Modelling seamount diversity and biogeography

Derek Tittensor

With: M. Clark, J. McPherson, R. Myers, A. Rogers, K. Stocks, & CenSeam DAWG

28th March 2006, ISA, Kingston, Jamaica
Overview

1. Modelling patterns of diversity and endemism on seamounts
   (pessimism)

2. Habitat suitability modelling for seamount corals
   (optimism)
Section 1: Modelling diversity and endemism
Sampling issues on seamounts

Data originally from Shirshov Institute, Russia. Some of the most comprehensive sampling of fish and invertebrates on seamounts in Seamounts Online (http://seamounts.sdsc.edu/) using standardised sampling gear is at these locations.
Four methods of sampling on these seamounts

- **Fish**
- **Invertebrates**

Size of circle represents number of samples

Zeros & taxonomic scope key issue for seamount data
Variation in sampling

- There were ~10 different sampling methods used to collect fish & invertebrates from these seamounts.

- This figure doesn’t include differences between the same sampling method; e.g. different mesh sizes on nets.

- This is one of the biggest challenges when synthesizing seamount data for a large-scale analysis – very difficult to correct for variation in sampling effort.
Species accumulation curves


No sign of an asymptote

Jumeau West
Species accumulation curves for different sampling methods:

Vastly different number of invertebrates collected by different sampling methods.
Species accumulation curves

![Graph showing species accumulation curves with two lines representing different sampling methods: Fish - Bottom Trawl and Inverts - Fish Trawl. The x-axis represents the cumulative number of samples, and the y-axis represents the mean number of species. The graph includes error bars indicating variability.]

Great Bol’shaya
How to work with this data?

- Seamounts are extremely undersampled. Can we do anything about this?

- **Rarefaction** to standardise sampling effort – but does not provide useful information when sampling methods are different.

- **Non-parametric estimators** (e.g. Chao1, Chao2). Typically do not converge with data patterns such as those shown.

  *(pessimism)*
Chao 2 non parametric estimator

Fish collected by bottom trawl, Great Bol'shaya

No. of species

Number of samples

Species Observed
Chao 2 prediction
Chao 2 non parametric estimator

Fish collected by bottom trawl, Great Bol'shaya

- Species Observed
- Chao 2 prediction
- 95% c.i.
Endemism on seamounts

- What are the factors driving patterns of endemism?
- Can we construct theoretical models of endemism on seamounts?
A hierarchical model of endemism

Species richness:
- Entire marine environment
- Ocean o
- Seamount chain c
- Seamount s

Endemic species:
- Endemic to >1 ocean
- Endemic to ocean
- Endemic to seamount chain
- Endemic to seamount

INCREASING SPATIAL SCALE
A key question...

How many of these endemics are true endemics, and how many are a product of incomplete sampling?

Misclassifications will have a big effect on the power of models to explain patterns.
Modelling endemism

- Does terrestrial island biogeography theory provide a suitable testbed for constructing simple models of endemism on seamounts?

- What factors may be important in determining % endemics on seamounts? Isolation, age, depth, size…?
Endemism upon seamounts

Simple plots to visually assess the effects of age, depth & geographical isolation

Nasca & Sala y Gomez ridges
Seamount age vs. endemicals

Percent endemics - fish and invertebrates

Needs a GLM to properly assess fit

These plots will change if endemics are reclassified
Seamount depth vs. endemics

Percent endemics - fish and invertebrates

Needs a GLM to properly assess fit
Distance from continental margin (geographic isolation) vs. endemics

Certainly not the full story
e.g. Tasmanian Seamounts

IBGT does not appear to be a good fit.

Percent endemics - fish and invertebrates

Needs a GLM to properly assess fit
To summarise

- Problems with correcting for sampling effort. This is a major issue.
- General patterns of endemism & the factors responsible are difficult to establish.
- These simple models (based on island biogeography) do not appear to provide a good fit to seamounts. Very data limited.

(pessimism)
Section 2: Modelling global habitat suitability for Scleractinian corals on seamounts
The underlying principle of habitat modelling

observed distribution + environmental factors = predicted distribution
Central question:

Can we predict seamounts likely to provide suitable coral habitat?
SAUP Seamounts by depth

Data from SAUP
Scleratininia by depth

Best sampled corals on seamounts but note huge spatial gaps in coverage

... are depths merely reflecting sampling bias?
We only have presence data; no absences. The zeros problem again.
ENFA – Environmental Niche Factor Analysis

Inputs: ecogeographical variables (EGV’s) such as temperature, salinity, chlorophyll; and a species presence map.

Summarises all variables into a few uncorrelated factors.

Takes only presence data into account.

Compares the species distribution to the ‘global’ (available) environmental habitat distribution.

Hirzel et al., Ecology (2002)
Species niche is a subset of the global environment.

- **Marginality** (deviation from the global mean)
- **Specialisation** (niche breadth)

\[
\text{Marginality} = \frac{\left| \mu_G - \mu_S \right|}{1.96 \sigma_G}
\]

\[
\text{Specialisation} = \frac{\sigma_G}{\sigma_S}
\]

Hirzel et al., Ecology (2002)

e.g. Brotons et al. 2004 Ecography 27: 437-448
ENFA - continued

In many respects similar to a PCA, but eigenvectors are assigned ecological meaning: first represents 100% of marginality, others the remaining specialization.

LIMITATIONS OF ENFA

- Assumes that ecogeographical variables (EGV’s) are multinormally distributed & represent important factors.
- Threshold selection for model is not simple (converting from habitat suitability % to p/a).
- *Sample range must reflect actual species range.*

Hirzel et al., Ecology (2002)
The general idea

Globally 1 degree gridded data for 0 – 5500m from the World Ocean Atlas, GLODAP project & elsewhere

Ecogeographical variables

Suitable habitat prediction

Species presence

Scleratinia by depth on a 1 degree grid
Coral habitat prediction

- Model suitable locations for Scleratinia globally against an environmental background of the global ocean down to 5500m.
- Then restrict it only to those locations that are known to have seamounts in the appropriate depth range. Cannot map directly to seamounts due to SAUP and coral data mismatches.
- Remember, we are only predicting suitable Scleratinia habitat. We do not know if it will actually contain coral.
Scleratinia Results

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Value</th>
<th>Expl.SPEC.</th>
<th>Cum.Expl.Specialisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.657</td>
<td>0.343</td>
<td>0.343</td>
</tr>
<tr>
<td>2</td>
<td>8.741</td>
<td>0.346</td>
<td>0.689</td>
</tr>
<tr>
<td>3</td>
<td>3.086</td>
<td>0.122</td>
<td>0.811</td>
</tr>
<tr>
<td>4</td>
<td>1.936</td>
<td>0.077</td>
<td>0.888</td>
</tr>
<tr>
<td>5</td>
<td>1.265</td>
<td>0.050</td>
<td>0.938</td>
</tr>
</tbody>
</table>

Score matrix:

<table>
<thead>
<tr>
<th></th>
<th>1 (34%)</th>
<th>2 (35%)</th>
<th>3 (12%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total CO2</td>
<td>-0.43</td>
<td>0.24</td>
<td>-0.44</td>
</tr>
<tr>
<td>Depth</td>
<td>-0.43</td>
<td>0.21</td>
<td>0.10</td>
</tr>
<tr>
<td>Temperature</td>
<td>0.41</td>
<td>-0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>% O2 sat.</td>
<td>0.39</td>
<td>0.85</td>
<td>-0.75</td>
</tr>
<tr>
<td>Alkalinity</td>
<td>-0.33</td>
<td>-0.08</td>
<td>0.00</td>
</tr>
<tr>
<td>Sfc. Chloro.</td>
<td>0.29</td>
<td>-0.02</td>
<td>-0.03</td>
</tr>
<tr>
<td>Dis. O2</td>
<td>0.27</td>
<td>-0.39</td>
<td>0.48</td>
</tr>
<tr>
<td>Salinity</td>
<td>0.23</td>
<td>-0.03</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Marginality: 1.411
Specialisation: 1.776

Remember that first factor accounts for all of the species marginality.
NOTE ON DISAGREEMENTS

Predicted habitat suitability at 500m, Scleractinia
Predicted habitat suitability at 1000m, Scleractinia
Predicted habitat suitability at 1500m, Scleractinia
Predicted habitat suitability at 2000m, Scleractinia
Octocorallia

- Presence data much more limited
- Model likely to have less power
- Model at a very preliminary stage
### Octocorallia Results

#### Eigenvalues

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.305</td>
<td>0.305</td>
</tr>
<tr>
<td>2</td>
<td>0.341</td>
<td>0.646</td>
</tr>
<tr>
<td>3</td>
<td>0.213</td>
<td>0.859</td>
</tr>
<tr>
<td>4</td>
<td>0.058</td>
<td>0.917</td>
</tr>
<tr>
<td>5</td>
<td>0.035</td>
<td>0.952</td>
</tr>
</tbody>
</table>

#### Score matrix

<table>
<thead>
<tr>
<th></th>
<th>1 (31%)</th>
<th>2 (34%)</th>
<th>3 (21%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>0.62</td>
<td>-0.53</td>
<td>0.33</td>
</tr>
<tr>
<td>Depth</td>
<td>-0.51</td>
<td>-0.77</td>
<td>-0.04</td>
</tr>
<tr>
<td>Salinity</td>
<td>-0.43</td>
<td>0.07</td>
<td>-0.16</td>
</tr>
<tr>
<td>Dis. O2</td>
<td>-0.29</td>
<td>0.19</td>
<td>0.65</td>
</tr>
<tr>
<td>Total CO2</td>
<td>-0.26</td>
<td>0.23</td>
<td>-0.17</td>
</tr>
<tr>
<td>Alkalinity</td>
<td>-0.13</td>
<td>-0.17</td>
<td>0.32</td>
</tr>
<tr>
<td>% O2 sat.</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.55</td>
</tr>
<tr>
<td>Sfc. chloro</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Remember that first factor accounts for all of the species marginality.

**Marginality:** 0.769  
**Specialisation:** 1.800
Predicted habitat suitability at 1000m, Octocorallia
Predicted habitat suitability at 1500m, Octocorallia
The next steps...

- This workshop is a perfect opportunity to ‘ground truth’ these models

- Match to fishing effort & seamount density. (Spatial autocorrelation issues – can deal with these in a mixed-model spatial regression).
Map of seamount density per 1 degree grid cell.
Model calibration and verification

Cross-verification: 4 bins, 10 replicates

Scleractinia
Octocorallia
Other potentially important factors

- Current velocity – filter feeders. There may be a scaling issue here as small-scale turbulence may be very different from regional current average
- Substrate type
- Seamount diameter/height as a measure of patch size
- Distance to nearest seamount chain
- Many other possibilities
What else can we do?

- Compare outputs from multiple appropriate models (e.g. maximum entropy models for absence only data) for verification purposes (model averaging)

- Compare to data from other (non-seamount) deep sea Scleractinia; differences, similarities

- Community based models use commonly associated species as a ‘proxy’ for presence records (optimism)
In Conclusion

- Data quantity and differences in sampling methodology are two key limiting factors for modelling diversity on seamounts.
- Need to further develop statistical tools for these kinds of data.
- Having data with presence/absence (i.e. zeros) opens up a much wider variety of modelling techniques.
- Apply appropriate analysis techniques for the quality and quantity of the data available.
Acknowledgements

- FMAP – Future of Marine Animal Populations
- CenSeam (especially DAWG)
- ISA
- Tony Koslow