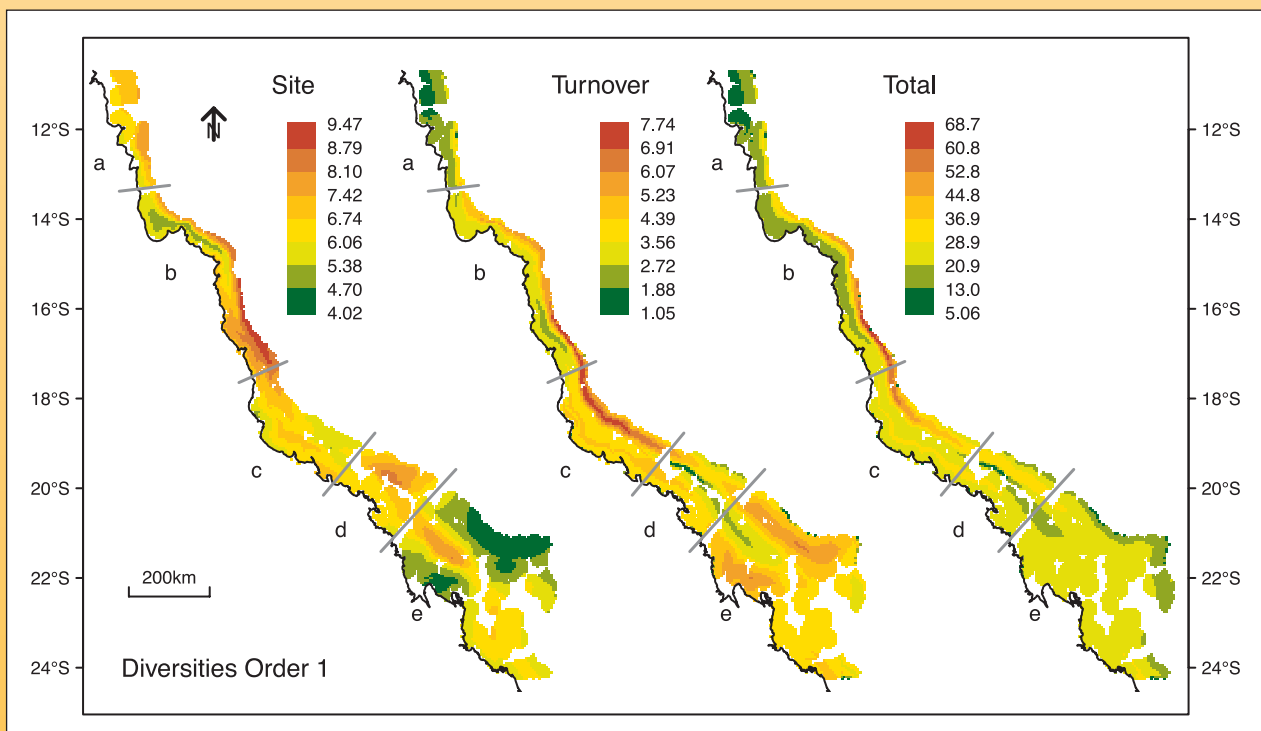




## Relating Fish Diversity to Environmental Drivers and the Great Barrier Reef Zoning



Glenn De'ath, Mike Cappo, Katharina Fabricius,  
Alister Cheal and Greg Coleman



# **Relating Fish Diversity to Environmental Drivers and the Great Barrier Reef Zoning**

Glenn De'ath<sup>1</sup>, Mike Cappo<sup>1</sup>, Katharina Fabricius<sup>1</sup>,  
Alister Cheal<sup>1</sup> and Greg Coleman<sup>1</sup>

<sup>1</sup>Australian Institute of Marine Science, PMB 3, Townsville Qld 4810, Australia

Corresponding author: Glenn De'ath; email: [g.death@aims.gov.au](mailto:g.death@aims.gov.au)  
Australian Institute of Marine Science, PMB 3, Townsville Qld 4810, Australia.



**Australian Government**

---

**Department of Sustainability, Environment,  
Water, Population and Communities**

Supported by the Australian Government's  
National Environmental Research Program  
Project 5.1: Biodiversity of the GBR

© The Australian Institute of Marine Science

National Library of Australia Cataloguing-in-Publication entry:

978-1-921359-81-1

This report should be cited as:

Relating Fish Diversity to Environmental Drivers and the Great Barrier Reef Zoning. Glenn De'ath, Mike Cappel, Katharina Fabricius, Alister Cheal and Greg Coleman. Report to the National Environmental Research Program. Reef and Rainforest Research Centre Limited, Cairns (38pp.).

Published by the Reef and Rainforest Research Centre on behalf of the Australian Governments National Environmental Research Program (NERP) Tropical Ecosystems (TE) Hub.

The Tropical Ecosystems Hub is part of the Australian Government's National Environmental Research Program. The NERP TE Hub is administered in North Queensland by the Reef and Rainforest Research Centre Limited (RRRC). The NERP Tropical Ecosystems Hub addresses issues of concern for the management, conservation and sustainable use of the World Heritage listed Great Barrier Reef (GBR) and its catchments, tropical rainforests including the Wet Tropics World Heritage Area (WTWHA), and the terrestrial and marine assets underpinning resilient communities in the Torres Strait, through the generation and transfer of world-class research and shared knowledge.

This publication is copyright. The Copyright Act 1968 permits fair dealing for study, research, information or educational purposes subject to inclusion of a sufficient acknowledgement of the source.

The views and opinions expressed in this publication are those of the authors and do not necessarily reflect those of the Australian Government or the Minister for Sustainability, Environment, Water, Population and Communities.

While reasonable effort has been made to ensure that the contents of this publication are factually correct, the Commonwealth does not accept responsibility for the accuracy or completeness of the contents, and shall not be liable for any loss or damage that may be occasioned directly or indirectly through the use of, or reliance on, the contents of this publication.

Cover image: Site, turnover and total diversities of seafloor fish on the Great Barrier Reef.

This report is available for download from the NERP Tropical Ecosystems Hub website: <http://www.nerptropical.edu.au/research>

01 June 2013

# Contents

|  |           |
|--|-----------|
| <b>Summary</b>   | <b>7</b>  |
| <b>Introduction</b>  | <b>7</b>  |
| <b>What is Diversity?</b>                                  | <b>8</b>  |
| <b>Data</b>  | <b>9</b>  |
| <b>Methods</b>   | <b>11</b> |
| <b>Spatial Distribution of Site Abundances</b>             | <b>12</b> |
| <b>Spatial and Environmental Drivers of Diversity</b>      | <b>13</b> |
| Distributions of species and site abundances and presences | 13        |
| Measures of diversity                                      | 13        |
| <b>MDM Analysis of Inter-reef Fish Data</b>                | <b>15</b> |
| Spatial distribution of diversities                        | 15        |
| Visualising the diversity model                            | 16        |
| Order 1 diversity  | 16        |
| Order 0 and 2 diversities                                  | 19        |
| Spatial distribution of selected species                   | 19        |
| Effects of environmental drivers                           | 21        |
| Joint effects of spatial and environmental drivers         | 23        |
| Effects on individual species                              | 23        |
| <b>Assessment of GBR Zoning</b>                            | <b>25</b> |
| <b>Discussion</b>  | <b>28</b> |
| <b>Acknowledgements</b>                                    | <b>28</b> |
| <b>Appendices</b>  | <b>30</b> |
| A1: Diversity and the Multinomial Diversity Model          | 30        |
| Diversity  | 30        |
| The Multinomial Diversity Model                            | 31        |
| Parameteric Entropy and Diversity                          | 31        |
| The Multinomial Model                                      | 32        |
| Linking Entropy to the Multinomial Model                   | 33        |
| A2: Analysis of a Simple Data Set                          | 35        |
| A3: Diversity of Order $\neq 1$                            | 37        |
| A4: The Across-Along Spatial Coordinates of the GBR        | 38        |

# List of Figures

|    |   |    |
|----|---|----|
| 1  | The locations of the fish surveys on the Great Barrier Reef. The diameters of the circles are proportional to the BRUVS site diversities (range 1.2–24.3), and show strong spatial variation but also local correlation. Inshore sites have lower site diversity than offshore sites, other than in the southern regions where highest diversity occurs on the mid-shelf. . . . . | 11 |
| 2  | Spatial distribution of predicted site total abundance of inter-reef fish on the GBR. The spatial coordinates of the models were based on relative distances across and along the GBR. Predictions are limited to locations with a precision of <2 standard errors. . . . .   | 12 |
| 3  | Individual species total abundances (A, C) and presences (B, D) (black) for sites (A, B) and species (C, D) and their accumulative values (red). The extreme distribution of fish counts is clearly apparent (A, C). . . . .  | 13 |
| 4  | ${}^1D$ diversity of inter-reef fish on the GBR. The three panels represent site, turnover and local diversities. Within each map, the 5 regions (a-e) are defined by regional similarities in spatial patterns of diversity along the GBR. . . . .   | 17 |
| 5  | ${}^0D$ diversity of inter-reef fish on the GBR. The three panels represent site, turnover and local diversities. Within each map, the 5 regions (a-e) are defined by regional similarities in spatial patterns along the GBR. . . . .  | 17 |
| 6  | ${}^2D$ diversity of inter-reef fish on the GBR. The three panels represent site, turnover and local diversities. Within each map, the 5 regions (a-e) are defined by regional similarities in spatial patterns along the GBR. . . . .  | 18 |
| 7  | $\log_2$ ratios ( ${}^0D/{}^2D$ ) of diversity of inter-reef fish on the GBR. The two panels represent site, turnover and local diversities. Within each map, the 5 regions (a-e) are defined by regional similarities in spatial patterns along the GBR. . . .   | 18 |
| 8  | Spatial distribution of the estimated proportional abundances (%) of <i>N.furcosus</i> , <i>E.naucrates</i> and <i>S.queenslandicus</i> . The strong cross-shelf patterns underpin the similar patterns observed in the diversity maps. Legend scales are on a logarithmic scale to better represent the low abundance areas . . . . .  | 20 |
| 9  | Spatial distribution of the estimated proportional abundances (%) of ( <i>D.russelli</i> , <i>P.nagasakiensis</i> and <i>S.leptolepis</i> ). The latter two species have highly localised high populations. Legend scales are on a logarithmic scale to better represent the low abundance areas . . . . .  | 20 |
| 10 | Spatial distribution of depth (A) and the three environmental drivers, grain size (B), mud (C) and carbonate (D). . . . .   | 21 |
| 11 | Distribution of ${}^1D$ depth (m) (A), grain size (B), mud (%) (C) and carbonate (%) (D). Site diversity (+), smoothed site diversity (red) and environmental driver diversity (blue) along the 4 environmental gradients. The vertical distance between the red and blue lines represents local turnover. . . . .  | 22 |

|    |  |    |
|----|--|----|
| 12 | Entropies for the 10 most abundant taxa. The entropies of the 5 models are indicated: total ( $\vdash$ ), site ( $\vdash$ ), spatial-environment ( $\circ$ ) spatial ( $\triangle$ ) and environment ( $\nabla$ ). The horizontal distances between the $\vdash$ and each of $\vdash$ , $\circ$ , $\triangle$ and $\nabla$ represent the entropies accounted for by the sites, spatial-environment, spatial and environment models for each of the 10 species. . . . . | 24 |
| 13 | The Great Barrier Reef Marine Park zoning. . . . .   | 25 |
| 14 | Diversities of orders 0, 1 and 2 for the three zones: Marine National Park (MNP), Habitat Protection (HP) and General Use (GU). . . . .  | 26 |
| 15 | Identifying the dominant species for 2 or 3 groups: (A) MNP <i>vs.</i> HP – GU and (B) MNP <i>vs.</i> HP <i>vs.</i> GU. . . . .  | 27 |
| 16 | Response curves of the four species along the gradient X for the data from Table 7. The four black lines (different combinations of dashes and dots) represent the predicted probabilities of the species occurring at locations along the gradient. The solid grey line indicates the estimated site diversity along the gradient. . . . .  | 35 |
| 17 | The across-along coordinate system of the Great Barrier Reef showing the across (A) and along (B) values in terms of longitude and latitude coordinates. . . . .   | 38 |

## List of Tables

|   |   |    |
|---|---|----|
| 1 | The hierarchy of parametric diversities of ${}^0D$ , ${}^1D$ and ${}^2D$ of the fish species data. . . . .  | 14 |
| 2 | The analysis of entropy and diversity table. . . . .  | 15 |
| 3 | Analysis of entropy diversities of orders 0, 1 and 2. For each order of diversity three models are fitted: total, spatial and site. Columns contain the models, the entropy ( ${}^1H$ ) and change in entropy $\Delta^1H$ between models, the diversity ${}^1D$ and ratio $\Delta^1D$ , the diversity ${}^0D$ , ratio $\Delta^0D$ , the diversity ${}^2D$ , and ratio $\Delta^2D$ and the ratio ${}^0D/{}^2D$ . . . . . | 15 |
| 4 | Analysis of entropy and diversities environmental and spatial models. Columns indicate the terms in the model, the entropy ( ${}^1H$ ) and change in entropy $\Delta^1H$ between models, the diversity ${}^1D$ and change in diversity $\Delta^1D$ , and the significance of the permutation tests. . . . .   | 23 |
| 5 | Analysis of entropy and diversity of order 1 and significance of model comparisons. The 4 fitted models are: (1) constant, (2) comparison of MNP and HP – GU, (4) 3-way comparison of MNP, HP and GU, and (4) site. Tests of significance are based on permutation tests (n=1000). . . . .  | 26 |
| 6 | Diversities ( $\alpha\beta\gamma$ ) of orders 1 and 0 for all sites, and for sites grouped by the four zones: Marine National Park (MNP), combined Habitat Protection and General Use (HP – GU), Habitat Protection (HP) and General Use (GU). . . . .  | 26 |
| 7 | Example data used to illustrate the variation of diversity of four species S1 – S4 for the three predictors of the MDM: a constant (1), sites (Sites) and a continuous gradient (X). . . . .  | 35 |

|   |  |    |
|---|--|----|
| 8 | The analysis of deviance and diversity for the example data. The predictors for each of the three models are defined, together with the degrees of freedom (df), deviance (Dev), changes in df ( $\Delta$ df) and deviance ( $\Delta$ Dev), diversity ( ${}^1D$ ) and proportional change in diversity ( $\Delta$ ${}^1D$ ). . . . . | 36 |
|---|--|----|

## Summary

The Great Barrier Reef (GBR) has experienced a 50.8% decline (28.0%–13.8%) in coral cover from 1985–2012. In order to facilitate recovery and effectively manage this ecosystem, we need to go beyond simple measures of reef health such as coral cover, and understand the spatial and temporal dynamics of the reef biota, and their relationships to environmental drivers.

Diversity is a key ecological concept in this process, and this work describes and models the diversity of the GBR using a novel statistical methodology developed within this project, namely the multinomial diversity model (MDM).

We outline the foundations and capabilities of the MDM and use it to investigate (1) the spatial distribution of diversity of inter-reef fishes and its dependency on environmental drivers, and (2) variation in diversity across the zoning of the Great Barrier Reef. Using the MDM, total diversity can be decomposed into site and turnover diversities that are simple to interpret in terms of the spatial and environmental drivers.

Spatial variation of diversity was mapped and regions of high and low levels of site diversity and species turnover were identified. Site diversity was highest in the outer GBR around 15°N and lowest in the southern inner and offshore regions. Species turnover was highest in the central offshore region and lowest in the inner northern reefs. Environmental and spatial predictors jointly accounted for 51.4% of the total variation of diversity, and separately they accounted for 36.7% and 34.9% respectively.

Diversities were compared between the Marine National Park (MNP), Habitat Protection (HP) and General Use (GU) zones, and were highest in HP zones. Species differences were investigated and the dominant species of each type of zone were identified.

## Introduction

The Great Barrier Reef (GBR) is the world's largest coral reef ecosystem, containing ~3000 individual coral reefs within an area of 345,000 km<sup>2</sup>. Its outstanding universal values were recognised by World Heritage listing in 1981. GBR reefs have been classified as the world's least threatened (De'ath and Fabricius, 2010) due to their distance from the relatively small human population centres, and strong legal protection through the implementation of a comprehensive zoning system (Commonwealth of Australia, 2003). Despite these advantages relative to other coral reef systems, the GBR has experienced a decline in coral cover from 28.0% to 13.8% (0.53% yr<sup>-1</sup> from 1985–2012, a loss of 50.7% of initial coral cover (De'ath, 2012). This loss has been attributed to large-scale acute disturbances, especially tropical storms, bleaching, and population outbreaks of the coral-eating crown-of-thorns starfish *Acanthaster planci* (COTS) which may also increase in frequency and intensity in response to human activities (Fabricius et al., 2010).

In order to effectively manage the GBR, we need to understand the historic changes and current



state not only in terms of the abundances of reef and inter-reef biota, but also in terms of the temporal and spatial change in diversity. The concepts and tools necessary to undertake such an investigation have been developed as part of this project. The principal methodology is the multinomial diversity model (MDM, (De’ath, 2012)), a new method for relating diversity to complex environmental, spatial and temporal predictors.

Based on the MDM, the objectives of this work were three-fold: (1) to investigate spatial patterns of diversities of inter-reef fish species for the whole GBR, and (2) to relate these patterns to (a) spatial location and environmental drivers, and (b) the GBR zoning plan.

In the final phase of this project, a comprehensive analysis of the spatial and temporal change of reef benthos and fish will be undertaken in relationship to the GBR zoning.

## What is Diversity?

Species diversity is a key ecological concept for the understanding and management of ecosystems. It is defined as the effective number of species, or other types (e.g. genera, families), that are represented in a collection of individuals. For simplicity we refer simply to species in this description. The diversity of a collection depends on the number of species and their proportional abundances (PAs). Diversity increases with (1) the number of species and (2) the evenness of their distribution, and for a given number of species, diversity is maximised when all species are equally abundant. Diversity is expressed as the ‘effective number of species’, which refers to the number of equally-abundant species needed to obtain the same mean PA as that observed in the data of interest.

Diversities also have the characteristic of order, which is simply a relative weighting of abundances. Species can be weighted by raising the PAs to the power  $q$ , called the *order of diversity*, and although  $q$  can take any value; 0, 1 or 2 are typically used. The value 0 converts PAs to presence-absence, 2 gives greater weight to more abundant species and 1 weights all species equally. Diversities of orders 0, 1 and 2 are typically referred to as richness, Shannon and Simpson, respectively.

Although high diversity is generally a desirable property of an ecosystem, this is not always the case. For example, increasing prevalence of ‘undesirable’ species may increase the diversity of a community, but it will usually not increase its ‘value’ or ‘health’. Thus the choice of species is central to meaningful analyses.

The multinomial diversity model (MDM, Appendix 1) is a new method for relating Shannon diversity to complex environmental, spatial and temporal predictors (De’ath, 2012). It is based on a parameterised formulation of Shannon entropy and diversity, and a novel link between entropy and the log-likelihood of the multinomial model. The parameterised forms of diversity of orders 0, 1 and 2 are used in this work and are denoted as  ${}^0D$ ,  ${}^1D$  and  ${}^2D$ .  ${}^1D$  is mathematically identical to Shannon diversity, but we use this label to indicate its role in the MDM.

The parameteric Shannon entropy  ${}^1H$  and diversity  ${}^1D$  of  $N$  sites each with  $S$  species, propor-

tional abundances  $p_{ij}$  and parameters  $\pi_{ij}$  are defined:

$${}^1H = - \sum_{j=1}^N \sum_{i=1}^S p_{ij} \ln \pi_{ij}; \quad {}^1D = \exp({}^1H) \quad (1)$$

Entropies are additive, and this is the scale of our calculations. Diversities are simply the exponential of entropies and hence are multiplicative. Diversity is the scale on which we express results as the effective number of species, thereby providing meaningful ecological interpretation. For model interpretation, such as assessing the relative effects of different predictors, or the influence of individual sites or species, either entropy or diversity (or both) can be used, depending on the objectives and perspective of the analyst.

The MDM relates diversity to the predictors by minimising the diversity of the estimated species values. Model effects can be expressed as changes in diversity. Diversity can be partitioned within and between sites, species and models, and changes in diversity can be attributed to model predictors. This greatly enhances our capacity to model complex data sets, yet also provide simple interpretations.

Parameterised diversities are equivalent to their traditional equivalents in the way they reflect more or less abundant species. For example,  ${}^0D$  is equivalent to richness and  ${}^2D$  is equivalent to diversity based on Simpson's index. They also have the same interpretations in terms of the effective number of species, but they differ from their traditional counterparts that are limited to simple hierarchies in that they can be used within the MDM to relate diversity to complex predictors.

By formulating diversity as a statistical model, diversity is simplified both conceptually and analytically, and diversity analyses are extended beyond traditional simple hierarchies of  $\alpha$ ,  $\beta$ ,  $\gamma$  and measures of turnover. The MDM inherits the properties of generalised linear models, and thus proven methods can be used for model selection, and graphical and numerical interpretation. A weighted version of the Shannon diversity model is used to extend the MDM to diversities of orders 0 and 2 (richness and Simpson; Appendices 1 & 3).

## Data

From 2003–2005, inter-reef fish were surveyed using a fleet of Baited Remote Underwater Video Stations (BRUVS) (Cappo et al., 2007). Biologically informed stratification of major physical variables was used to sample as many different habitat types as possible. BRUVS were used to provide fish counts from 365 sites (Fig. 1) and 39,989 individuals from 347 species of vertebrates were recorded. The bony fishes were from 10 Orders, dominated by Perciformes (267 species), Tetraodontiformes (27), Anguilliformes (6), Aulopiformes (3), Scorpaeniformes, Clupeiformes, Beryciformes with 2 species, and Siluriformes, Pleuronectiformes and Gasterosteiformes each with a single species. The chondrichthyans were well represented by the Carcharhiniformes (15 species), Rajiformes (13) and Orectolobiformes (3). There were 5 species of sea snakes

from the Family Hydrophiidae. This work focuses only on species, but subsequent work will investigate how diversities vary with taxonomic and functional hierarchies. That will enable us to address questions such as "Do species within an given genus or family respond similarly to an environmental diver(s)?"

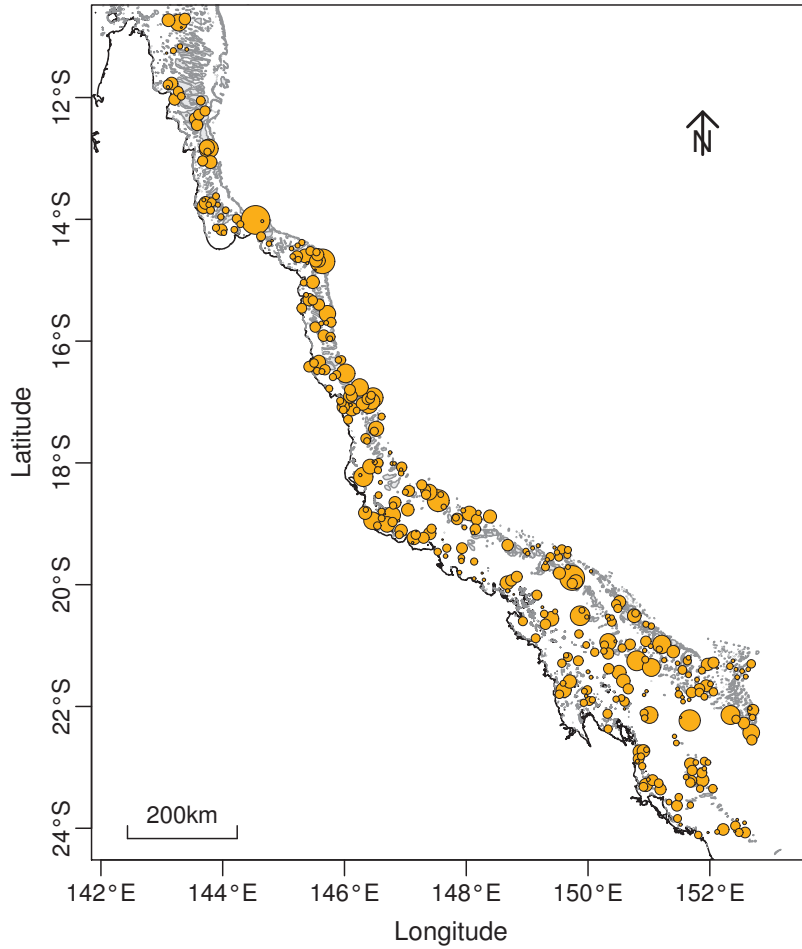


Figure 1: The locations of the fish surveys on the Great Barrier Reef. The diameters of the circles are proportional to the BRUVS site diversities (range 1.2–24.3), and show strong spatial variation but also local correlation. Inshore sites have lower site diversity than offshore sites, other than in the southern regions where highest diversity occurs on the mid-shelf.

These fish data are highly variable and sparse. A total of 347 species were observed across the 365 sites. Mean site abundance was 115.2 (range: 1–5842) and the mean abundance of individual species was 0.316 (range: 0–1440). The mean number of species per site was 14.6 (range = 1–192) and thus the average probability of a species being observed was 0.04 (i.e. 96% of the species data were zeros). Coupled with the extreme variability of the observations, this sparseness results in difficulties for any data analysis.

## Methods

All diversity modelling in this work is based on the MDM (De’ath, 2012). The R statistical software package (R Core Team, 2013) and the R package MDM (De’ath, 2011) were used for all data analyses.

## Spatial Distribution of Site Abundances

We first use univariate statistical smoothers to model the spatial distribution of species site abundance (mean = 109.5, range = 8, 1485). The maximum abundance occurs in the mid-shelf at 18°S to 20°S, whereas lowest abundances occur on the inner and outer regions from 16°S to 18°S.

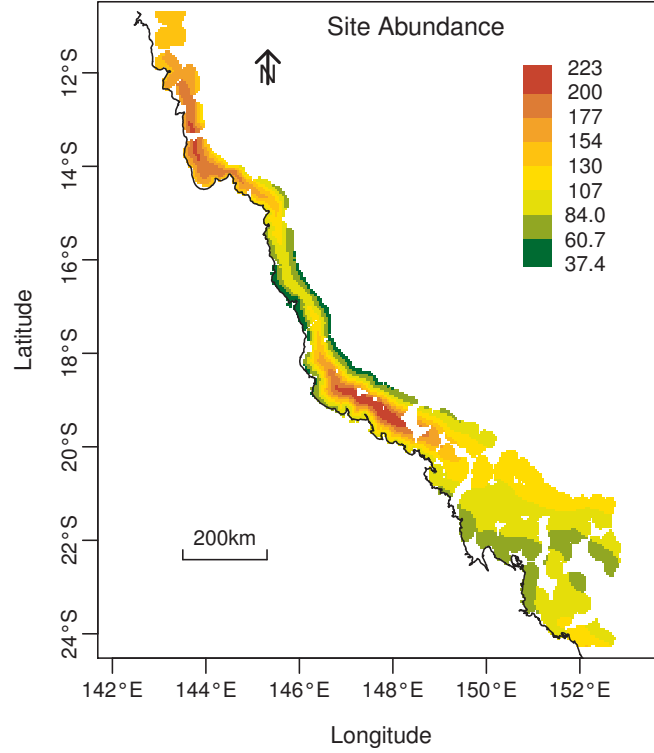


Figure 2: Spatial distribution of predicted site total abundance of inter-reef fish on the GBR. The spatial coordinates of the models were based on relative distances across and along the GBR. Predictions are limited to locations with a precision of  $<2$  standard errors.

The spatial distribution of abundances provides useful information, however it tells us nothing about local species turnover or local total diversity. By local, we mean averaged at a spatial scale chosen to most effectively represent the local total diversity that comprises both local site and turnover diversities. This is a major advantage of the new formulation of diversity and its link to the multinomial model (MDM). We can estimate local site and total diversities, and turnover of diversity, at any spatial scale. The predictors are not restricted to spatial models, and we can also relate diversity to environmental drivers or any other predictors, and combine spatial and environmental predictors in a single model to determine the best predictors of diversity. All spatial and environmental effects can also be easily interpreted in terms of numbers of effective species (Appendix 1).

# Spatial and Environmental Drivers of Diversity

## Distributions of species and site abundances and presences

The site and species abundances and their cumulative site totals (Fig. 3A) show the dominance of the highly abundant sites and species. Site dominance is not an issue for diversity analyses since diversity is based on relative probabilities of occurrence, the site totals being standardised to sum to one (Appendix 1). It is an issue for species however. For presence-absence the dominance of the highly abundant sites and species is greatly lessened (Fig. 3A – D). Of the 365 species, 200 occurred on 5 or less of the 365 sites and accounted for only 7.6% of the total abundance and 4.0% of the total presences.

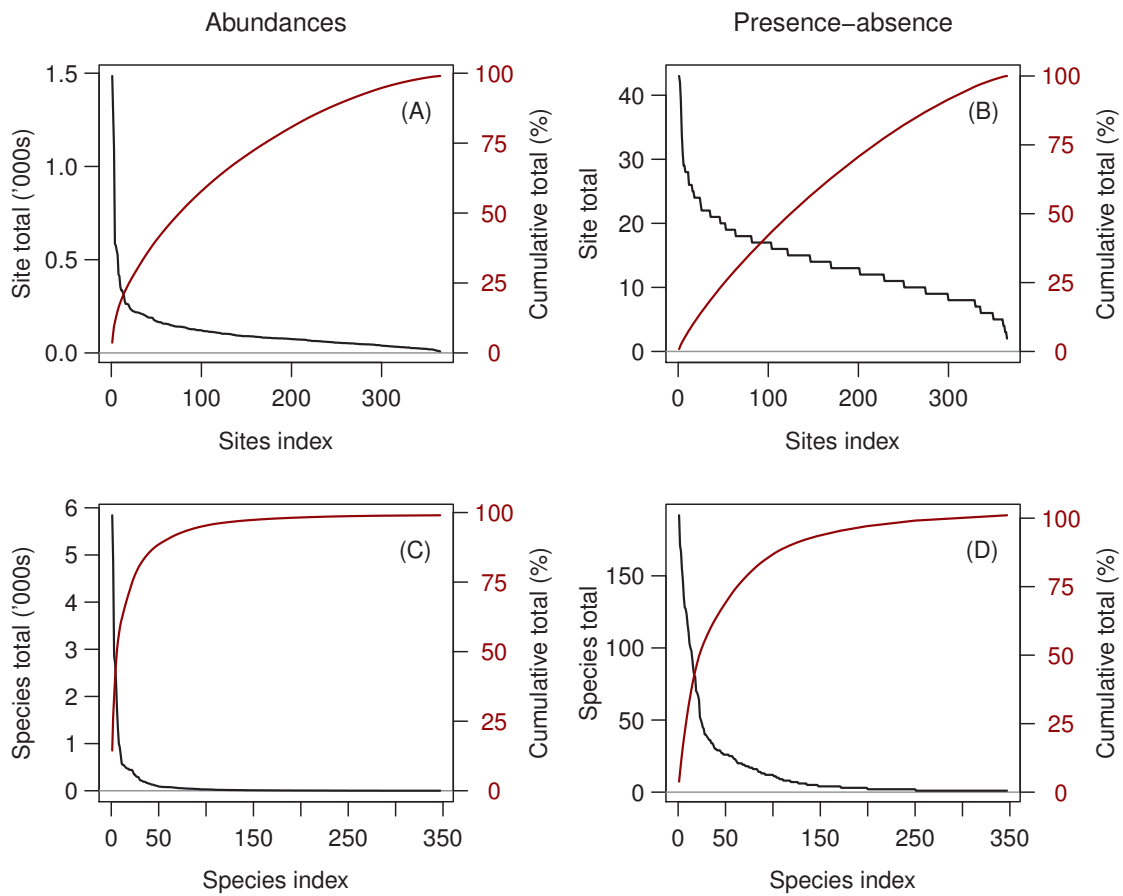


Figure 3: Individual species total abundances (A, C) and presences (B, D) (black) for sites (A, B) and species (C, D) and their accumulative values (red). The extreme distribution of fish counts is clearly apparent (A, C).

## Measures of diversity

Diversity of order 1 (Shannon) is arguably the most important and useful measure of diversity, and the MDM and traditional measures coincide, although we note that many analyses that claim to use Shannon diversity are actually using the index of Shannon diversity (De'ath, 2012). MDM and traditional measures of diversity are identical for order 1, and irrespective of whether

we use MDM or traditional measures of diversity, the values of diversity will always be ordered as  ${}^0D > {}^1D > {}^2D$  (Table 1).

| Diversity type       | ${}^0D$ | ${}^1D$ | ${}^2D$ |
|----------------------|---------|---------|---------|
| Total ( $\gamma$ )   | 103.4   | 55.3    | 34.5    |
| Turnover ( $\beta$ ) | 8.2     | 8.9     | 12.3    |
| Site ( $\alpha$ )    | 12.6    | 6.3     | 2.8     |

Table 1: The hierarchy of parametric diversities of  ${}^0D$ ,  ${}^1D$  and  ${}^2D$  of the fish species data.

Site diversity ( $\alpha$ ) is defined as the mean of the diversities calculated for each site, total diversity ( $\gamma$ ) is the diversity of the species means across the whole dataset, and turnover ( $\beta$ ) is the ratio  $\gamma/\alpha$  (Table 1, Appendix 1). The fact that  $\beta$  is a ratio of diversities, and not simply a diversity, is important.

# MDM Analysis of Inter-reef Fish Data

We begin the MDM analyses by contrasting it with the traditional hierarchy (Table 1). The MDM analysis of entropy and diversity (Table 2) illustrates the relationships between entropies and diversities.

| Model    | ${}^1H$ | $\Delta^1H$ | ${}^1D$ | $\Delta^1D$ |
|----------|---------|-------------|---------|-------------|
| Constant | 4.01    |             | 55.3    |             |
| Sites    | 1.83    | 2.18        | 6.3     | 8.9         |

Table 2: The analysis of entropy and diversity table.

Note that  ${}^1D$  and  $\Delta^1D$  are the exponentials of  ${}^1H$  and  $\Delta^1H$  (Eq. 1). Entropies are additive and the change in  $\Delta^1H$  between the models is the arithmetic difference, whereas diversities are multiplicative and hence the change in diversity  $\Delta^1H$  is the ratio. For each model,  ${}^1H$  and  ${}^1D$  represent mean site entropy and diversity. Conversely,  $\Delta^1H$  and  $\Delta^1D$  represent the change in entropy between two configurations of PAs and thus represent turnover. This distinction between the entropy (diversity) of a single model and the difference (ratio) in entropies (diversities) of two models is central to the interpretation of diversity analyses using the MDM.

## Spatial distribution of diversities

We fitted a hierarchy of three MDM models to the species data: (1) the sites model ( $\alpha$ ) – fits the species data exactly, (2) the spatial model fits an 8 degrees of freedom 2-dimensional smoother (across–along) to each species, and (3) the constant model ( $\gamma$ ) fits a constant for each species across all sites. This was done for diversity orders 0, 1 and 2. This has the effect of partitioning what was labelled  $\beta$  diversity (8.89, Tables 1 & 2) into the two ratios of diversity. Thus for  ${}^1D$  we have  $2.27 \cdot 3.89 = 8.89$  (Tables 1 & 3). This also applies to  ${}^0D$  and  ${}^2D$ .

| Model    | ${}^1H$ | $\Delta^1H$ | ${}^1D$ | $\Delta^1D$ | ${}^0D$ | $\Delta^0D$ | ${}^2D$ | $\Delta^2D$ | ${}^0D/{}^2D$ |
|----------|---------|-------------|---------|-------------|---------|-------------|---------|-------------|---------------|
| Constant | 4.01    |             | 55.3    |             | 103.4   |             | 34.5    |             | 2.99          |
| Spatial  | 3.19    | 0.82        | 24.4    | 2.27        | 49.9    | 2.07        | 8.41    | 4.10        |               |
| Site     | 1.83    | 1.36        | 6.26    | 3.89        | 12.6    | 3.96        | 2.81    | 2.99        | 4.48          |

Table 3: Analysis of entropy diversities of orders 0, 1 and 2. For each order of diversity three models are fitted: total, spatial and site. Columns contain the models, the entropy ( ${}^1H$ ) and change in entropy  $\Delta^1H$  between models, the diversity  ${}^1D$  and ratio  $\Delta^1D$ , the diversity  ${}^0D$ , ratio  $\Delta^0D$ , the diversity  ${}^2D$ , and ratio  $\Delta^2D$  and the ratio  ${}^0D/{}^2D$ .

The spatial models of orders 0, 1 and 2 explain 34.6%, 37.6% and 56.2% of the total entropy of the species data. The significance of the spatial model can be assessed with permutation tests, and for diversities of all orders, the spatial effects were significant ( $P < 0.001$ ). Given the extreme species distributions, their high variability and the large numbers of absences, these MDM models are surprisingly effective predictors of the species abundances.



The entropies and diversities of the various models need to be labelled for the purpose of description and the following convention will be adopted. The constant and site models will be referred to as "Total" and "Site" diversity, and the diversities of other models will be referred to by their predictors, in this case "Spatial". Changes in entropies and diversities will be referred to by the two names of the model that define the change, e.g. "Constant–Spatial" turnover or "Spatial–Site turnover, with the term "turnover" indicating that they are related to turnover of species between the given pair of models.

## Visualising the diversity model

### Order 1 diversity

Fitting the MDM spatial model provides predictions of site, spatial and total diversities. The three predicted diversities over the GBR are shown in Fig. 4. At each point in space, turnover is estimated by the local total diversity / site diversity. The regions a–d were chosen to comprise relatively consistent patterns of diversity across all orders of diversity. Site diversity increases with distance from the coast in a–b, but in c–e a modal cross shelf distribution is apparent. For total diversity, the pattern is again increasing cross-shelf, but this extends through a–d with a flat cross-shelf distribution in e. The variation in total diversity is ten-fold, far greater than that of site diversity.

The five regions can be summarised as:

- a** : moderate site diversity weakly increasing cross-shelf; lowest turnover and total diversity
- b** : highest site diversity; high variation in total diversity; strong increasing cross-shelf gradients
- c** : weak modal site diversity; increasing cross-shelf total and site diversities with highest turnover (6-fold) offshore
- d** : weakest cross-shelf gradients and few extreme diversities
- e** : lowest and convex cross-shelf distribution of site diversities; low and even distribution of total diversity

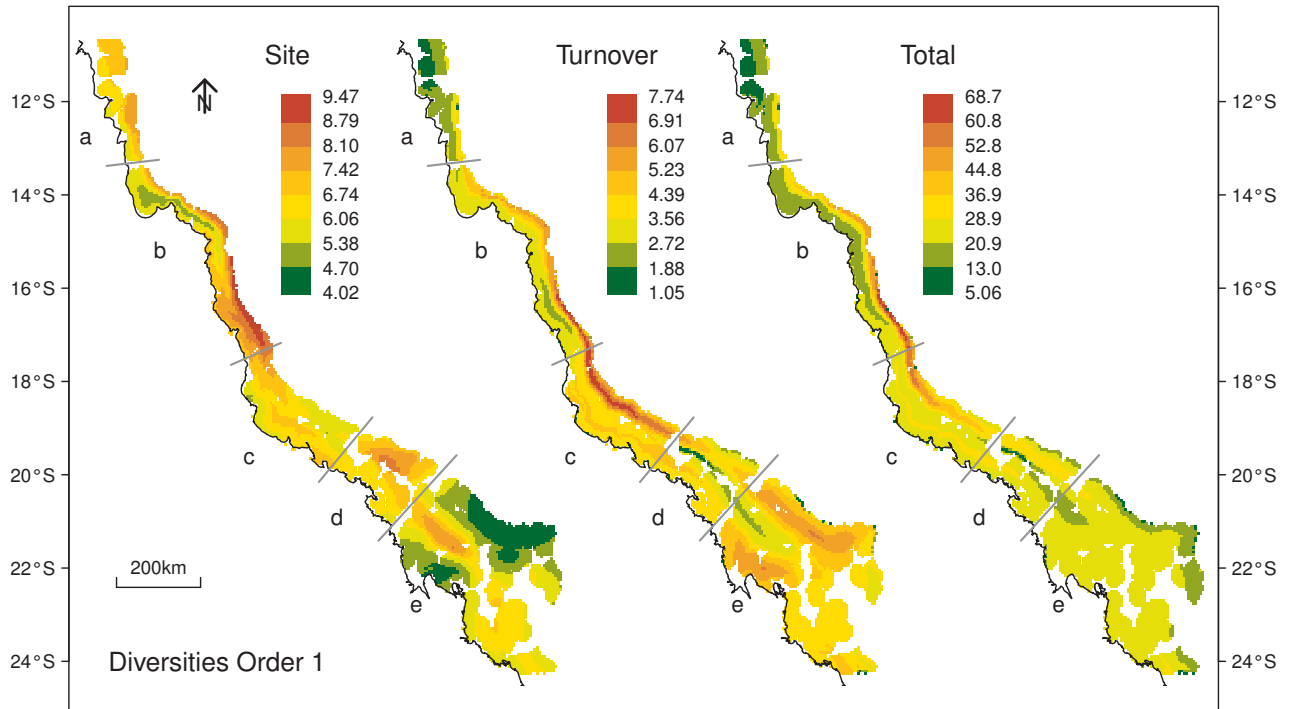


Figure 4:  $^1D$  diversity of inter-reef fish on the GBR. The three panels represent site, turnover and local diversities. Within each map, the 5 regions (a-e) are defined by regional similarities in spatial patterns of diversity along the GBR.

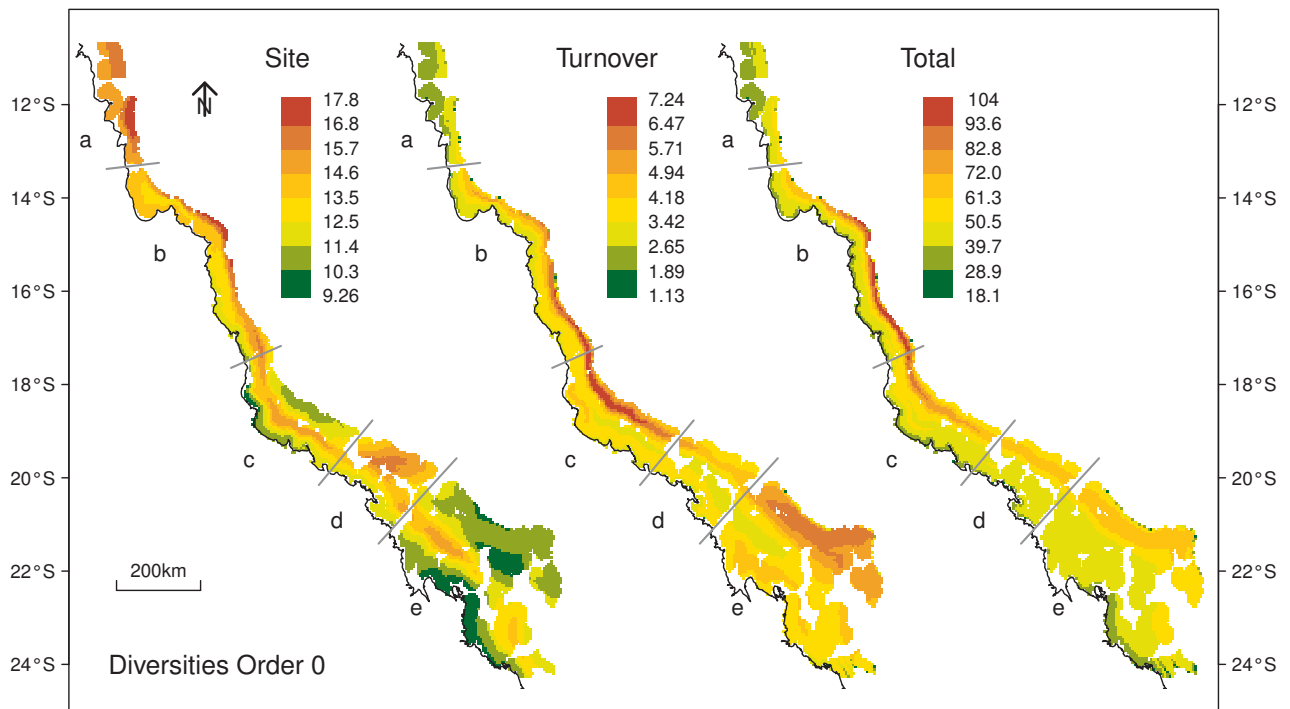


Figure 5:  $^0D$  diversity of inter-reef fish on the GBR. The three panels represent site, turnover and local diversities. Within each map, the 5 regions (a-e) are defined by regional similarities in spatial patterns along the GBR.

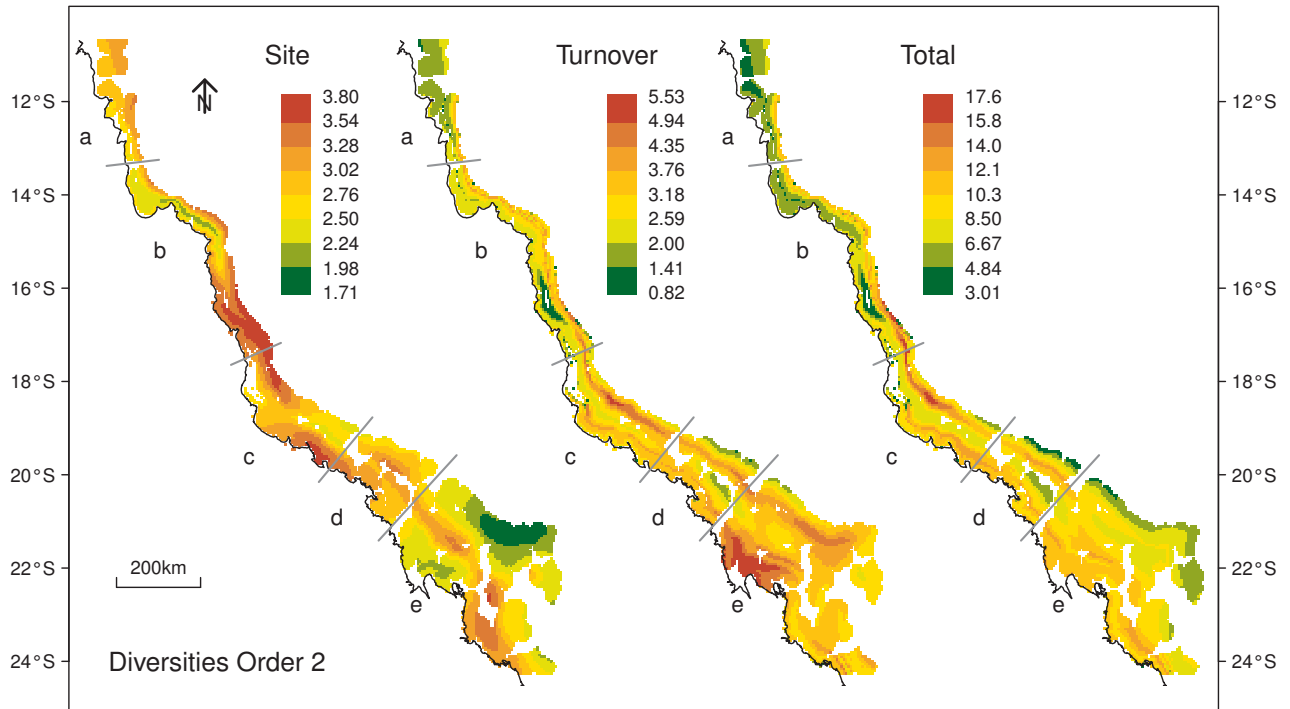


Figure 6:  $^2D$  diversity of inter-reef fish on the GBR. The three panels represent site, turnover and local diversities. Within each map, the 5 regions (a-e) are defined by regional similarities in spatial patterns along the GBR.

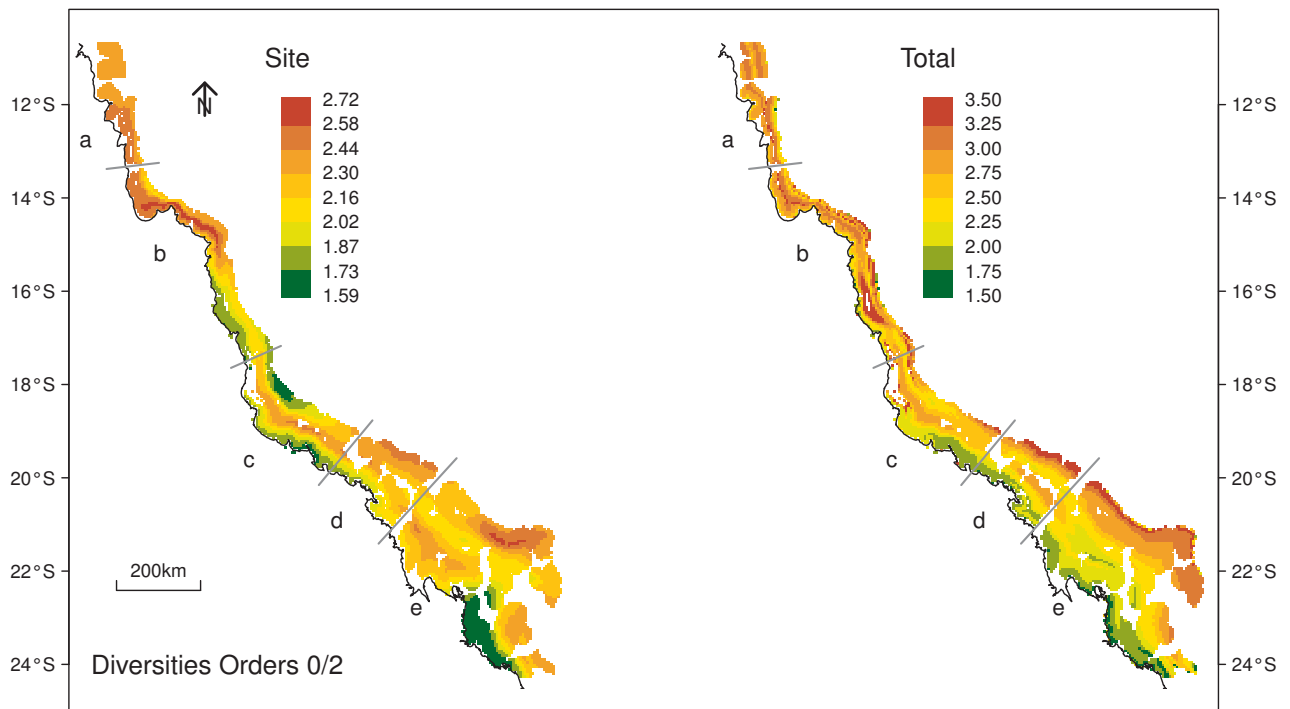


Figure 7:  $\log_2$  ratios ( $^0D/^2D$ ) of diversity of inter-reef fish on the GBR. The two panels represent site, turnover and local diversities. Within each map, the 5 regions (a-e) are defined by regional similarities in spatial patterns along the GBR.

## Order 0 and 2 diversities

${}^0D$  is determined solely by presence-absence of species, whereas  ${}^2D$  is largely determined by the high abundance species. Examining these orders of diversity and comparing to them each other, and to  ${}^1D$ , can be highly informative as to how the relative abundances vary across species.

The spatial patterns of  ${}^0D$  do not differ greatly from those of  ${}^1D$ , whereas spatial patterns of  ${}^2D$  show greater differences to  ${}^1D$  (Figs. 5, 6). The high levels of diversity in b extend across the shelf and there are relatively higher levels of diversity on the inshore in e. The latter is reflected in high turnover in inner e. The cross-shelf gradients in total diversity also differ with lower value in the outer shelf for d-e.

A simple but effective way to enhance the interpretation of differences across the orders of diversity is to map the ratios  ${}^0D/{}^2D$ . These ratios will usually be  $> 1$  and high ratios correspond to communities dominated by a few species and low ratios correspond to an even distribution of species. We express this on the  $\log_2$  scale as a one unit increase (decrease) on this scale represents a doubling (halving) of turnover. These are presented for site and total diversities (Fig. 7) and are highly informative. For site diversity, low values ( $< 1.8$ ) of the diversity indicate relatively even distributions in three inshore and one offshore regions, whereas higher values ( $> 2.25$ ) occur in the north and the southern outer shelf. For total diversity, the patterns are somewhat simpler with a even distribution across the northern shelf and a cross shelf gradient in the south with low inshore and high offshore turnover.

This approach to modelling diversity can provide a wealth of information on the spatial distribution of the three types for diversity. It can (1) estimate of each of the three types of diversity at each point in space, (2) identify diversity gradients and regions of uniformity and (3) estimate the uniformity of the species abundance by the ratios  ${}^0D/{}^2D$ .

## Spatial distribution of selected species

The MDM estimates the proportional abundances (PAs) individual species, and hence we can also map them for any given species. These are shown for *N.furcosus*, *E.naucrates*, *S.queenslandicus*, *D.russelli*, *P.nagasakiensis* and *S.leptolepis* (Figs. 8, 9). The spatial distributions vary greatly across the 6 species, but all vary most strongly across the shelf. *E.naucrates* has the broadest cover of the 6 species though it is rare on the outer shelf other than in the far north. *S.queenslandicus* and *S.leptolepis* favour in-shore environments, *N.furcosus* and *D.russelli* favour the mid-shelf and *P.nagasakiensis* favours the inshore.

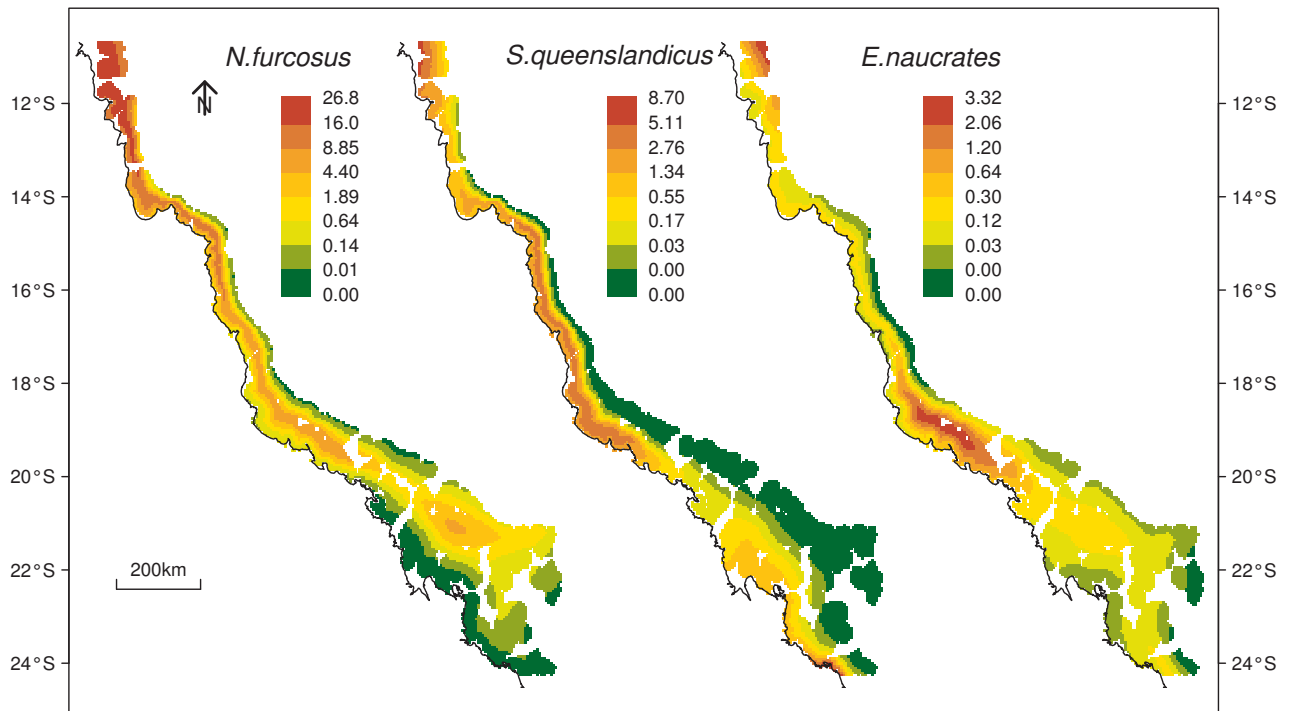


Figure 8: Spatial distribution of the estimated proportional abundances (%) of *N.furcosus*, *E.naucrates* and *S.queenslandicus*. The strong cross-shelf patterns underpin the similar patterns observed in the diversity maps. Legend scales are on a logarithmic scale to better represent the low abundance areas

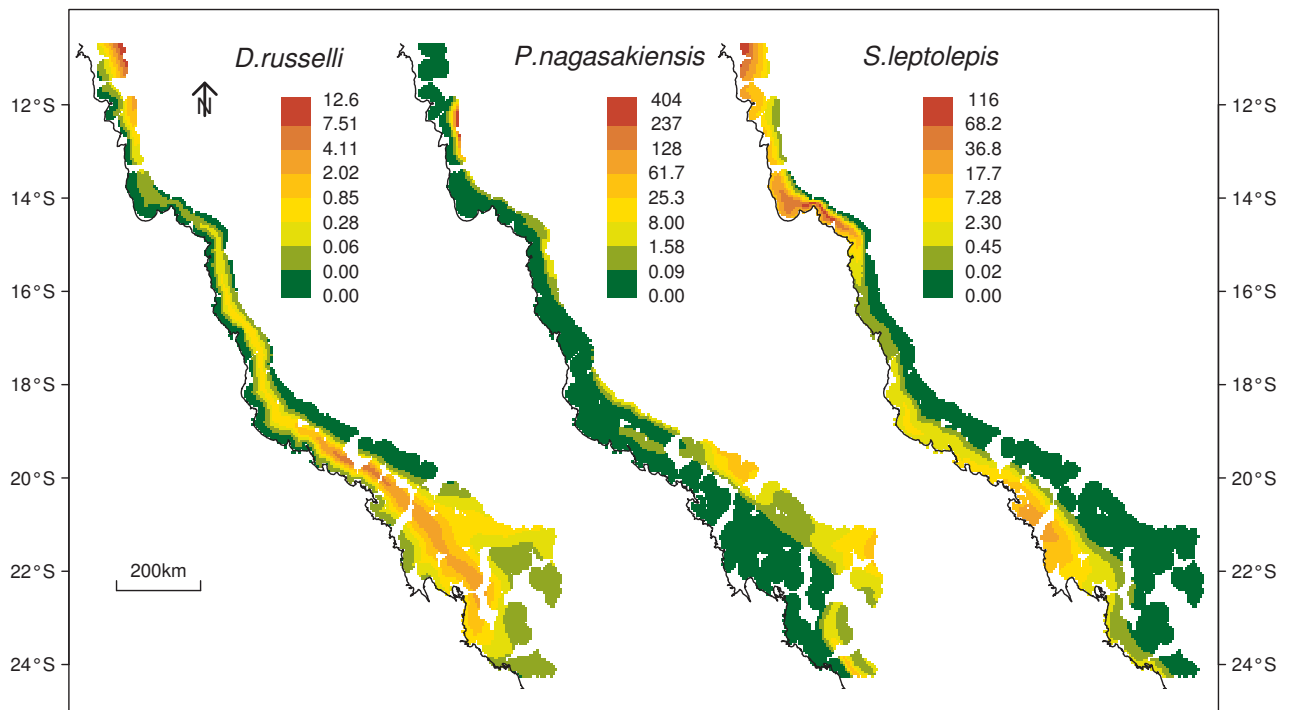


Figure 9: Spatial distribution of the estimated proportional abundances (%) of (*D.russelli*, *P.nagasakiensis* and *S.leptolepis*). The latter two species have highly localised high populations. Legend scales are on a logarithmic scale to better represent the low abundance areas

## Effects of environmental drivers

The environmental drivers of diversity were selected on their predictive ability. Four of the 38 drivers were selected and their spatial distribution and effects on diversity were quantified and are illustrated below.

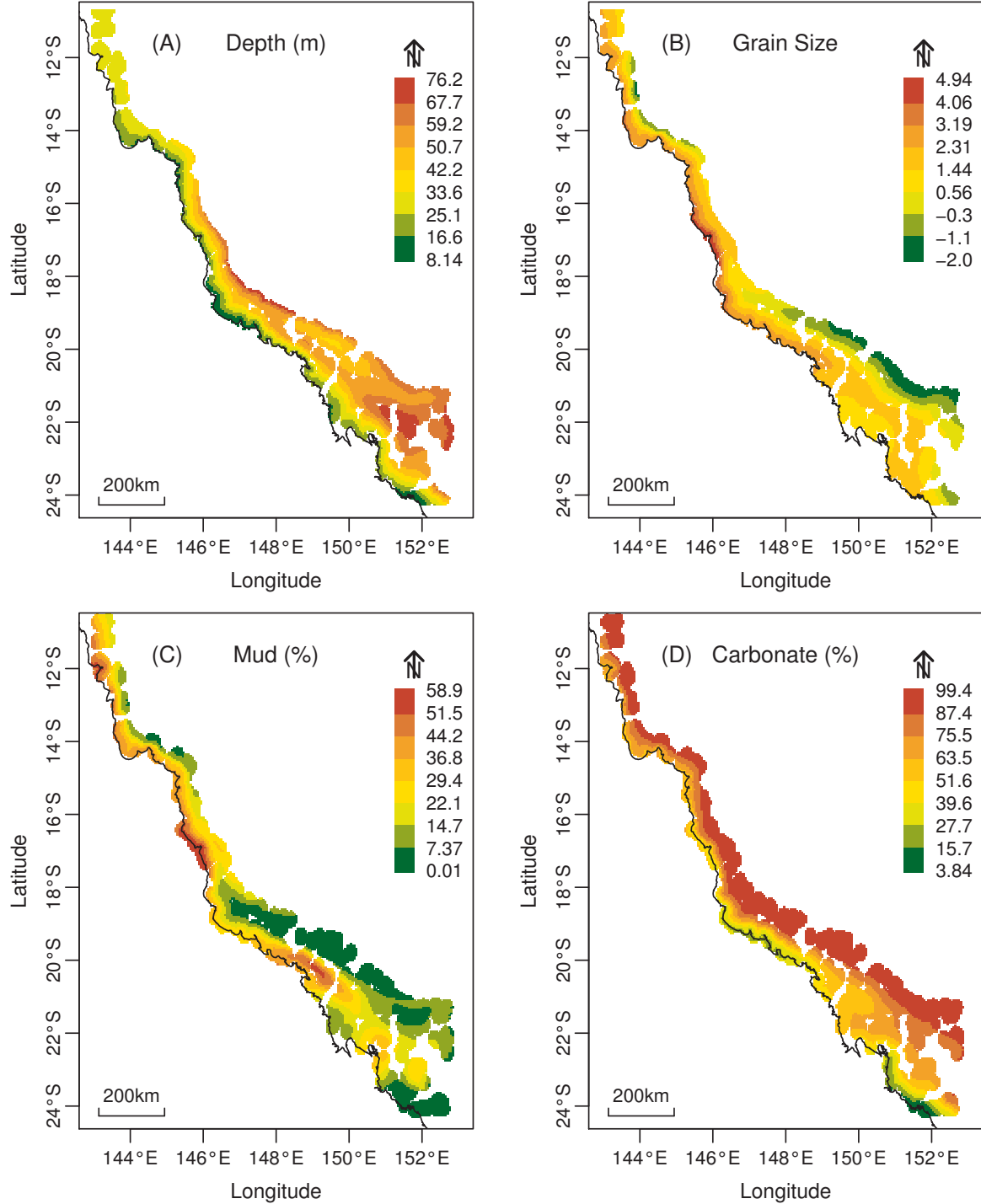


Figure 10: Spatial distribution of depth (A) and the three environmental drivers, grain size (B), mud (C) and carbonate (D).

The spatial distribution of the 4 environmental drivers depth of grain size, mud and carbonate (Fig.10) change predominantly across the shelf. Shallower depths are more prevalent in the

far northern region. Grain size varies greatly at local scales but small particles dominate the off-shore. Mud dominates the inshore in the central and northern shelf, with little mud in the offshore southern region. Carbonate is high over most of the shelf other than on the inner shelf and parts of the southern central shelf.

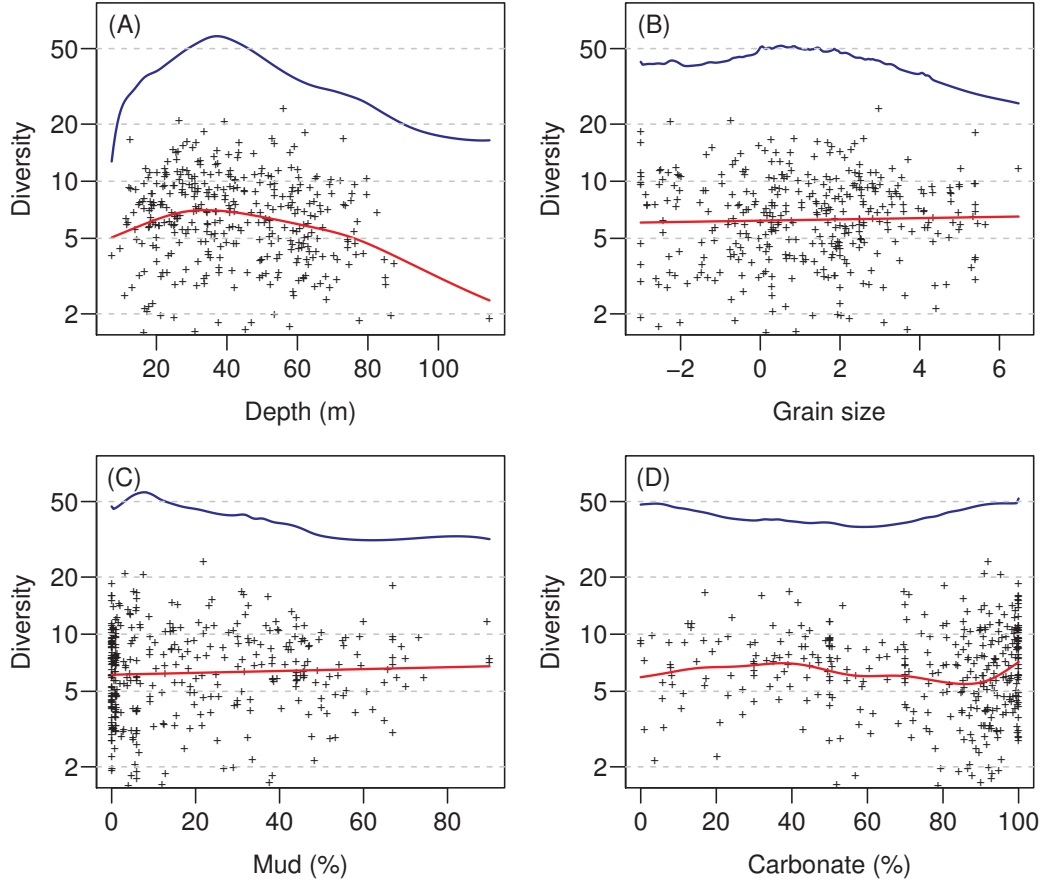


Figure 11: Distribution of  $^1D$  depth (m) (A), grain size (B), mud (%) (C) and carbonate (%) (D). Site diversity (+), smoothed site diversity (red) and environmental driver diversity (blue) along the 4 environmental gradients. The vertical distance between the red and blue lines represents local turnover.

The effects of the 4 environmental drivers for  $^1D$  show how site diversity (+), smoothed site diversity (red) and driver diversity (blue) vary along the gradients of depth, grain size, mud and carbonate (Fig. 11). Turnover is represented by the vertical distance between the driver and site diversity. The effects are plotted on the entropy scale (log of diversity) and thus a unit change on the vertical corresponds to a fixed proportional change in diversity (e.g. 5–10 is equivalent to 10–20).

Changes in diversities are most pronounced along the depth gradient (depth diversity: range = 12.7–58.1 and site diversity: range = 2.4–7.0), with maximum and minimum values of depth diversity occurring at depths of 37 and 7m. Smoothed site diversity and depth-site turnover are maximised at depths of 33 and 37m, close to the maximum depth diversity. For grain size, the range of diversity was 25.7–51.8 with highest values at average grain sizes and lowest values for large grains. Smoothed site diversity had negligible variation and averaged 6.3 (range 6.1–6.5), and thus turnover covaried with grain size diversity. Mud diversity declined with increasing

mud, other than for a small rise at levels of mud  $<7\%$ . Smoothed site diversity increased linearly with mud and hence turnover was highest at 7.6% and lowest at 90% mud cover. Carbonate diversity was lowest ( $\sim 37$ ) at intermediate levels of carbonate and maximised ( $\sim 51$ ) at both lowest and highest levels.

## Joint effects of spatial and environmental drivers

The environmental effects account for 36.7% of the total entropy,  ${}^1H$ , and the joint effects of the environmental and spatial drivers account for 51.4% (Table 4). If the spatial effects are added before the environmental drivers, they account for the 34.9% of the total entropy. Thus 16.5% ( $51.4 - 34.9$ ) of the entropy can uniquely accounted for by environment, and 14.7% ( $51.4 - 36.7$ ) can uniquely accounted for by spatial. The common entropy, explainable by both environment and spatial drivers, is thus 20.2% The degrees of freedom of the environment and spatial drivers are just 4.38% of the total and thus these effects, either jointly or separately, are highly significant.

| Model               | ${}^1H$ | $\Delta^1H$ | ${}^1D$ | $\Delta^1D$ | P        |
|---------------------|---------|-------------|---------|-------------|----------|
| Constant            | 4.01    |             | 55.3    |             |          |
| Environment         | 3.21    | 0.80        | 24.8    | 2.23        | $<0.001$ |
| Environment+Spatial | 2.89    | 0.32        | 18.0    | 1.38        | $<0.001$ |
| Site                | 1.83    | 1.06        | 6.26    | 2.88        |          |

Table 4: Analysis of entropy and diversities environmental and spatial models. Columns indicate the terms in the model, the entropy ( ${}^1H$ ) and change in entropy  $\Delta^1H$  between models, the diversity  ${}^1D$  and change in diversity  $\Delta^1D$ , and the significance of the permutation tests.

## Effects on individual species

The entropy of any fitted model is the sum of a sites by species matrix (Eq. 1) and we can decompose this matrix by its rows or columns to give use the sites or species components. We can also partition entropy (or its components) across models, to give us the entropies broken down by any subsets of sites, species or models.

For example, we can use these properties to show how species are explained by the spatial and environmental drivers. For the 10 most abundant species we have estimated and plotted the entropy components for 5 models (Fig. 12): total ( $\dashv$ ), site ( $\vdash$ ), spatial-environment ( $\circ$ ) spatial ( $\triangle$ ) and environment ( $\nabla$ ). The distances between the  $\dashv$  and each of  $\vdash$ ,  $\circ$ ,  $\triangle$  and  $\nabla$  are the entropies accounted for by the sites, spatial-environment, spatial and environment models for each of the 10 species. The spatial-environment model ( $\circ$ ) explains up to 50% of the entropy of species. For *N.furcosus* and *S.leptolepis* the spatial model explains more entropy than environment model, whereas for *N.theodorei* and *P.paradiseus*, environment accounts for more. For most of the species, the sum of the entropies explained by the spatial and environmental predictors exceeds that of the spatial-environment model. This is due to the correlation between the two sets of predictors. However, this is not always the case, with



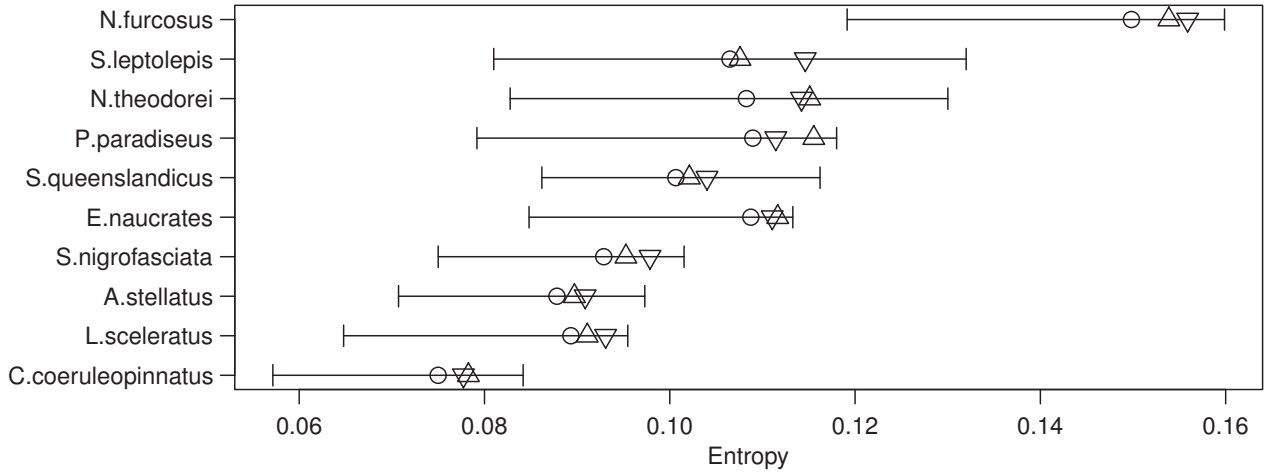


Figure 12: Entropies for the 10 most abundant taxa. The entropies of the 5 models are indicated: total (┼), site (┼), spatial-environment (○) spatial (△) and environment (▽). The horizontal distances between the ┼ and each of ┼, ○, △ and ▽ represent the entropies accounted for by the sites, spatial-environment, spatial and environment models for each of the 10 species.

the sum of the individual effects of the two predictors for *N.furcosus* and *P.paradiseus* being approximately equal to their joint effect.

This capacity to break down the entropy of a model, or a collection of models, in numerous ways in order to explore how individual or collections of species and sites relate to the spatial and environmental drivers of the model(s) represents a powerful and flexible toolbox for understanding the drivers of change in diversity.

# Assessment of GBR Zoning

One of the principal objectives of GBR zoning (Fig. 13) is to conserve biodiversity, and the MDM is a useful toolbox to assess the spatial and temporal effectiveness of such protected areas. The GBR protected areas are intended to be representative of the GBR as a whole, and thus they should contain as many of the species and environment types of the whole GBR as is possible, but also be distributed in a broad spatial pattern.

This analysis shows a way of assessing the effectiveness of the zoning to capture the spatial diversity of the fish data. Although these data are limited to a spatial assessment, later work will show how the temporal change can be addressed using the fish and benthos data of the AIMS Long Term Monitoring Program (LTMP).

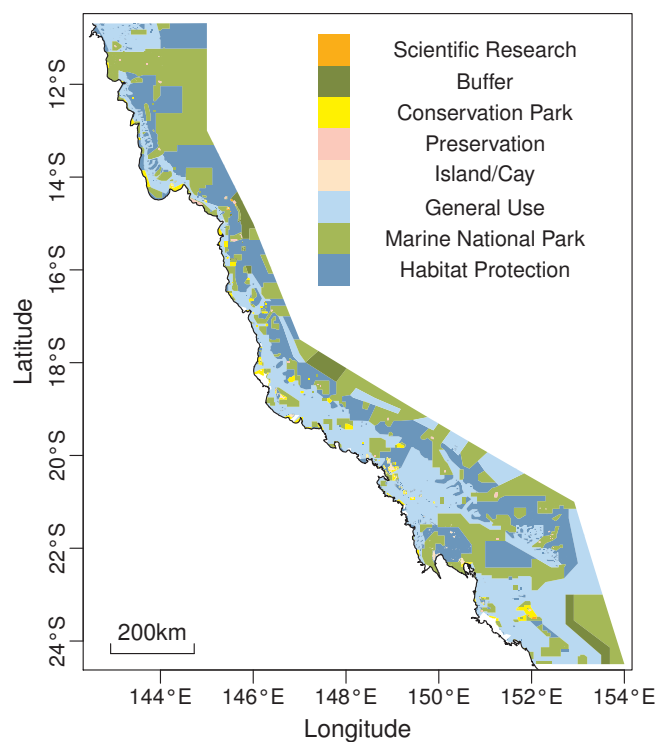


Figure 13: The Great Barrier Reef Marine Park zoning.

The survey sites gave good coverage of the Great Barrier Reef World Heritage Area (Fig. 1 & Fig. 13) and the three major zoning types – General Use, Habitat Protection and Marine National Park. Give the assigned usage of Conservation Park, Preservation and Buffer zones, the 7 sites located in these zones were added to the Marine National Park group, giving site totals for the 3 zoning types of 165 (General Use; GU), 123 (Habitat Protection; HP) and 70 (Marine National Park; MNP).

The diversities  $^0D$ ,  $^1D$  and  $^2D$  were calculated for all sites, MNP, HP and GU combined and HO and GU separately (Table 8) and their distributions are shown in Fig. 14. The analysis of  $^1H$  and  $^1D$  (Table 5) show a small non-significant difference between MNP and the pooled HP – GU sites, but a significant difference between the three groups MNP, HP and GU. The latter is due to the higher diversity of HP (Table 3). The comparison of MNP, HP and GU based on

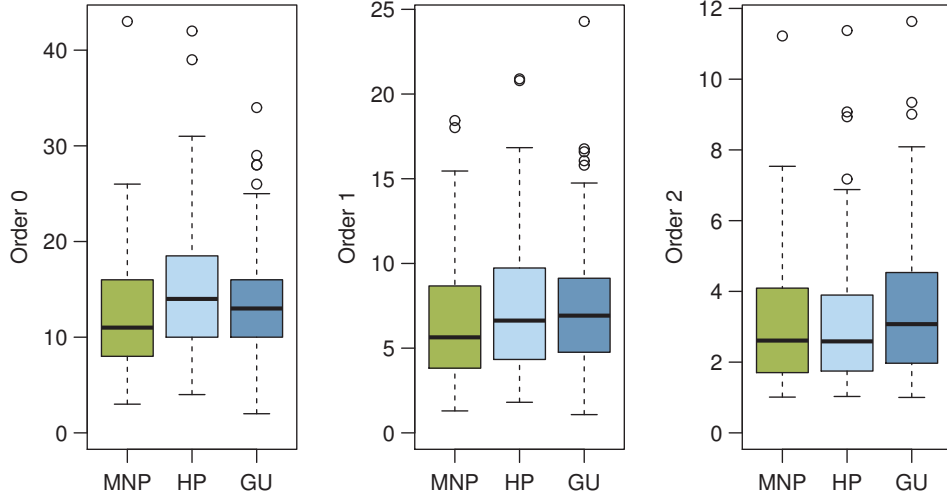


Figure 14: Diversities of orders 0, 1 and 2 for the three zones: Marine National Park (MNP), Habitat Protection (HP) and General Use (GU).

${}^0D$  shows a higher relative diversity in HP compared to MNP and GU. This suggests that HP and GU have many more rare species than MNP, but that the number of more common species differs less between the 3 groups. Site diversities were highly variable across all groups (Fig. 14), and were  $\sim 20\%$  higher for both HP and GU compared to MNP (Table 3).

| Model           | ${}^1H$ | $\Delta^1H$ | ${}^1D$ | $\Delta^1D$ | P    |
|-----------------|---------|-------------|---------|-------------|------|
| Constant        | 4.01    |             | 55.3    |             |      |
| MNP vs HP – GU  | 3.97    | 0.04        | 52.9    | 1.05        | 0.67 |
| MNP vs HP vs GU | 3.83    | 0.14        | 46.0    | 1.15        | 0.03 |
| Site            | 1.83    | 2.00        | 6.3     | 7.35        |      |

Table 5: Analysis of entropy and diversity of order 1 and significance of model comparisons. The 4 fitted models are: (1) constant, (2) comparison of MNP and HP – GU, (4) 3-way comparison of MNP, HP and GU, and (4) site. Tests of significance are based on permutation tests (n=1000).

| Diversity type       | ${}^1D$ |      |         |      |      | ${}^0D$ |      |         |       |      |
|----------------------|---------|------|---------|------|------|---------|------|---------|-------|------|
|                      | All     | MNP  | HP – GU | HP   | GU   | All     | MNP  | HP – GU | HP    | GU   |
| Total ( $\gamma$ )   | 55.3    | 43.1 | 53.9    | 48.1 | 43.0 | 101.5   | 73.0 | 105.1   | 115.2 | 73.9 |
| Turnover ( $\beta$ ) | 8.8     | 8.5  | 8.5     | 7.8  | 6.8  | 8.1     | 6.8  | 8.1     | 8.5   | 5.9  |
| Site ( $\alpha$ )    | 6.3     | 5.5  | 6.4     | 6.4  | 6.4  | 12.6    | 10.8 | 13.0    | 13.6  | 12.6 |

Table 6: Diversities ( $\alpha\beta\gamma$ ) of orders 1 and 0 for all sites, and for sites grouped by the four zones: Marine National Park (MNP), combined Habitat Protection and General Use (HP – GU), Habitat Protection (HP) and General Use (GU).

We can also identify dominant species of MNP, HP and GU (Fig. 15). In (A) the comparison is MNP *vs.* HP – GU. Species dominant in MNP include *D.russelli*, *S.queenslandicus*,

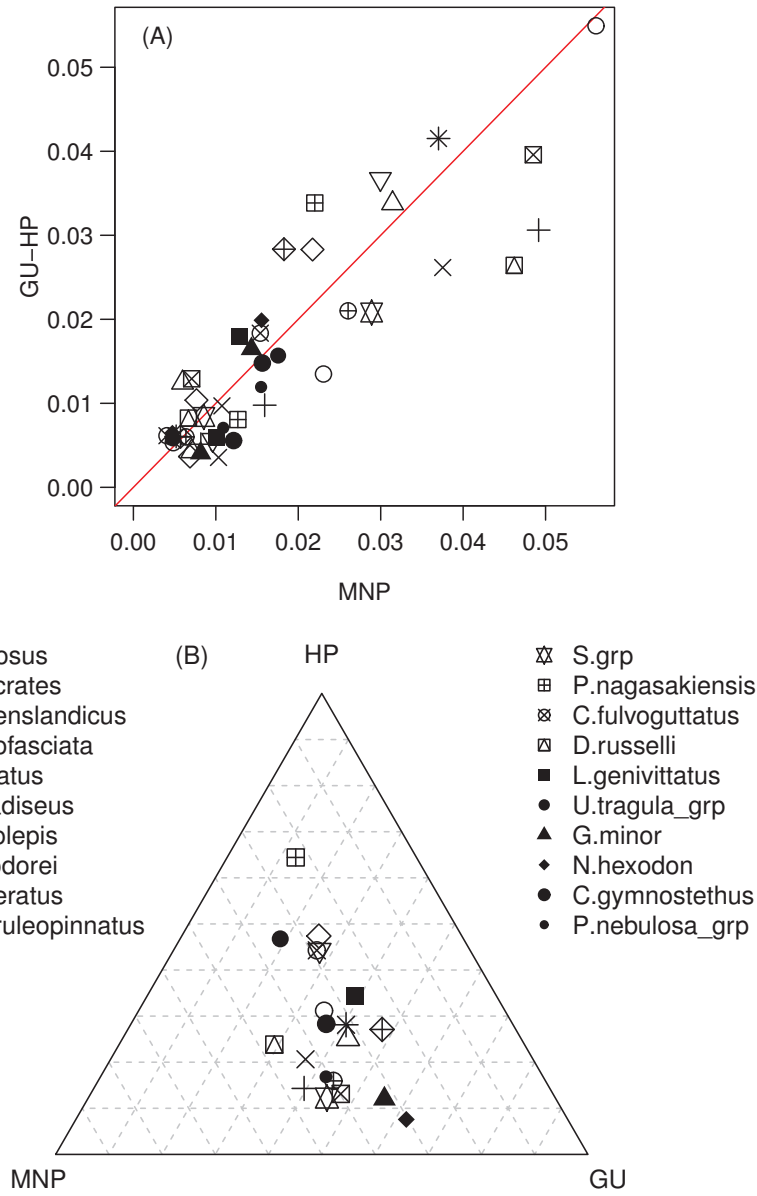


Figure 15: Identifying the dominant species for 2 or 3 groups: (A) MNP *vs.* HP – GU and (B) MNP *vs.* HP *vs.* GU.

*S.nigrofasciata* and *P.nagasakiensis* and in HP – GU *N.theodorei*, *P.nagasakiensis*, *L.scleratus* and *P.paradiseus*. In (B) the three way comparison is MNP *vs.* HP *vs.* GU. Species dominant in MNP include *S.queenslandicus* and *S.nigrofasciata*, in HP *P.nagasakiensis* and *A.stellatus*, and in GU *N.hexodon* and *G.minor*.

## Discussion

Diversity is a key concept for understanding and managing complex ecosystems. The multinomial diversity model (MDM) is a new statistical model of diversity, developed as part of this project, that takes the analysis and understanding of diversity to new levels. The MDM relates diversity to multiple predictors, and expresses model effects as changes in entropy and diversity. The entropies and diversities of a collection of MDMs can be partitioned across species, sites and predictors. This greatly enhances our capacity to model complex data sets, and yet also provide simple interpretations. The MDM is a generalised linear model. These models are based on established statistical theory, have a proven track record in applied data analysis for more than 35 years, and are the primary tools of professional statisticians and data analysts. The application of the MDM to studies of diversity will benefit from that basis and experience.

We have shown how the MDM can quantify and graphically represent spatial and environmental variation in diversity, thereby leading to better understanding of change in complex ecosystems. This model-based approach has enabled us to assess the relative effects of drivers, either singly or jointly, and to represent local effects along gradients in term of total, site and turnover diversities. This work focused on species data, but the MDM can also be used to assess the effects of drivers on hierarchies such as species-genera-families. This will be examined in subsequent work.

The MDM was also used to examine differences in diversity across the principal zonings of the GBR, and key species of these differences were identified. Subsequent work will extend this work to assess temporal change of LTMP fish and benthos data with respect to the zonings.

The remainder of this project will focus on two tasks:

1. Completion and publication of the inter-reef fish diversity presented in this work.
2. Analysis of spatial and temporal change of the Long Term Monitoring Program reef benthos and fish data.

In addition to the methods used in this work, we will also develop a method for estimating diversity within and between taxonomical hierarchies, i.e. address questions such as "How does diversity change from species to genera to families?"

## Acknowledgements

This work was supported by the Australian Institute of Marine Science, and the National Environmental Research Program of the Australian Government.

## References

- Cappo, M., G. De'ath, and P. Speare. 2007. Inter-reef vertebrate communities of the great barrier reef marine park determined by baited remote underwater video stations. *Marine Ecology Progress Series*, **350**:209–221.
- Commonwealth of Australia. 2003. Great Barrier Reef Marine Park Zoning Plan. Great Barrier Reef Marine Park Authority, Townsville, Australia: 211 pp.
- De'ath, G. 2011. MDM: Multinomial Diversity Model. R package version 1.1.
- De'ath, G. 2012. The multinomial diversity model: Linking shannon diversity to multiple predictors. *Ecology*, **323**:116–119.
- De'ath, G. and K. Fabricius. 2010. Water quality as a regional driver of coral biodiversity and macroalgae on the great barrier reef. *Ecological Applications*, **20**:840–850.
- Fabricius, K. E., K. Okaji, and G. De'ath. 2010. Three lines of evidence to link outbreaks of the crown-of-thorns seastar *Acanthaster planci* to the release of larval food limitation. *Coral Reefs*, **29**:593–605.
- R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

# Appendices

## A1: Diversity and the Multinomial Diversity Model

### Diversity

Ecological diversity can refer to many types of classification (e.g. taxonomic, functional, genetic) made at various levels of sampling. For simplicity, we refer to the distribution of  $S$  species across a collection of  $N$  sites. Diversity is defined in terms of species proportional abundances (PAs) and it has four key properties: (1) diversity can vary from 1 for a single species occurrence to  $S$  for equally abundant species, and is interpreted as the *effective number of species* (Tuomisto 2010a, 2011), or the *numbers equivalent* (Jost 2007); (2) diversity concurs with the widely accepted doubling principle (Hill 1973), which states that given a community with  $S$  species, then the addition of another  $S$  species with the same PAs should double the original diversity; (3) the diversity of a site only depends on the species present, and the inclusion of absent species does not alter its diversity; (4) species can be weighted by raising the PAs to the power  $q$ , called the *order of diversity*, and although  $q$  can take any value; 0, 1 or 2 are typically used. The value 0 converts PAs to presence-absence, 2 gives greater weight to more abundant species and 1 weights all species equally. Diversities of orders 0, 1 and 2 are referred to as richness, Shannon and Simpson, respectively.

In this work, we emphasise and clarify the distinction between diversity and entropy. Making this distinction also helps clarify previous confusion about these two concepts (Jost 2007). The concept of entropy is used in many disciplines, including physics, chemistry and economics (Borges and Roditi 1998; Havrda and Charvat 1967; Kaniadakis *et al.* 2005). Of the many forms of entropy, Shannon entropy is the best known and most widely used, and is central to this study. To emphasise their relationship, we define the Shannon entropy,  ${}^1H$ , and Shannon diversity,  ${}^1D$ , of a single site with proportional abundances  $p_i$ , by:

$${}^1H = - \sum_{i=1}^S p_i \ln p_i; \quad {}^1D = \exp({}^1H) \quad (2)$$

Entropy and diversity for orders  $q \neq 1$  can similarly be defined:

$${}^qH = \frac{1}{(1-q)} \ln \sum_{i=1}^S p_i^q; \quad {}^qD = \exp({}^qH) \quad (3)$$

where  ${}^qH$  is known as Renyi entropy. Entropies are additive, and this is the scale of our calculations. Diversities are simply the exponential of entropies for all values of  $q$ , and hence are multiplicative. Diversity is the scale on which we express results as the effective number of species, thereby providing meaningful ecological interpretation. For model interpretation, such as assessing the relative effects of different predictors, or the influence of individual sites or species, either entropy or diversity (or both) can be used, depending on the objectives and

perspective of the analyst.

Although Eqs. 2 & 3 define diversity for a single site, they do not define diversity for a collection of sites, nor do they tell us how to manipulate diversities. A common approach to this issue is through the concepts of  $\alpha$ ,  $\beta$  and  $\gamma$  diversities (Jost 2007; Tuomisto 2010ab). Following this approach, but working with entropy, we use  ${}^1H$  to define these diversities.

For a collection of  $N$  sites,  ${}^1H_\alpha$  is simply the mean of the site entropies:

$${}^1H_\alpha = - \sum_{j=1}^N \sum_{i=1}^S p_{ij} \ln p_{ij} / N \quad (4)$$

and hence, it follows that  ${}^1D_\alpha$  is the geometric mean of the site diversities (Eq. 2).  ${}^1H_\gamma$  is the entropy of PAs averaged across sites, and can also be expressed in an analogous form to  ${}^1H_\alpha$  (Eq. 4):

$${}^1H_\gamma = - \sum_{i=1}^S p_{i.} \ln p_{i.} = - \sum_{j=1}^N \sum_{i=1}^S p_{ij} \ln p_{i.} / N \quad (5)$$

where  $p_{i.}$  is the proportional abundance of species  $i$  averaged across the  $N$  sites. The turnover between sites is defined by  ${}^1D_\beta = {}^1D_\gamma / {}^1D_\alpha$  (Jost 2007), and hence it follows that  ${}^1H_\beta = {}^1H_\gamma - {}^1H_\alpha$ .

The shared formulation of  $\alpha$  and  $\gamma$  Shannon entropies as double summations over species and sites (Eqs. 4 & 5) is an essential step to the parameterised form of  ${}^1H$ , which in turn links to the multinomial model that we will use to relate the species proportional abundances to the environmental predictors.

## The Multinomial Diversity Model

The MDM has three components: (1) parametric entropy and diversity, (2) the multinomial model, and (3) the relationship between parametric entropy and the log-likelihood of the multinomial model.

## Parameteric Entropy and Diversity

The central objective of this study is to incorporate diversity into a statistical model, and to then use established statistical theory and practices to model relationships between diversity and environmental predictors. To meet this objective, we need to parameterise the definitions of entropy and diversity (Eq. 2). We do this by observing (1) the terms  $p_{ij} \ln p_{ij}$  and  $p_{ij} \ln p_{i.}$  (Eqs. 4 & 5) have the first  $p_{ij}$  in common and thus represent the data, and (2) the  $p_{ij}$  and  $p_{i.}$  of the logarithmic terms represent 'estimated values of  $p_{ij}$ ' for the  $\alpha$  and  $\gamma$  models respectively. We thus substitute the parameter  $\pi_{ij}$  for  $p_{ij}$  in the expression  $\ln p_{ij}$ , and define the parametric



forms of entropy  ${}^1H$ , and diversity  ${}^1D$ :

$${}^1H(\pi_{ij}; p_{ij}, w_j) = - \sum_{j=1}^N \sum_{i=1}^S w_j p_{ij} \ln \pi_{ij}/N \quad (6)$$

$${}^1D(\pi_{ij}; p_{ij}, w_j) = \exp({}^1H(\pi_{ij}; p_{ij}, w_j)) \quad (7)$$

We have also included site weights,  $w_j$ , to account for factors such as varying precision and sampling effort. Species weights,  $w_i$ , can also be included to account for factors such as varying size or mass of species. The weights would then take the form  $w_{ij} = w_i w_j$ .

Although parametric Shannon entropy and diversity are a generalisation of the definition of diversity (Hill 1973, Jost 2007), we retain the labels  ${}^1H$  and  ${}^1D$  to avoid unnecessary proliferation of terminology. The use of the superscript also avoids confusion with the general use of the terms 'entropy' and 'diversity', typically denoted by  $H$  and  $D$  respectively.

This parameterisation of  ${}^1H$  and  ${}^1D$  (Eqts. 6 & 7) enables us to represent the diversity of general configurations of PAs such as  $\pi_{ij} = f(\mathbf{x}_j, \theta_i)$  where  $\pi_{ij}$  depends on some specified function of the predictors  $\mathbf{x}$  and parameters  $\theta$ . It also provides the crucial link to the multinomial model, as shown below.

### The Multinomial Model

The multinomial model (MM) is a generalised linear model (Nelder & Wedderburn 1972; McCullagh & Nelder 1989) and is used to relate counts or proportions of multiple response data to a set of predictors. It extends the well-known logistic regression model from two to two or more response categories, and is thus ideal for relating species PAs to environmental and other predictors. The MM is given by  $\pi_{ij} = f(\mathbf{x}_j, \theta_i)$  where:

$$\pi_{ij} = \frac{\exp(\eta_{ij})}{\sum_{i=1}^S \exp(\eta_{ij})} ; \quad \eta_{ij} = \sum_{k=1}^K x_{jk} \theta_{ik} \quad (8)$$

The term  $\pi_{ij}$  represents the PA of species  $i$  at site  $j$ ,  $\eta_{ij}$  represents the linear effect of the predictors  $x_{jk}$ ,  $k = 1 : K$ , and the parameter  $\theta_{ik}$  represents the systematic effect of the predictor  $k$  on species  $i$ . The terms  $\exp(\eta_{ij})$  are constrained to values  $\geq 0$ , and dividing them by their sum constrains all  $\pi_{ij}$  to lie in the interval  $[0,1]$  and sum to one for each site as required for PAs (Eq. 8). We use the estimated values of  $\pi_{ij}$ , denoted  $\hat{\pi}_{ij}$ , to compare the probabilities of occurrence between species and/or sites, and these can be expressed as ratios of probabilities, odds or log-odds. Log-odds are an attractive way to express results since relationships between the log-odds of occurrence of two species and the predictors are linear, and can be expressed:

$$\eta_{ij} = \ln(\pi_{ij}/\pi_{i1}) = \sum_{k=1}^K x_{jk} \theta_{ik} \quad (9)$$

where  $\eta_{ij}$  is the log-odds of an observation for site  $j$  being species  $i$  for  $i = 2, 3, \dots, S$  compared to being species 1. Thus we can interpret MMs in terms of either probabilities (Eq. 8) or log-odds (Eq. 9). The expression for  $\eta_{ij}$  also links to the widely-used Gaussian species distribution model as described below.

## Linking Entropy to the Multinomial Model

We now link the parametric definition of entropy and diversity to the MM in order to complete the diversity model. The link is provided by the relationship between between entropy and the log-likelihood of the MM. Omitting constant terms, the weighted log-likelihood,  $LL$ , of the MM is given by:

$$LL(\pi_{ij}; p_{ij}) = \sum_{j=1}^N \sum_{i=1}^S w_j p_{ij} \ln \pi_{ij} \quad (10)$$

Thus the crucial link between  ${}^1H$  and  $LL$ , from Eqs. 6 & 10, is given by:

$${}^1H = -LL/N \quad (11)$$

This link is the foundation for modelling diversity in terms of environmental, spatial and temporal predictors, and is the crucial component of the MDM. It appears to be previously undiscovered in the literature.

The MDM provides maximum likelihood estimates of the PAs,  $\hat{\pi}_{ij}$ , conditional on the values of the predictors. The estimated  $\hat{\pi}_{ij}$  of the fitted MDM also have minimum  ${}^1H$  and  ${}^1D$ . Additionally, comparisons of two or more MDMs, translate into differences in  ${}^1H$ , or equivalently, into ratios of  ${}^1D$ . Thus the effects of individual predictors or collections of predictors can be quantified as  ${}^1H$ , converted to  ${}^1D$ , and interpreted as the effective number of species. Non-nested comparisons of models and comparisons within models, e.g. between levels of categorical predictor, can also be quantified as  ${}^1H$  as discussed below.

Because of this link, the MDM is an extremely effective and efficient toolbox for diversity analysis. It also is structured in the same way as the aspect of nature that it models. Environmental drivers affect species directly and diversity is altered as a result of those influences. That process is reflected in the MDM whereby we model change in species PAs due to environmental drivers, and that change translates into diversity.

As a practical consideration in model fitting, we note that deviance, and not log-likelihood, is the most widely used measure of fit of generalised linear models, and for the MDM equals  $-2LL$ . Thus, from Eq. 11, we can also express entropy as  ${}^1H = \text{Deviance}/2N$ .

In summary, the MDM is defined by three components:

1. The definitions of parametric entropy,  ${}^1H$ , and parametric diversity,  ${}^1D$  (Eqs. 6 & 7).  ${}^1H$  is the scale of calculations and is additive.  ${}^1D$  is the scale on which we express results as the effective number of species and is multiplicative.

2. The MM that estimates the relationships between the species PAs and the predictors (Eq. 8).
3. The relationship between  ${}^1H$  and the log-likelihood function of the MM (Eq. 11) that gives best estimates of PAs,  $\hat{\pi}_{ij}$ , for any chosen predictors, and simultaneously minimises  ${}^1H$  and  ${}^1D$ .

## A2: Analysis of a Simple Data Set

The analysis of this simple data set illustrates: (1) basic calculations of  ${}^1D$ , (2) equivalences between  $\alpha$ ,  $\beta$  and  $\gamma$  diversities and the deviances of MDM models, (3) relationships between the predictors and the species response curves, and (4) the differences between modelling site diversity ( $\alpha$ ) and the turnover of diversity between sites ( $\beta$ ). The data comprise single species observed on each of four sites with no species in common across sites (Table 7). The site ( $\alpha$ ) and total ( $\gamma$ ) diversities of the species data are 1 and 4 respectively and hence  $\beta$  diversity equals 4. This is the case irrespective of the order of diversity.

| 1 ( $\gamma$ ) | Site ( $\alpha$ ) | X | S1 | S2 | S3 | S4 |
|----------------|-------------------|---|----|----|----|----|
| 1              | A                 | 1 | 1  | 0  | 0  | 0  |
| 1              | B                 | 2 | 0  | 1  | 0  | 0  |
| 1              | C                 | 3 | 0  | 0  | 1  | 0  |
| 1              | D                 | 4 | 0  | 0  | 0  | 1  |

Table 7: Example data used to illustrate the variation of diversity of four species S1–S4 for the three predictors of the MDM: a constant (1), sites (Sites) and a continuous gradient (X).

The equivalence of the calculations of  $\alpha$ ,  $\beta$  and  $\gamma$  diversities and calculations of  ${}^1D$  from the MDM can be shown as follows. We fit two MDMs to the species data S1–S4. Model 1, with the constant predictor 1 ( $\gamma$  model), fits the overall mean of each species (0.25, 0.25, 0.25, 0.25) to all sites, whereas Model 2, with predictor Sites ( $\alpha$  model), fits the observed species data exactly. The diversities of the two MDM models are 4 and 1 respectively (Table 8). and  ${}^1D_\gamma = 4$  and  ${}^1D_\alpha = 1$ . Finally,  ${}^1D_\beta = {}^1D_\gamma / {}^1D_\alpha = 4$ .

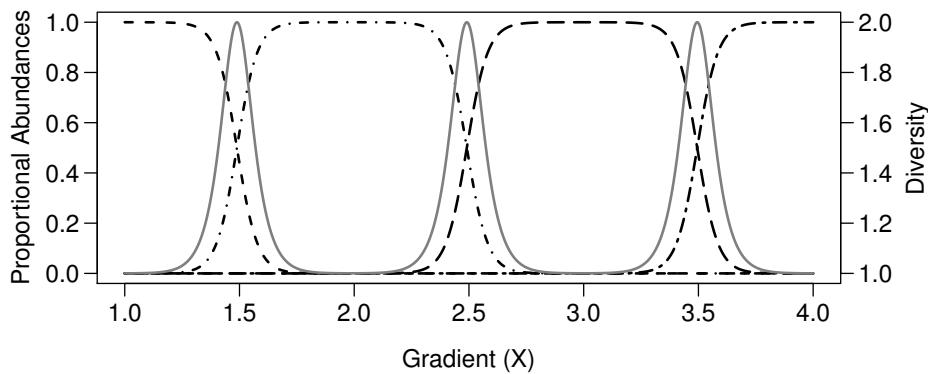


Figure 16: Response curves of the four species along the gradient X for the data from Table 7. The four black lines (different combinations of dashes and dots) represent the predicted probabilities of the species occurring at locations along the gradient. The solid grey line indicates the estimated site diversity along the gradient.

We can also fit the MDM to these data for the single quantitative predictor X (Tables 7 & 8). Since each species only occurs at one site, an exact fit is obtained, and the four sites are located along the gradient in the order of the quantitative predictor (Fig. 16). Estimated species response curves show PAs close to one for values near to their locations, 1, 2, 3 and

| Model                 | df | Dev   | $\Delta$ df | $\Delta$ Dev | $D$ | $\Delta^1D$ |
|-----------------------|----|-------|-------------|--------------|-----|-------------|
| Constant ( $\gamma$ ) | 9  | 11.09 |             |              | 4   |             |
| X                     | 6  | 0     | 1           | 11.09        | 1   | 4           |
| Sites ( $\alpha$ )    | 0  | 0     | 2           | 0            | 1   | 1           |

Table 8: The analysis of deviance and diversity for the example data. The predictors for each of the three models are defined, together with the degrees of freedom (df), deviance (Dev), changes in df ( $\Delta$ df) and deviance ( $\Delta$ Dev), diversity ( $^1D$ ) and proportional change in diversity ( $\Delta^1D$ ).

4, and close to zero elsewhere. Diversity rises from one at site locations to two at mid-points between site locations where neighbouring species have equal PAs of 0.5.

The analysis of diversity (Table 8) also illustrates the relationships between deviances and diversities. For each of the three models,  $^1D$  is calculated from single configuration of the PAs and represent mean site entropy and diversity. Conversely,  $\Delta^1D$  represents the change between two configurations of PAs and thus represents turnover. This distinction between the diversity of a single model and the ratio of diversities of two models is central to the interpretation of diversity analyses using the MDM.

This simple example also illustrates the difference between an MDM analysis of diversity, and analyses that directly relate site diversity to predictors. In this example, site diversity takes the value one for all sites, and does not vary with either site identity (A – D) or X. Hence any regression of site diversity on these predictors will show no change across sites. This contrasts with the MDM analyses above that account for turnover of species in addition to change in site diversities. Regressing site diversity on predictors does useful information, and can complement an MDM analysis, but it does not account for the turnover of species which is crucial in many studies.

### A3: Diversity of Order $\neq 1$

The MDM is based on  ${}^1D$ , but diversities of other orders ( $\neq {}^1D$ ), including  ${}^0D$  (richness) and  ${}^2D$  (Simpson) are also widely used by ecologists, and extending the MDM to include these diversities would significantly extend its usefulness. Unfortunately, the functional form of  $\neq {}^1H$  cannot be directly linked to the log-likelihood of the multinomial model, and hence we cannot directly extend the MDM to include  $\neq {}^1D$  in the same manner as  ${}^1D$ . To overcome this problem, we include a weighting term of the form  $p^q$  in  ${}^1H$  in such a way that the modified form of  ${}^1H$  has similar properties to  ${}^qH$ , but also retains its link to the MDM. The key to achieving this is based on l'Hopital's Theorem which shows that as  $q \rightarrow 1$ ,  ${}^qH$  takes the form:

$${}^qH = - \sum_{i=1}^S p_i^q \ln p_i / \sum_{i=1}^S p_i^q$$

Thus, replacing the  $p_i$  of  $\ln p_i$  with  $\pi_i$  to give us the parametric form, we define parametric entropy and diversity of order  $q$ ,  ${}^{q1}H$  and  ${}^{q1}D$ :

$${}^{q1}H = - \sum_{i=1}^S q p_i \ln \pi_i; \quad {}^{q1}D = \exp({}^{q1}H) \quad (12)$$

where

$$q p_i = p_i^q / \sum_{i=1}^S p_i^q \quad (13)$$

We use the notation to  ${}^{q1}H$  and  ${}^{q1}D$  to denote these weighted Shannon entropies and diversities and to distinguish them from  ${}^qH$  and  ${}^qD$ . This new formulation of diversity closely mimics  ${}^qD$ , but also enables us to relate  ${}^{q1}D$  to complex environmental drivers, beyond the currently limited application of  ${}^0D$  and  ${}^2D$  to simple hierarchies and turnover.

The properties of  ${}^{q1}D$  as an alternative to  ${}^qD$  are compelling. For  $q = 1$  it simplifies to  ${}^1D$ , and for all values of  $q$  it inherits all of the desirable properties of  ${}^qD$ , such as: (1) obeying the doubling rule, (2) for  $N$  evenly distributed species the diversity will be  $N$  for all values of  $q$ , and will be 1 for sites with a single species occurrence, and (3)  ${}^{01}D \geq {}^{11}D \geq {}^{21}D$ . Additionally,  ${}^{q1}D$  can be used within the MDM model to relate diversity to environmental drivers. Empirical comparisons of  ${}^{q1}D$  and  ${}^qD$  (Appendix Two) suggest that (1)  ${}^1D$  is typically the most powerful measure to detect differences in diversity, and (2) that differences between  ${}^{q1}D$  and  ${}^qD$  are typically small compared to their variation. Although it would be premature to suggest the replacement of  ${}^qD$  by  ${}^{q1}D$ , it is clearly reasonable to use  ${}^{q1}D$  within the MDM to weight diversity by relative species abundance; a step that substantially broadens the scope of the MDM and addresses the needs of ecologists.

## A4: The Across-Along Spatial Coordinates of the GBR

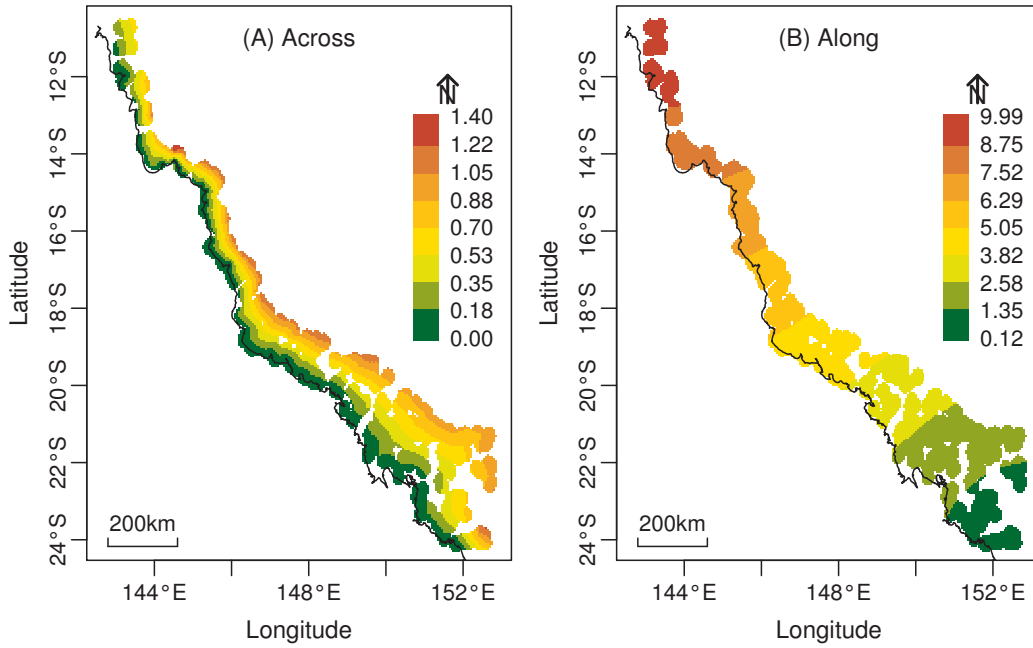


Figure 17: The across-along coordinate system of the Great Barrier Reef showing the across (A) and along (B) values in terms of longitude and latitude coordinates.

The geo-locations of the sites were defined by mean depth and relative distance across and along the GBRMP. Distance across takes the values 0 on the coast and 1 on the outermost edge of the continental shelf (80 m isobath), and distance along the shelf takes the value 0s on the southern edge of the GBRMP and 1 on the northern edge (Fig. 17).

The across-along spatial coordinate system provides a better basis for both explaining and predicting biological and environmental processes of the GBR (e.g. Fabricius & Death 2001) since such processes are often strongly influenced by the coastline and edge of the outer reefs, and by many of the reef structures that tend to align parallel to those two boundaries. The across coordinates also scale the reef to be equally wide (from coast to outermost edge of the continental shelf) along its whole length.