2005. Experiments versus models : New phenomena , inference and surprise:317–329.

Mouquet, N., and M. Loreau. 2003. Community Patterns in Source-Sink Metacommunities. The American Naturalist 162:544–557.

Netto Paulo da Cunha, S. A. and L. 1994. Effects of sediment disturbance on the structure of benthic fauna in subtropical tidal creek of southeastern Brazil. Marine Ecology Progress Series, 106: 239-247, 4 figures, 3 tables. Marine Ecology Progress Series.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs,

and H. Wagner. 2017. vegan: Community Ecology Package.

Ovaskainen, O., H. J. de Knegt, and M. del M. Delgado. 2016. Quantitative ecology and evolutionary biology: integrating models with data:285.

Palmer, M. 1988. Dispersal of marine meiofauna: A review and conceptual model explaining passive transport and active emergence with implications for recruitment. Marine Ecology Progress Series 48:81–91.

Park, S. J., R. A. J. Taylor, and P. S. Grewal. 2013. Spatial organization of soil nematode communities in urban landscapes: Taylor's Power Law reveals life strategy characteristics. Applied Soil Ecology 64:214–222.

Parrott, L., and R. Kok. 2000. Incorporating complexity in ecosystem modelling. Complexity International 7:1–19.

Peres-Neto, P. R., P. Legendre, S. Dray, and D. Borcard. 2006. Variation Partitioning of Species Data Matrices : Estimation and Comparison of Fractions. Ecology 87:2614–2625.

Pilditch, C. A., S. Valanko, J. Norkko, and A. Norkko. 2015. Post-settlement dispersal: The neglected link in maintenance of soft-sediment biodiversity. Biogical Letters 11:1–6.

Plante-Cuny, M. R. 1978. Recherches sur la production primaire bentique en milieu marin tropical 1. Variations de la production primaire et des teuners en pigments photosyntétiques sur quelques fonds sableaux. Valeur des résultats obtenus par la méthode du C14. Cah. ORSTOM Sér. Océanogr 11:317–348.

La Point, T. W., and J. A. Perry. 1989. Use of experimental ecosystems in regulatory decision making. Environmental Management 13:539–544.

R Development Core team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rádková, V., J. Bojková, V. Křoupalová, J. Schenková, V. Syrovátka, and M. Horsák. 2014. The role of dispersal mode and habitat specialisation in metacommunity structuring of aquatic macroinvertebrates in isolated spring fens. Freshwater Biology 59:2256–2267.

Radway, J. C. 2008. Functioning of Microphytobenthos in Estuaries . Based on a symposium held in Amsterdam, The Netherlands, 21–23 August 2003 . Edited by Jacco C Kromkamp, Jody F C de Brouwer, Gérard F Blanchard, Rodney M Forster, and Véronique Créach. Amsterdam (The Nether. The Quarterly Review of Biology 83:130–130.

Railsback, S. F. 2008. Getting "results": the pattern-oriented approach to analyzing natural systems with individual-based models. Natural Resource Modeling 14:465–475.

Rao, C. 1964. The use and interpretation of principal component analysis in applied research. Sankhyā: The Indian Journal of Statistics, Series A 26:329–358.

Reddy, T. A. 2011. Applied data analysis and modeling for energy engineers and scientists. Springer Science & Business Media.

Regan, H. M., Y. Ben-Haim, B. Langford, W. G. Wilson, P. Lundberg, S. J. Andelman, and M. A. Burgman. 2005. Robust decision-making under severe uncertainty for conservation management. Ecological Applications 15:1471–1477.

Rodil, I. F., P. Lucena-Moya, H. Jokinen, V. Ollus, H. Wennhage, A. Villnäs, and A. Norkko. 2017. The role of dispersal mode and habitat specialization for metacommunity structure of shallow beach invertebrates. Plos One 12:e0172160.

Rouse, G., and F. Pleijel. 2001. Polychaetes. Oxford university press.

Ruiz-Abierno, A., and M. Armenteros. 2017. Coral reef habitats strongly influence the diversity of macro- and meiobenthos in the Caribbean. Marine Biodiversity 47:101–111.

Rumohr, H. 1999. ICES Techniques in Marine Environmental Sciences. Soft bottom macrofauna: Collection, treatment, and quality assurance of samples. International Counsil for the Exploration of the Sea:1–20.

Rundell, R. J., and B. S. Leander. 2010. Masters of miniaturization: Convergent evolution among interstitial eukaryotes. BioEssays 32:430–437.

Santos, A. C. C., R. B. Choueri, G. de Figueiredo Eufrasio Pauly, D. Abessa, and F. Gallucci. 2018. Is the microcosm approach using meiofauna community descriptors a suitable tool for ecotoxicological studies? Ecotoxicology and Environmental Safety 147:945–953.

Schmera, D., J. Podani, Z. Botta-Dukát, and T. Erős. 2018. On the reliability of the Elements of Metacommunity Structure framework for separating idealized metacommunity patterns. Ecological Indicators 85:853–860.

Schratzberger, M., J. M. Gee, H. L. Rees, S. E. Boyd, and C. M. Wall. 2000. The structure and taxonomic composition of sublittoral meiofauna assemblages as an indicator of the status of marine environments. Journal of the Marine Biological Association of the United Kingdom 80:969–980.

Schratzberger, M., P. Whomersley, K. Warr, S. G. Bolam, and H. L. Rees. 2004. Colonisation of various types of sediment by estuarine nematodes via lateral infaunal migration : a laboratory study. Marine Biology 145:69–78.

Sherman, K. M., and B. C. Coull. 1980. The response of meiofauna to sediment disturbance. Journal of Experimental Marine Biology and Ecology 46:59–71.

Shi, Y., P. Grogan, H. Sun, J. Xiong, Y. Yang, J. Zhou, and H. Chu. 2015. Multi-scale variability analysis reveals the importance of spatial distance in shaping Arctic soil microbial functional communities. Soil Biology and Biochemistry 86:126–134.

Shurin, J. B., K. Cottenie, and H. Hillebrand. 2009. Spatial autocorrelation and dispersal limitation in freshwater organisms. Oecologia 159:151–159.

Smith, T. W., and J. T. Lundholm. 2010. Variation partitioning as a tool to distinguish between niche and neutral processes. Ecography 33:648–655.

Snelgrove, P. V. R. 1997. The Importance of Marine Sediment Biodiversity in Ecosystem Processes. Royal Swedish Academy of Sciences 26:578–583.

Somerfield, P. J., and R. M. Warwick. 2013. Meiofauna Techniques. Pages 253–284 Methods for the Study of Marine Benthos. John Wiley & Sons, Ltd, Oxford, UK.

Souza, W. P. 1984. The role of disturbance in natural communities. Annual Review of Ecology and Systematics 15:353–391.

Srivastava, D. S., J. Kolasa, J. Bengtsson, A. Gonzalez, S. P. Lawler, T. E. Miller, P. Munguia, T. Romanuk, D. C. Schneider, and M. K. Trzcinski. 2004. Are natural microcosms useful model systems for ecology ? 19.

Suguio, K. 1973. Introdução à sedimentologia. Blücher, Edigar, São Paulo.

Supp, S. R., and S. K. M. Ernest. 2014. Species-level and community-level responses to disturbance: A cross-community analysis. Ecology 95:1717–1723.

Supp, S. R., D. N. Koons, and S. K. M. Ernest. 2015. Using life history trade-offs to understand core-transient structuring of small mammal community. Ecosphere 6:187.

Swihart, R. K., J. J. Lusk, J. E. Duchamp, C. E. Rizkalla, and J. E. Moore. 2006. The roles of landscape context, niche breadth, and range boundaries in predicting species responses to habitat alteration. Diversity <html_ent glyph="@amp;" ascii="&"/> Distributions 12:277–287.

Tahseen, Q. 2012. Nematodes in aquatic environments : adaptations and survival 3:13–40.

Thompson, R., and C. Towsend. 2006. A truce with neutral theory:local determinist factors species traits and dispersal limitation together determine patters of diversity in stream invertebrates. Journal of Animal Ecology 75:476–484.

Toms, J. D., and M. L. Lesperance. 2003. Piecewise regression: A tool for identifying ecological thresholds. Ecology 84:2034–2041.

Ugland, K. I., and J. S. Gray. 1982. Lognormal Distributions and the Concept of Community Equilibrium. Oikos 39:171.

Ullberg, J., and E. Ólafsson. 2003. Free-living marine nematodes actively choose habitat when descending from the water column. Marine Ecology Progress Series 260:141–149.

Vanschoenwinkel, B., C. De Vries, M. Seaman, and L. Brendonck. 2007. The role of metacommunity processes in shaping invertebrate rock pool communities along a dispersal gradient. Oikos:1–12.

Vellend, M., D. S. Srivastava, K. M. Anderson, C. D. Brown, J. E. Jankowski, E. J.

Kleynhans, N. J. B. Kraft, A. D. Letaw, A. A. M. Macdonald, J. E. Maclean, I. H. Myers-Smith, A. R. Norris, and X. Xue. 2014. Assessing the relative importance of neutral stochasticity in ecological communities. Oikos 123:1420–1430.

Vergnon, R., E. H. van Nes, and M. Scheffer. 2012. Emergent neutrality leads to multimodal species abundance distributions. Nature Communications 3:663.

Vieira, D. C., and G. Fonseca. 2013. The Importance of Vertical and Horizontal Dimensions of the Sediment Matrix in Structuring Nematodes Across Spatial Scales. PLoS ONE 8:e77704.

Vinet, L., and A. Zhedanov. 2010. A “missing” family of classical orthogonal polynomials. Biological Conservation 110:305.

Warwick, R. M., S. L. Dashfield, and P. J. Somerfield. 2006. The integral structure of a benthic infaunal assemblage. Pages 12–18 Journal of Experimental Marine Biology and Ecology.

Wiegand, T., F. Jeltsch, I. Hanski, and V. Grimm. 2003. Using pattern-oriented modeling for revealing hidden information : a key for reconciling ecological theory and application. Oikos 100:209–222.

Wieser, V. W. 1953. Die Beziehung zwisehen Mundloblengestalt, Ernäbrungsweise und Vokommen bei frcilebenden marinen Nematoden. Arkiv Für Zoologir Zoologi 4.

Wilson, M., and T. Kakouli-Duarte. 2013. Nematodes as Environmental Indicators. Page Journal of Chemical Information and Modeling.