An Exploratory Analysis of the Effects of Sampling in Marine Surveys for Biodiversity Estimation

Hideyasu SHIMADZU¹  Ross DARNELL²

¹ Geoscience Australia  ² CSIRO Mathematical Information Sciences
Outline

1. Background
2. Data
3. Impact of sampling process
4. Summary
Background

**CERF project:** [http://www.marinehub.org/](http://www.marinehub.org/)
The Commonwealth Environment Research Facilities (CERF) Marine Biodiversity Hub prediction project analyses patterns and dynamic of marine biodiversity to determine the appropriate units and models for effectively predicting Australia’s marine biodiversity.

The project administered through the Australian Government Department of the Environment, Water, Heritage and the Arts.

**Major contributors:** University of Tasmania (UTas); CSIRO Wealth from Oceans Flagship; Geoscience Australia (GA); Australian Institute of Marine Science (AIMS); Museum Victoria (MV).

**Aims:** Modelling the relationship between biodiversity (e.g., presence/absence, richness, biomass etc) and the environment (e.g., depth, %carbonate, %mud, $O_2$).
Niche theory

Hutchinson (1957)

Niche: The full range of environmental conditions under which an organism can exist.

Note: Niche is a sub-space of a hyper space spanned by some physical covariates (environmental conditions).
Great Barrier Reef Data

Pitcher et al. (2007)

Survey period: 2003 – 6
Number of observation sites: 1252
Methods: Trawl, Sled
Number of species observed: 2862

Note
Also physical covariates are given.
Sampling process

Species Class | Weight measure | Frozen, Preserved in Formalin or Ethanol | Species identification | Weight measure in laboratory

1/2 Subsampling

Species classes
Algae, Brachiopoda, Bryozoa, Crustacea, Porifera, Seagrass, etc.
Subsampling

Widely used method in marine surveys. A part of the whole catch is sorted. Many variety of subsampling procedures (survey specific).

Note: Very little research on subsampling (eg Heales et al., 2003).
Model

Species presence/absence

Occupancy – Detectability – Sampling (Compound events)

\[ \Pr (S_3 = s_3) = \sum_{S_2} \sum_{S_1} \Pr (S_1, S_2, S_3) = \sum_{S_2} \sum_{S_1} \Pr (S_1) \Pr (S_2|S_1) \Pr (S_3|S_2, S_1) \]
Our approach

Restrict our attention on the effect of subsampling:

- Fully sampled vs. subsampled;
- Sled;
- Site matching
Site matching

For survey sites \( \{j : j = 1, 2, \ldots, n\} \), matched pair sites such that one is fully sampled and the other is subsampled are

\[ \left\{ (j, j') : \min_{j' \in \mathcal{J}'} \| z_j - z_{j'} \|, \ j \in \mathcal{J} \right\}, \]

where \( z_j \) is a vector of physical covariates at site \( j \).

Given a pair of species presence/absence \((Y_j, Y_{j'})\), odds ratio is given as

\[ \text{OR} = \frac{\sum_j y_j}{|\mathcal{J}|} \bigg/ \frac{\sum_{j'} y_{j'}}{|\mathcal{J}'|}, \]

where \(| \cdot |\) denotes the number of elements of a set.
Subsampled vs. fully sampled

When subsample was taken species are more likely to be observed...
<table>
<thead>
<tr>
<th>Odds ratio</th>
<th>Scientific name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.801</td>
<td>Magnoliophyta Liliopsida Hydrocharitales Hydrocharitaceae Halophila ovalis</td>
</tr>
<tr>
<td>1.643</td>
<td>Chlorophyta Chlorophyceae Caulerpales Codiceae Halimeda gigas</td>
</tr>
<tr>
<td>1.587</td>
<td>Chlorophyta Chlorophyceae Caulerpales Codiceae Halimeda borneenses</td>
</tr>
<tr>
<td>1.516</td>
<td>Bryozoa Gymnolaemata Cheilostomata Margarrettidae Margaretta spp</td>
</tr>
<tr>
<td>1.390</td>
<td>Bryozoa Gymnolaemata Cheilostomata Schizoporellidae Stylopoma spp</td>
</tr>
<tr>
<td>1.377</td>
<td>Chlorophyta Chlorophyceae Caulerpales Codiceae Udotea flabellum</td>
</tr>
<tr>
<td>1.376</td>
<td>Chlorophyta Chlorophyceae Caulerpales Caulerpaceae Caulerpa taxifolia</td>
</tr>
<tr>
<td>1.367</td>
<td>Chlorophyta Chlorophyceae Cladophorales Boodleaceae Phyllodictyon sp1</td>
</tr>
<tr>
<td>1.341</td>
<td>Chlorophyta Chlorophyceae Caulerpales Caulerpaceae Caulerpa brachypus</td>
</tr>
<tr>
<td>1.320</td>
<td>Phaeophyta Phaeophyceae Dictyotales Dictytotaceae Lobophora variegata</td>
</tr>
<tr>
<td>Odds ratio</td>
<td>Scientific names</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>0.014</td>
<td><em>Echinodermata Echinoidea Clypeasteroida Laganidae Laganidae</em></td>
</tr>
<tr>
<td>0.016</td>
<td><em>Porifera Demospongiae Hadromerida Tethyidae Xenospongia patelliformis</em></td>
</tr>
<tr>
<td>0.034</td>
<td><em>Arthropoda Crustacea Decapoda Calappidae Calappa</em></td>
</tr>
<tr>
<td>0.038</td>
<td><em>Mollusca Bivalvia Ostreoida Pectinidae Amusium pleuronectes</em></td>
</tr>
<tr>
<td>0.041</td>
<td><em>Arthropoda Crustacea Decapoda Xanthidae Actumnus squamosus</em></td>
</tr>
<tr>
<td>0.052</td>
<td><em>Arthropoda Crustacea Decapoda Parthenopidae Parthenope longimanus</em></td>
</tr>
<tr>
<td>0.052</td>
<td><em>Mollusca Bivalvia Pterioida Malleidae Malleus albus</em></td>
</tr>
<tr>
<td>0.056</td>
<td><em>Echinodermata Holothuroidea Holothuroidea Holothuroidea Holothuroidea</em></td>
</tr>
<tr>
<td>0.060</td>
<td><em>Mollusca Gastropoda Sorbeoconcha Bursidae Bufonaria rana</em></td>
</tr>
<tr>
<td>0.060</td>
<td><em>Chordata Actinopterygii Tetraodontiformes Monacanthidae Paramonacanth</em></td>
</tr>
</tbody>
</table>
Summary

- Challenges in marine survey data;
- The impact of subsampling;
- Need more investigation of the factor related to odds ratio.
Thank you for your kind attentions. Comments and suggestions are welcome!

Ross DARNELL  
Ross.Darnell@csiro.au

Hideyasu SHIMADZU  
Hideyasu.Shimadzu@ga.gov.au